CONSISTENCY AND COMPUTATION

Wittgenstein on the
Diagonal Argument

Dissertation zur Erlangung des akademischen Grades
Doktor der Philosophie (Dr. phil.)

Vorgelegt im Fachbereich 02
Geistes- und Kulturwissenschaften
der Universität Kassel
Von Frederic Kettelhoit

Datum der Disputation:
6. Juli 2022

ABSTRACT

The diagonal arguments by Cantor, Gödel and Turing constitute fundamental results that demonstrate limits to what is mathematically possible. These results can easily appear to us as limits that *must* hold, as if they were laws of nature that govern the ideal world of platonic numbers and logic, giving them an "ultraphysical" appearance of rigidity and hardness.

In all three cases, Wittgenstein critically examines the formal ideal of consistency and points out that the conclusions of the diagonal arguments only seem inevitable if we are not prepared to accept the contradictory result of the diagonalisation as an object in the formal system.

Wittgenstein's intent is not to advocate for a trivialist or paraconsistent treatment of inconsistency, since such an interpretation of the mathematical results would be just as dogmatic and philosophically one-sided as the interpretations that Wittgenstein is critically examining. From his perspective, consistency is not an ideal in and of itself, but merely a principle that has proven itself so useful in a large variety of language games that we accept it as an unquestioned rule even in cases where the situation is radically different.

In Wittgenstein's philosophy of mathematics, the actual language games that might appear to act only as examples or as motivation for their later formalisation are not merely primitive secondary stimuli for the primary formal system, they are instead essential for an understanding of the formal system to begin with, because the actual language games in all their variety lead to a surveyable representation of our concepts in a way that does not reveal itself by merely considering the uniform treatment in the formal system.

Wittgenstein's investigation of the three diagonal arguments shows what is at stake in these particular proofs: Although the mathematical proofs themselves are perfectly valid, we have the tendency to interpret them not as merely demonstrating logical impossibilities, but as "ultraphysical" impossibilities, comparable to laws of nature, governing the ideal realm of mathematics. Such a misleading picture is the result of a "one-sided diet", because we lack surveyability: We fail to grasp the concepts in all their various uses and in the context of how they fit into our form of life. The antidote is to describe them in a surveyable representation, sometimes by imagining different forms of life.

ZUSAMMENFASSUNG

Die Diagonalargumente von Cantor, Gödel und Turing sind grundlegende Ergebnisse, die die Grenzen des mathematisch Möglichen aufzeigen. Diese Ergebnisse können leicht als Schranken erscheinen, die gelten *müssen*, als wären sie Naturgesetze, die die ideale Welt der platonischen Zahlen und der Logik regieren, wodurch sie den "ultraphysischen" Anschein von Starrheit und Härte erhalten.

In allen drei Fällen setzt sich Wittgenstein kritisch mit dem formalen Ideal der Konsistenz auseinander und weist darauf hin, dass die Schlussfolgerungen der Diagonalargumente nur dann unausweichlich erscheinen, wenn wir nicht bereit sind, das widersprüchliche Ergebnis der Diagonalisierung als Gegenstand des formalen Systems zu akzeptieren.

Wittgenstein will nicht für eine trivialistische oder parakonsistente Behandlung der Inkonsistenz eintreten, da eine solche Interpretation der mathematischen Ergebnisse so dogmatisch und philosophisch einseitig wäre wie die von Wittgenstein kritisch hinterfragten Interpretationen. Aus seiner Sicht ist Konsistenz kein Ideal an sich, sondern lediglich ein Prinzip, das sich in einer Vielzahl von Sprachspielen als so nützlich erwiesen hat, dass wir es als unhinterfragte Regel auch in Fällen akzeptieren, in denen die Situation radikal anders ist.

In Wittgensteins Philosophie der Mathematik sind die konkreten Sprachspiele, die nur als Beispiele oder als Motivation für ihre spätere Formalisierung zu dienen scheinen, nicht bloß primitive Sekundärreize für das primäre formale System, sondern sie sind für das Verständnis des formalen Systems überhaupt erst wesentlich, weil die konkreten Sprachspiele in ihrer ganzen Vielfalt zu einer übersichtlichen Darstellung unserer Begriffe in einer Weise führen, die sich bei bloßer Betrachtung der einheitlichen Behandlung im formalen System nicht erschließt.

Wittgensteins Untersuchung der drei Diagonalargumente zeigt uns, was in diesen besonderen Beweisen auf dem Spiel steht: Obwohl die mathematischen Beweise selbst vollkommen valide sind, neigen wir dazu, sie nicht als bloße Demonstration logischer Unmöglichkeiten zu interpretieren, sondern als "ultraphysische" Unmöglichkeiten, vergleichbar mit Naturgesetzen, die das ideale Reich der Mathematik regieren. Ein solch irreführendes Bild ist das Ergebnis einer "einseitigen Diät", da uns Übersichtlichkeit fehlt: Wir begreifen die Begriffe nicht in all ihren verschiedenen Verwendungen und vor dem Kontext, wie sie in unsere Lebensform passen. Das Gegenmittel ist eine Beschreibung mithilfe einer übersichtlichen Darstellung, manchmal dadurch, dass man sich andere Lebensformen vorstellt.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Prof. Dr. Stefan Majetschak for his supervision, his advice and support during my work on this thesis as well as his help regarding all administrative hurdles along the way. I would also like to express my sincere thanks to Prof. Dr. Felix Mühlhölzer for his helpful feedback on drafts of this thesis, especially regarding the finer mathematical details of Wittgenstein's philosophy of mathematics.

I am deeply grateful for the generous financial support of the Otto-Braun-Fonds, without which this work would not have been possible. I would further like to thank the University of Kassel and especially Dr. Isabelle Schulze for answering all questions concerning the scholarship.

Additionally, I would like to express my sincere thanks to the Wittgenstein Archives Bergen for welcoming me during a research visit in 2019 and 2020, as well as for the stimulating and invaluable discussions that I had there with Prof. Dr. Alois Pichler, Dr. Jasmin Trächtler, Heinz Wilhelm Krüger and others.

I would further like to express my gratitude to the friends who shaped this thesis through their perspectives in our discussions or by reading drafts of the various chapters. I would especially like to thank Jasmin Trächtler, Diego Cabrera, Ricardo Henriques and Marie Weigert.

Lastly, I want to thank my family and especially my parents, who are probably at least as thankful as I am that this chapter is finally drawing to a close. The work of the last years would not have been possible without your support and advice. Written acknowledgements are long overdue and yet completely inadequate to express how grateful I am for your support, not just in this context, but for all the past years. Thank you!

(P.S. Apologies to Ludwig Wittgenstein, who would have surely hated this thesis, with all of its concessions to academic philosophy and style.)

CONTENTS

INTRODUCTION

> You might say, "How is it possible that there should be a misunderstanding so very hard to remove?"
>
> It can be explained partly by a difference of education.
>
> Partly by a quotation from Hilbert: "No one is going to turn us out of the paradise which Cantor has created."
>
> I would say, "I wouldn't dream of trying to drive anyone out of this paradise." I would try to do something quite different: I would try to show you that it is not a paradise — so that you'll leave of your own accord. I would say, "You're welcome to this; just look about you." [*LFM XI*, p. 103]

Some of the most confounding remarks in the philosophical writings of Ludwig Wittgenstein are puzzling not because it is hard to understand *what* Wittgenstein wants to say, but rather because it is often unclear *why* he wants to say it. This is especially true for the *Remarks on the Foundations of Mathematics*,[1] which were compiled from a heterogeneous collection of posthumously published documents and in the process sometimes heavily edited. Although they might appear as a coherent work authored by Wittgenstein, he himself never intended to write a self-contained philosophical work focused primarily on the philosophy of mathematics. The closest comparable attempt is the "pre-war version"[2] of the *Philosophical Investigations*, whose second part would have focused mostly on the philosophy of mathematics, but which was abandoned in favour of the version of the *PI* that we know today.

Given these circumstances, it seems puzzling that Wittgenstein chose to write extensively on rather specialised issues and authors in mathematics, such as Cantor's diagonal argument and Gödel's incompleteness theorem. These remarks share very few obvious connections with Wittgenstein's more general philosophy and might appear to be hardly related even to the rest of his more mathematical writings. It is then perhaps no surprise that these more specialised investigations drew the most criticism during the initial reception of the *RFM* and that Wittgenstein was accused of being out of his depth concerning some of the more advanced mathematical details of the proofs. Are these collections of remarks in the end only the missteps of an otherwise brilliant philosopher, all the more understandable in light of the fact that the writings in question were perhaps never meant for publication?

This thesis wants to offer an alternative viewpoint, namely that what might at first appear to be extremely specialised and only

1 From here on abbreviated as "*RFM*", see "Writings and Lectures by Wittgenstein" at the end of this thesis for a list of all abbreviations.

2 The "pre-war version" corresponds roughly to §§1-88 of the final version of the *PI*, written in 1936 and typed up in 1937 (see Stern, 2004, p. 15).

vaguely related collections of remarks are in fact closely connected not just among each other, but also to Wittgenstein's philosophy as a whole. Wittgenstein's writings on Cantor's diagonal argument, Gödel's incompleteness theorems and to a lesser extent even his remarks on Turing are linked not only through the obvious fact that all of these mathematical proofs are applications of the *diagonal method*, but also in their tendency to give rise to similar conceptual confusions. The following chapters attempt to show that the applications of *diagonalisation* and the *diagonal method* are not topics of niche interest, but rather symptomatic examples for many of the philosophical problems that Wittgenstein aimed to investigate and clarify throughout his life.

In light of the harsh criticism that Wittgenstein's writings on these diagonal arguments have faced, it is fruitful to state the philosophical assumptions and the overall perspective of the following chapters before jumping into the more mathematical aspects of the proofs. Most importantly among these assumptions, Wittgenstein's goal of *non-interference in mathematical matters* will be presupposed as a guiding principle of all of the remarks on the various diagonal arguments and will be assumed to hold in all of the following discussion. This goal can briefly be summarised as follows: A philosophical investigation in the sense of Wittgenstein cannot and must not contest the validity of mathematical proofs or calculations, but can only clarify conceptual misunderstandings, which arise at the frontier between the formalisms of mathematical calculi and their interpretation in 'prose' (Ms-127, 185.2 / *RFM V* §46; Ms-124, 138.3 / *RFM VII* §41), borrowing concepts from our everyday language. This goal or guiding principle is most clearly expressed in Wittgenstein's *Lectures on the Foundations of Mathematics* from 1939:

> That is not what I am going to do at all. In fact, I am going to avoid it at all costs; it will be most important not to interfere with the mathematicians. I must not make a calculation and say, "That's the result; not what Turing says it is." Suppose it ever did happen — it would have nothing to do with the foundations of mathematics. [*LFM I*, p. 13]

This goal of non-interference is also evident in various remarks in the *Nachlass*, for example in Ts-227a, 89.2 / *PI* §124 or in Ms-124, 82.2–82.3 / *RFM VII* §19.

Previous interpreters have often read Wittgenstein either as deliberately ignoring his own stated goal or as falling short of it, be it due to a shift in priorities and method or simply due to a lack of attention to detail.[3] The strong form of these and similar indictments make the

3 See for example Dummett, 1959, p. 326: "Certainly in his discussion of Cantor he displays no timidity about 'interfering with the mathematicians.'"; Steiner, 2001, p. 261: "Wittgenstein slips into trying to refute the theorem, in what he takes to be Gödel's proof, itself!"; Steiner, 2001, p. 263: "In other words, Wittgenstein – in defiance of his own doctrines and against his better judgment – attempted to refute an informal version of a mathematical proof."; Berto, 2009, p. 194: "Furthermore, I will not trust Wittgenstein's own declarations, according to which his remarks should not have any strictly mathematical import."

task of the reader considerably easier, as many of the more idiosyncratic remarks can be chalked up to a misplaced desire of contesting standard mathematical practice, but they hardly hold up in light of Wittgenstein's own goal and his considerable attention to detail in his other writings, as this thesis attempts to show. It must be pointed out, however, that the weaker form of this criticism, namely that some of the rather dogmatic and idiosyncratic remarks are a consequence of a temporary lapse or of a lack of attention to detail, cannot be entirely discarded, for two reasons:

First, it cannot be denied that Wittgenstein's thinking experienced a transition from more dogmatic to less dogmatic remarks over time and that this same tendency is at work in his mathematical writings, many of which were written before the (mostly undogmatic) final version of the *PI*. It would thus be unreasonable to assume that all of the more mathematical remarks in the years 1937–1944 could live up to the high standard of the very late writings of Wittgenstein.

Second, it is important to keep in mind that with the exception of the first part of the *RFM*, which is based on typescript Ts-222, none of the 'parts' of the *RFM* ever advanced beyond the manuscript stage. In the case of Wittgenstein, who usually worked on and reworked remarks in multiple manuscripts over many years, this is an indication that none of the remarks that will be discussed in the following chapters can be considered to have the same level of quality as those in the *PI*. The most 'polished' of Wittgenstein's remarks usually originated in small pocket notebooks, were then included, rearranged and reworded in one or more larger notebooks, before making their way into one or more typescripts, so that it is not uncommon for some remarks in the *PI* to have moved through 5 or more stages before reaching their final version. In the case of most of the parts of the *RFM*, nearly all remarks have at least 'survived' the draft stage of the pocket notebooks, but usually appear only in one, sometimes in two large notebooks.

It would therefore be futile to defend all of Wittgenstein's remarks in his more mathematical writings, as some of them may simply be the by-product of a developing train of thought, his way of experimenting with different ideas. However, this does not imply *carte blanche* and should instead lead to a more charitable reading, where remarks are discarded only if there does not appear to be a reading that is in line with Wittgenstein's conception of philosophy. In many cases, his more idiosyncratic remarks are not a consequence of misunderstandings or negligence on the part of Wittgenstein, but simply the result of his radically different outlook, compared to the large majority of working mathematicians and even other philosophers.

Before taking a closer look at some of the most relevant concepts of Wittgenstein's philosophy in the context of the next chapters, it can be helpful to sketch out the approach of this thesis in very broad strokes, with the main argument roughly summarised as follows:

It is only possible to clarify the philosophical issues at stake in the different diagonal proofs if these proofs are investigated in the context of their specific applications and uses. A *general* investigation that tries to understand these proofs through their shared mathematical *essence* will fail to do them justice philosophically, at least from the perspective of Wittgenstein's philosophy. When considered against the backdrop of their use, the different diagonal arguments share the tendency to present a *logical* impossibility as a kind of limitative result for what we can do in theory and even more so in practice, which then appears as a sort of "*ultraphysical*" impossibility. From the perspective of Wittgenstein, however, the impossibilities demonstrated by these diagonal proofs cannot be *physical* impossibilities (and neither "ultraphysical" impossibilities), only *logical* impossibilities. A philosophical investigation in the sense of Wittgenstein can clarify this difference by providing *surveyability* of the way these proofs use existing non-mathematical concepts and of the way these proofs are used in our (non-mathematical) practice. This is often achieved by imagining different forms of life, where the limits demonstrated by these proofs have a different role or standing. It will then become clear that the limits proved with the help of the diagonal method are *rules* of our language. A philosophical investigation of these proofs will leave each mathematical proof as it is, but can dispel conceptual confusion arising from a mistaken interpretation and application of the proof.

0.1 LOGICAL AND ULTRAPHYSICAL IMPOSSIBILITY

One of the central aspects of Wittgenstein's radically different outlook is his distinction between *logical* and *physical* possibility and impossibility. While the roots of this distinction can arguably be traced back to the distinction between logical and empirical propositions that is first developed in the *Tractatus*, one of the earliest notable and explicit mentions of the distinction between logical and physical impossibility appears as chapter 27 in the *Big Typescript*, where Wittgenstein also speaks of "ultraphysical" possibility and impossibility:

> 27) "Logische Möglichkeit und Unmöglichkeit". – Das Bild des 'Könnens' ultraphysisch angewandt. (Ähnlich: "Das ausgeschlossene Dritte".) [Ts-213, IIr.5]
>
> ---
>
> 27) "Logical Possibility and Impossibility". - The picture of 'being able to' applied ultraphysically. (Similar: "The Excluded Middle.")

The relationship between these three forms of possibility and impossibility can be roughly summarised as follows: A *physical* impossibility is a limit imposed on us by the laws of physics, such as the speed of light as the maximum speed that anything known to us could travel. Surpassing the speed of light is therefore (to our knowledge) a physical impossibility, because our *experience* tells us that achieving faster-than-light travel is impossible. However, if our experience changed and someone were to discover the possibility of faster-than-light travel, we would be prepared to revise our views and call faster-than-light travel a physical possibility.

In contrast, a *logical* impossibility is a possibility that is *ruled out* in the literal sense of the word, it is in other words excluded by the *grammatical rules* that govern our language use. A proposition such as "There is no reddish green" (Ms-133, 25r.2; Ts-229, 332.2; Ts-233a, 71.4; Ts-245, 244.7 / *Z* §346 / *RPP I* §624) is an example of a logical impossibility, because there is nothing that we *would ever call* a "reddish green", no matter what scientific discoveries are made in the future. Such a colour has no sense in our language and is excluded by the rules of our language games of colour concepts. Compared to physical impossibilities, it is much harder to see whether and how a logical *impossibility* could ever become a logical *possibility*: It seems impossible to even *imagine* what we could call a "reddish green" or how a logical law such as the law of the excluded middle, $p \vee \neg p$, could ever *not* hold.[4] While the concepts of logical and physical (im-)possibility run through the whole *Nachlass*, appearing as late as 15.4.1951 in Ms-176, 50r.3 / *LW II* and thus less than a month before Wittgenstein's death, the concept of "ultraphysical" possibility and impossibility occurs much more rarely in his writings, being confined mostly to the period of the *Big Typescript* and a single remark in *RFM I*. Strictly speaking, "ultraphysical" (im-)possibility is not a distinct concept, but rather Wittgenstein's term for mistaking what is actually a *logical* possibility or impossibility for a physical one:

> Ich darf aber doch nur folgern, was wirklich *folgt*! – Soll das heißen: nur das, was den Schlußregeln gemäß folgt; oder soll es heißen: nur das, was *solchen* Schlußregeln gemäß folgt, die irgendwie mit einer Realität übereinstimmen? Hier schwebt uns in vager Weise vor, daß diese Realität etwas sehr abstraktes, sehr allgemeines und sehr hartes ist. Die Logik ist eine Art von Ultra-Physik, die Beschreibung des 'logischen Baus' der Welt, den wir durch eine Art von Ultra-Erfahrung wahrnehmen (mit dem Verstande etwa). Es schweben uns hier vielleicht Schlüsse vor wie dieser: "Der Ofen raucht, also ist das Ofenrohr wieder verlegt." (Und so wird dieser Schluß gezogen! Nicht so: "Der Ofen raucht, und wenn immer der Ofen raucht, ist das Ofenrohr verlegt; also .....".) [Ms-117, 1.1, Ts-221a/b, 143.3, Ts-222, 11.3 / *BGM I* §8]

> But still, I must only infer what really *follows*! – Is this supposed to mean: only what follows, going by the rules of inference; or is it supposed to mean: only what follows, going by *such* rules of inference as somehow agree with some (sort of) reality? Here what is before our minds in a vague way is that this reality is something very abstract, very general, and very rigid. Logic is a kind of ultra-physics, the description of the 'logical structure'

4 Imaginability can thus be understood as a "criterion for logical possibility" (Trächtler, 2020), but not in the sense of "a matter of someone's power of imagination or phantasy" (Trächtler, 2020, p. 171), rather as being "bound to and restricted by language and grammar itself" (Trächtler, 2020, p. 174).

> of the world, which we perceive through a kind of ultra-experience (with the understanding e.g.). Here perhaps inferences like the following come to mind: "The stove is smoking, so the chimney is out of order again". (And *that* is how this conclusion is drawn! Not like this: "The stove is smoking, and whenever the stove smokes the chimney is out of order; and so...") [*RFM I* §8]

To misunderstand logical structure and thus logical possibilities and impossibilities as something "ultra-rigid"[5] (*LFM XX*, p. 197–99) or ultraphysical is according to Wittgenstein a source of conceptual confusion that is symptomatic for theories and quarrels in philosophy:

> In den Theorien und Streitigkeiten der Philosophie finden wir die Worte, deren Bedeutungen uns vom alltäglichen Leben her wohlbekannt sind, in einem ultraphysischen Sinne angewandt. [Ms-114, 10r.2; Ts-211, 747.2; Ts-212, 1191.1; Ts-213, 429r.3]

> In the theories and disputes of philosophy we find the words, whose meanings are well known to us from everyday life, applied in an ultraphysical sense. [Ms-114, 10r.2; Ts-211, 747.2; Ts-212, 1191.1; Ts-213, 429r.3]

Considering that according to Wittgenstein, the task of philosophy is not to propose theories (Ts-220, 76.2, Ts-239, 76.3, Ms-142, 102.2, Ts-227a/b, 84.2 / *PI* §109), an application of words in an "ultraphysical sense" is already a harsh indictment, but an earlier version of the remark is even more explicit, going so far as to call it a "wrong" application:

> In allen philosophischen Theorien finden wir Worte deren Sinn uns von den Phänomenen des täglichen Lebens her wohl bekannt ist in einem ultraphysischen Sinn, also falsch, angewandt. [Ms-107, 177.3]

> In all philosophical theories we find words whose sense is well known to us from the phenomena of everyday life applied in an ultraphysical sense, i.e. wrongly. [Ms-107, 177.3]

Why did Wittgenstein remove this mention of a "wrong" application from the subsequent versions? The most likely explanation is that

5 'Ultra-rigidity' and 'ultra-physicality' are closely connected to Wittgenstein's discussion of the "picture of a machine as a symbol of its mode of operation" in *PI* §§191–197 and more generally the notion of superlatives such as "Über-Regeln", "Über-Sprache" and "Über-Propositionales" (Schulte, 2021). Schultes interpretation seems generally compatible with the reading of "ultra-physical" possibility and impossibility proposed in this text, with a small exception: Schulte, 2021, p. 14 denies that the role of paraconsistent systems plays a role in Wittgenstein's thought and instead reads the mention of contradictory propositions as something "Überpropositionales" (Ms-125, 67r.2 / *RFM IV* §59) to mean that such such a 'contradiction' is excluded from logic altogether. This might be true for the specific remark discussed in this context by Schulte, but seems hard to defend against the backdrop of Wittgenstein's numerous other remarks on the position of contradictions in logic. Of course Wittgenstein was not a paraconsistent logician (nor a logician of any other school or conception of logic), but his remarks often argue precisely against prematurely excluding a contradiction from logic instead of treating it *within* logic (which then requires a paraconsistent approach). This does not make him a proponent of paraconsistency, but certainly a proponent for an openness towards a variety of logical systems.

calling it "wrong" is itself misleading: After all, calling something "ultraphysical" instead of merely "physical" is precisely meant to invoke that it is *not* physical, the term "ultraphysical" is thus not *wrong* in the strict sense. The problematic aspect is rather that this picture of the logical as 'sort-of-but-not-really-physical' leads to conceptual confusions, which can only be clarified through philosophical investigations and careful examinations of the use and grammar of the concepts involved. The misleading picture that such a conceptual confusion gives rise to is not wrong, it is only generalised beyond its means and thus appears as a general theory with *explanatory* power, when it should really remain restricted to a *description* of concrete cases.

That the picture of logic as something ultraphysical is by no means harmless becomes evident in another remark by Wittgenstein from 1931, two years after the remark from Ms-107 (with a similar remark also appearing later in Ms-112, 113v.5, but there without a mention of "Ultraphysik"):

> Die einzig würdevolle Aufgabe der Philosophie ist: den alten Götzen (der) {Ultraphysik // Philosophie} zu zerstören. (D.h. ihre einzige Verbindung mit Göttern.) [Ms-153a, 164r.3]

> The only dignified task of philosophy is: to destroy the old idol (of) {ultraphysics // philosophy}. (I.e. its only connection with gods.) [Ms-153a, 164r.3]

Coming back to the diagonal arguments discussed hereafter, one of the central theses of the following chapters is that the *applications* of these arguments by Cantor, Gödel and Turing (or sometimes by their interpreters) tend to be "ultraphysical" in the sense of Wittgenstein. Wittgenstein is interested in these matters not because his aim were to contest the validity of any of the proofs, but rather because these perfectly valid mathematical results are frequently applied in a way that makes them appear as demonstrating *ultraphysical* instead of merely *logical* impossibilities. One of the reasons is that the diagonal proofs by Cantor, Gödel and Turing all appear to demonstrate a *limit* to what is possible even under *ideal conditions in theory* and thus (this is the misleading part) all the more so in *practice*. It is clear that the limitative results of these proofs are not "physical" limits, but they seem to be so similar to limits such as the speed of light that they often appear to be "ultraphysical".

The impossibility of trisecting an angle using only Euclidean means shall serve as an introductory example that is much simpler than the diagonal arguments discussed hereafter: As Euclid's *Elements* demonstrate, there are many constructions that can be carried out using only a compass and a straightedge, such as the bisection of an angle or the construction of a regular pentagon. There are, however, also constructions that are *impossible* to carry out by using only a compass and a straightedge in the manner of Euclid, such as the trisection of an angle or the construction of a regular heptagon. The exact details of these proofs do not matter here, let us simply assume that we are presented with a valid proof of the impossibility of trisecting an angle.[6] Why do we give up any attempt to trisect an angle using only a compass and a straightedge once we understand the proof of the impossibility of such a construction? In contrast to the proof of the *possibility* of some particular construction, which guides us along a path of concrete steps until we finally reach the construction we set out to find and thus teaches us a certain method and capability that we can use henceforth, the proof of the *impossibility* of some particular construction does not teach us how to do something, but rather demonstrates a limit to our methods and capabilities, since it apparently predicts that no matter which path we embark on, we will never be able to reach our goal.

Seen from this angle, an impossibility proof might seem to survey a vast part of the mathematical landscape in a way that explores and tests an infinite number of possible routes in an instant, exhausting all avenues until we finally accept that there is no arguing with the proof and that any attempt to defy this mathematical prophecy is doomed to fail. Taken to the extreme, we might thus interpret such an impossibility proof as a *discovery* about the mathematical world, which adds to our knowledge about the *possibilities and limits in the ideal realm of Euclidean geometry*, a realm that is an idealised version of our practical reality. Faced with the *practical* task of trisecting an angle using only compass and straightedge, we could thus point to the impossibility proof as a justification that any attempt will fail and potentially give up the fruitless endeavour, based on the discovery that we made through the proof.[7] Seen from another angle however,

6 The choice of this particular example is no coincidence, Wittgenstein himself wrote about it on several occasions, most notably in Ts-227a, 199.3 / *PI* §334 and Ts-227a, 250.3 / *PI* §463. The example reflects many of Wittgenstein's views on impossibility results and therefore also applies to his remarks on diagonal arguments, but a full discussion would go beyond the scope of this thesis. See Floyd, 1995 for a discussion of the impossibility of trisecting the angle in the context of Gödel. As Floyd, 1995, p. 383 points out: "And yet, in *accepting* the proof [of the impossibility of trisecting the angle with straightedge and compass], we see that what they were trying to do was not only not done, but *could not possibly* (mathematically) be done." What Wittgenstein wants to investigate is exactly what it is that "could not possibly" be done, and how the merely parenthetical addition "(mathematically)" distinguishes this kind of impossibility from other forms of impossibility.

7 Wittgenstein's remarks contain numerous occurrences of mathematics as being invented instead of discovered, see for example *LFM I*, p. 22; Ts-222, 134.4 / *RFM I* §168; Ms-121, 43r.1 / *RFM II* §38; Ms-121, 27r.2. However, these remarks should only be read as an antidote against the one-sided view that mathematics is *always* discovered, not as the equally dogmatic view that mathematics is *always* invented. As Wittgenstein himself notes, sometimes it makes sense to say that we have *discovered* a new aspect in mathematics:

> I said: whoever invented calculation in the decimal notation surely made a mathematical discovery. But could he not have made this discovery all in Russellian symbols? He would, so to speak, have discovered a new *aspect*. [Ms-122, 87r.2 / *RFM III* §46]

the impossibility proof is merely a reflection of the *rules* that govern the Euclidean calculus. Proving that a trisection of an angle using only a compass and a straightedge is impossible comes down to saying that in Euclidean geometry there is nothing that we would *call* 'a trisection of an angle using only a compass and a straightedge' and that the proof simply shows how we use the word 'trisection' when we talk about Euclidean geometry. The proof itself does not and cannot govern the *application* of these ideal rules to any practical task at hand, because neither the proof nor the rules of the Euclidean calculus can on their own justify when and why these rules are applied. Both the calculus and the impossibility proof are thus *inventions* and there could certainly be something we would want to call 'a trisection of an angle using only compass and straightedge' if we invented another calculus and another proof.

This is not to say that the result of such an impossibility proof were arbitrary and that we could just as well have invented another calculus to prove the opposite result. The alternative to an interpretation of proofs as discoveries is not a free-for-all constructivism, where another calculus with different rules would work just as well as long as we all agreed on this new convention and decided to call something a valid trisection of an angle with compass and straightedge.[8] Such a reductionist view would ignore that the proof of the impossibility of trisecting an angle with compass and straightedge is useful precisely because *Euclidean geometry as a whole is useful* in a way that arbitrary other calculi are not. But viewing mathematics as invented instead of as discovered can nevertheless help us dispel the misleading picture of Euclidean proofs as something "ultraphysical", as if their application in the physical world were already implied in their logical rules.

These abstract reflections are of course hardly convincing for someone who is interested in carrying out practical constructions. Sure, we can certainly invent non-traditional applications for Euclidean geometry or invent a new calculus and try to apply it to our physical world, but is Euclidean geometry not in some sense the most 'natural' calculus for someone trying to trisect an angle using only a compass and a straightedge? Is the 'standard' impossibility proof not 'privileged' in a way that arbitrary other proofs would not be?

Let us consider the concrete example of someone who is tasked with the trisection of an angle using a compass and a straightedge

8 See also Floyd, 1995, pp. 390–391, which also highlights the "*decision*" involved (a wording that will become even more relevant in the context of Cantor and Gödel):

> Thus part of Wittgenstein's purpose in focussing on the formulation and the resolution of the trisection problem is to emphasize that there is no *absolute* requirement - mathematical or otherwise - that we restrict the conditions of "trisection" in the way we do. It is the *decision* to require that proofs be given within a particular setting, and that solutions take a particular form and be generally applicable which *generates* the unsolvable - this is, provably unsolvable, hence, resolvable - problem. Of course, this in no way renders the unsolvability of the task a matter of arbitrary human convention: God himself could not "trisect" an angle. But *we* can always, in Lakatos's words, bar - or create - what we *call* "monsters".

as a practical task. Armed with knowledge of the proof, this person might give up and point out the impossibility of such a construction. But what is it that is impossible here exactly? Is the answer perhaps that even if we might be able to trisect an angle more or less 'roughly' using these tools (through skill or sheer luck) we could never reach the same exact precision as in the case of the bisection of an angle? But given that the bisection of an angle is possible using Euclidean means and that we can *approximate* the trisection of an angle *via repeated bisections to an arbitrary degree*, even in theory, we are not practically limited in our ability to trisect angles using only Euclidean tools. Faced with the proof of the impossibility of trisecting an angle, someone could thus answer "I see that it is not possible to trisect the angle *in this way*, but here is how it *is* possible" and then proceed to 'trisect' an angle using repeated bisections as perfectly as practical physical limitations such as the size of an atom allow.

Of course such a 'trisection' would not be a trisection *in the Euclidean sense*, it would certainly not be *what we meant*.[9] Neither is such a 'trisection' in any way a *refutation* of the proof of the impossibility of trisecting an angle, the example is meant only as a very rudimentary way to illustrate that a theoretical impossibility need not necessarily restrict our practical possibilities in any way. Someone who proceeded to 'trisect' an angle through repeated bisection is definitely not trisecting according to the 'spirit' of Euclid's geometry, but does that mean that they are *wrong* to do so and lack the justification to regard their construction as a successful trisection, albeit not in the Euclidean sense?

This example of a trisection through an approximation of infinitely many bisections might at first seem far-fetched, but it is in other situations quite common in mathematics to treat an infinitely close approximation as equivalent to the result that is being approximated, for example the repeated division that leads to periodic decimal numbers such as $0.33333... = \frac{1}{3}$ or the geometric series as an infinite sum of terms such as $\frac{1}{3} = \frac{1}{4} + \frac{1}{16} + \frac{1}{64} + \frac{1}{256} + ...$, a series which corresponds quite directly to a trisection through repeated bisection. The point is not that we *should* call a trisection-through-approximation a trisection, only that under different circumstances we conceivably *could* call it so and that we might for practical reasons do so. To be explicit, the trisection-through-approximation example is of course

9 Cf. Floyd, 1995, p. 390:

> Hobbes not only boasted that he had trisected the angle, but also that he had squared the circle and doubled the cube: three equally impossible feats. What Hobbes really did was to give a method of construction *approximating* a solution. In fact it is possible to trisect in Euclid any arbitrary angle within close approximation. But this sort of "solution", however ingenious, was not (as we say) "what was wanted". We demanded (or wished to know about the possibility of) an *exact* solution.

Here "exact" is to be understood in the mathematical and therefore *theoretical* sense, because it is of course possible to approximate a trisection in a way that is *practically exact*.

not a refutation of the classic impossibility proof, its point is to clarify the *use* of the concept of trisection 'outside' the purely mathematical realm of Euclidean geometry, it is in other words concerned with the 'standing' of the impossibility of trisection in our lives.

Of course simply calling a trisection-through-approximation a trisection would amount to little more than a name change, which would affect us only in so far as it would then require us to distinguish between trisections-without-approximations (which would correspond to what can be constructed using only compass and straightedge without approximation) and 'all' trisections (which would include approximations). Such a trivial renaming would not clarify the concepts involved in any way and would certainly fall short of the sort of philosophical investigation that Wittgenstein was interested in, since "to imagine a language means to imagine a form of life" (Ts-227a, 15.4 / *PI* §19). Let us thus try to imagine a different *form of life*, where the trisection of angles is performed in such a way that the distinction between trisections in the sense of Euclid on one hand and trisections that are only possible through approximation on the other hand ceases to matter:

Imagine a group of people that had not learned about Euclidean geometry from the Greeks, but rather developed it much later, for example primarily as an enjoyable and systematic way to draw beautiful patterns, only after the invention of binary arithmetic and the introduction of rudimentary computers. For these people, there would be little difference in terms of practicality between a bisection that can be carried out by hand in a few steps and a trisection that needs to be painstakingly approximated through a large number of successive bisections, because all of their 'Euclidean' drawings would be evaluated and calculated by computers, for which numerous repeated bisections are hardly more complicated than a single bisection. We could even imagine that their computers were *programmed* (for this very specific task of drawing systematically) by specifying a list of instructions purely in terms of Euclidean commands such as 'place the compass at the point so-and-so', 'draw a circle with radius so-and-so'. Someone might then at some point have tried to draw a picture that required a trisection of an angle, failed to find a solution and subsequently raised the question whether a trisection using their computers was in fact possible. Is it not likely that these people might consider the repeated approximation of a trisection through successive bisections a valid solution to this problem? If we ever came into contact with them and explained the proof of the impossibility of trisecting an angle using compass and straightedge, they might perhaps comprehend and even accept the proof without difficulties, but would dismiss its significance, considering that for them it is a valid but ultimately useless exercise that misses the point of trisecting an angle, similar to how from our perspective their 'solution' misses the point of Euclidean geometry.

Several aspects of the above example have been borrowed from Wittgenstein, who frequently imagines a calculation or a proof to be used purely as an ornament in the form of "Tapetenmuster" / "wallpapers" (Ms-117, 159.3; Ms-124, 137.2; Ms-127, 195.4; Ts-221a/b, 165.2; Ts-222, 26.2, with the last two remarks being nearly identical to the earlier Ms-117, 38.2, where "Tapetenmuster" is still missing; *LFM III*, pp. 36–37; *LFM VI*, pp. 59–63; and in passing in *LFM VII*, p. 70; *LFM XII*, p. 120; *LFM XVIII*, p. 171), to emphasise "not what role it plays in mathematics; because this suggests a wrong picture", but rather "to know what part of speech it is", because it can be "an entirely different part of speech from what you would expect it to be [from its role] in mathematics" (*LFM XVIII*, p. 171). Even the idea of people using calculating machines as purely practical tools without any knowledge of their mathematical principles appears in Wittgenstein's writings, notably in Ms-126. Wittgenstein imagines "that a calculating machine had come into existence by accident" (Ms-126, 30.3 / *RFM V* §2) or "that calculating machines occurred in nature, but that people could not pierce their cases" and that "people use these appliances, say as we use calculation, though of that they know nothing" (Ms-126, 35.3 / *RFM V* §4). A related idea is that of a trained "human calculating machine" (Ms-126, 33.4 / *RFM V* §3), which mechanically follows a certain system of rules (see also Chapter 3 on Turing's computing machines as "humans who calculate"). All these examples evoke the image of a mathematical activity in the larger context of a practical use, with the emphasis on a purely mechanical following of rules and less regard for the mathematical insight or any other psychological process which might ordinarily be involved in such a calculation. It should be noted that these remarks appear shortly after an unpublished remark in which Wittgenstein considers a practical use for Cantor's diagonal argument and the impossibility of the trisection of an angle (Ms-126, 10.3), underlining that for Wittgenstein these impossibility results are closely connected with questions of mechanical rule following and the use of mathematical signs "in *mufti*", their "use outside mathematics" (Ms-126, 30.4 / *RFM V* §2).

0.2 SURVEYABILITY, PHILOSOPHY AND MATHEMATICS

Imagining a group of people who have no use for the proof of the impossibility of the trisection using only compass and straightedge can shed light on the role of imaginability in relation to logical and ultraphysical impossibility: Imaginability can serve as the criterion for logical impossibility because it is impossible for us to imagine that any discovery or experience would lead us to reject the impossibility of trisecting an angle or the law of the excluded middle, in contrast to the possibility of imagining that a future discovery in physics might lead to a rejection of the speed of light as the ultimate speed barrier. The former is impossible to imagine not because we might lack imagination but because such a possibility is *logically excluded* from our rules of language, whereas the latter is 'only' physically impossible. But while we cannot imagine a logical impossibility against the backdrop of *our form of life*, we can often imagine *different forms of life with different rules of language*, where a logical possibility with resemblances to our own concept is not logically excluded. Such a logical possibility would certainly not be identical to 'our' concept, but can sometimes show enough similarities to our own use of the concept that we might see why other people would want to say that something logically impossible (in the context of our form of life) is possible (in the context of their form of life), without declaring them as "mad" (*LFM XXI*, p. 202).

These two different aspects of imaginability, imaginability as a criterion for logical possibility and imaginability of other forms of life, explain why Wittgenstein should not be read as espousing a particular mathematical position. To imagine language games and forms of life in which formal reasoning proceeds without principles such as the law of the excluded middle or the law of non-contradiction is not by itself an advocacy for a logico-mathematical position such as intuitionism or dialetheism but only the attempt to present alternative uses for certain concepts.[10] It is clear that based on such a conception, philosophy of mathematics as understood by Wittgenstein cannot argue for or against particular mathematical positions, because what is at stake is not the *truth* but rather the *usefulness* of a logical possibility or impossibility. While neither mathematics nor philosophy can provide justification for this usefulness, philosophy can help alleviate the "one-sided diet" which is caused by a "thinking with only one kind of example" and is the "main cause of philosophical diseases" (Ms-116,

10 While many of Wittgenstein's remarks are very compatible with dialetheism, for example (the position that there are true contradictions and that the law of non-contradiction does not hold), such a dogmatic position is quite alien to Wittgenstein's philosophy. From the perspective of Wittgenstein, it is nonsensical to posit the *existence* of true contradictions as if they were some sort of ultraphysical entities existing in the platonic realm of mathematics, but it might be sensible in certain situations to play language games in which the law of non-contradiction is not applicable. One aspect that appears at different points in most of these diagonal proofs is their tendency to reject the contradictory conclusion and thus to exclude inconsistency from the system in favour of some sort of hierarchy of consistent but incomplete stages. It shall be argued that (at least the later, but in some aspects also the early) Wittgenstein's aversion to this kind of higher-order scheme motivates his philosophical investigation of these matters and explains some of the more notorious remarks on logical contradictions. This does not make Wittgenstein a dialetheist, but makes certain paraconsistent readings possible and even fruitful. He should nevertheless not be read to say that we *must* or *should* accept inconsistency, only that a different form of life could lead to a different conceptual decision when faced with a choice between inconsistency and incompleteness.

255.3; Ms-120, 135v.2; Ts-227a/b, 291.2; Ts-228, 53.3; Ts-230a/b/c, 11.5 / *PI* §593).

This method of imagining new uses is closely related to another fundamental concept of Wittgenstein's philosophy: "surveyability" and "surveyable representation". The corresponding German terms "übersehen", "übersichtlich", "übersehbar", "Übersicht" and "Übersichtlichkeit" have unfortunately traditionally been translated in a variety of ways, which has led to this concept being relatively overlooked in the anglophone reception of Wittgenstein, at least in comparison to concepts that were translated more uniformly, such as "language game" or "form of life".[11] The terms appear several times in the *PI*: As Wittgenstein mentions in *PI* §5 in reference to Augustine's picture of language, a language game can help to "clearly survey the purpose and functioning of the words". A more explicit discussion of what he means by "survey", "surveyable" and "surveyability" occurs in §92, where Wittgenstein opposes his philosophy, which aims to understand the "nature of language" as something "that already lies open to view, and that becomes *surveyable* through a process of ordering", to the common understanding of the essence of language as something "that lies *beneath* the surface". The central remark in the *PI* on surveyability and surveyable representation, however, is §122, one of a series of remarks revolving around Wittgenstein's philosophical method and the goal of philosophy in general:

> Es ist eine Hauptquelle unseres Unverständnisses, daß wir den Gebrauch unserer Wörter nicht *übersehen*. – Unserer Grammatik fehlt es an Übersichtlichkeit. – Die übersichtliche Darstellung vermittelt das Verständnis, welches eben darin besteht, daß wir die 'Zusammenhänge sehen'. Daher die Wichtigkeit des Findens und des Erfindens von *Zwischengliedern*.
>
> Der Begriff der übersichtlichen Darstellung ist für uns von grundlegender Bedeutung. Er bezeichnet unsere Darstellungsform, die Art, wie wir die Dinge sehen. (Ist dies eine 'Weltanschauung'?) [Ts-227a/b, 88.3 / *PU* §122]

> A main source of our failure to understand is that we don't have *an overview* of the use of our words. a Our grammar is deficient in surveyability. A surveyable representation produces precisely that kind of understanding which consists in 'seeing connections'. Hence the importance of finding and inventing *intermediate links*.
>
> The concept of a surveyable representation is of fundamental sign- ificance for us. It characterizes the way we represent things, how we look at matters. (Is this a 'Weltanschauung'?) [*PI* §122]

A full discussion of the role of surveyability for Wittgenstein's philosophy would go beyond the scope of this introduction,[12] but two aspects of the remark should be highlighted: First, although surveyability is not explicitly mentioned as often as other more famous concepts in the *PI*, it is nevertheless of "fundamental significance" and

11 See Majetschak, 2016 on the problematic translations of this concept and for a discussion of its central role in Wittgenstein's philosophy.

12 See Baker and Hacker, 2005a, pp. 259–65, Baker and Hacker, 2005b, pp. 307–334 and Majetschak, 2016 for an extensive discussion of *PI* §122.

can therefore be assumed to play an important role for Wittgenstein's philosophy of mathematics, given that his more mathematical remarks are not clearly delineated but rather form a part of a larger philosophical corpus. Second, Wittgenstein emphasises "the importance of finding and inventing *intermediate links*", and it is significant that Wittgenstein explicitly calls out the *invention* of new intermediate links, which becomes especially relevant in the context of his remarks on the philosophy of mathematics.[13] Although an earlier version of what later became §122 first appeared in Ms-110 in 1931 and was revisited by Wittgenstein in several manuscripts and typescript over the next years, the addition of "Erfindens" / "inventing" occurred only much later, only appearing in Ts-239, 77f.1 in 1937 and then in Ts-227a/b in 1944. The versions in Ms-110, 257.3; Ts-211, 282.2; Ts-212, 1144.2; Ts-213, 417r.3; Ts-220, 80.2; Ts-237, 80.1 and Ms-142, 107.2 only mention "Findens" / "finding". The original version, written in the context of Wittgenstein's remarks on James George Frazer's *Golden Bough*,[14] reads as follows:

> „Und so deutet das Chor auf ein geheimes Gesetz" möchte man zu der Frazerschen Tatsachensammlung sagen. Dieses Gesetz, diese Idee, *kann* ich nun durch eine Entwicklungshypothese {ausdrücken // darstellen} oder auch, analog dem Schema einer Pflanze durch das Schema einer religiösen Zeremonie oder aber durch die Gruppierung des Tatsachen-Materials allein, in einer *„übersichtlichen"* Darstellung.
>
> Der Begriff der übersichtlichen Darstellung ist für uns von grundlegender Bedeutung. Er bezeichnet unsere Darstellungsform, die Art wie wir die Dinge sehen. (Eine Art der ‚Weltanschauung' wie sie scheinbar für unsere Zeit typisch ist. Spengler)
>
> Diese übersichtliche Darstellung vermittelt das {Verstehen // Verständnis} welches eben darin besteht daß wir die „Zusammenhänge sehen". Daher die Wichtigkeit {der *Zwischenglieder* // des Findens von *Zwischengliedern*}. [Ms-110, 256.6–257.3 / *GB*]
>
> ---
>
> One would like to say of Frazer's collection of facts 'And so the choir points to a secret law'. It is *possible* to represent this law, this idea, by an hypothesis of development, or again, in analogy with the schema of a plant, by

13 As Majetschak, 2016, p. 73 (footnote) points out, "'übersichtliche Darstellung' might not only be translated as 'surveyable *re*-presentation' but also as 'surveyable presentation'. The structure presented by the form, which Wittgenstein has in mind, must not factually exist." Another closely related reason in favour of translating it simply as "surveyable presentation" can be added here: As Wittgenstein makes clear throughout the *Nachlass*, there is no *essence* behind many of the concepts in our language that we could point to as an object of reference, there is thus no 'thing' that the surveyable 'representation' could *represent*. Rather, the surveyable presentation is itself a part of our language and interwoven with the very concepts that it surveys. It can present concepts and their intermediate links, but is not secondary or superficial to anything supposedly primary or essential. Although "surveyable presentation" might therefore be a better translation than "surveyable representation", the latter term will be used in this text, purely for reasons of familiarity and consistency with other literature.

14 See Majetschak, 2012 on the influence of Frazer on Wittgenstein and the central importance of the "Remarks on Frazer's *Golden Bough*" for Wittgenstein's philosophical method.

> the schema of a religious ceremony, or by grouping the facts alone, in a *'surveyable'* representation.
>
> The concept of a surveyable representation is of fundamental significance for us. It characterizes the way we represent things, how we look at matters. (A kind of *'Weltanschauung'* as it is apparently typical for our time. Spengler)
>
> A surveyable representation produces precisely that kind of understanding which consists in 'seeing connections'. Hence the importance of finding intermediate links. [*GB*[15]]

In contrast to the version in Ts-227a/b, the links to the morphology of Goethe and Spengler are made explicit in Ms-110[16], but more importantly Wittgenstein emphasises the difference between "an hypothesis of development" or a general "schema" (of a "plant" or a "religious ceremony") on the one hand and his method on the other hand, which proceeds "by grouping the facts alone, in a *'surveyable'* representation". While a hypothesis of development or a schema could be said to capture the essence of a concept and might be used to *predict* which yet undiscovered specimen would fall under a particular concept, a surveyable representation in the sense of Wittgenstein does not attempt to provide such a predictive model of the world. While a hypothesis of development could potentially become useless if it turned out that one its presupposed intermediate links had never existed, the same is not true for a surveyable representation: Since such a representation is not 'judged' based on the merits of its predictive power, an intermediate link can even be merely "hypothetical", as the purpose of such a representation is not to predict something, but rather to "sharpen our eye for a formal connection", as Wittgenstein explains in the next remark:

> Ein hypothetisches Zwischenglied aber soll in diesem Falle nichts tun als die Aufmerksamkeit auf die Ähnlichkeit, den Zusammenhang, der *Tatsachen* lenken. Wie wenn man eine interne Beziehung der Kreisform zur Ellipse dadurch {illustrieren wollte // illustrierte} daß man eine Ellipse allmählich in einen Kreis überführt; *aber nicht um zu behaupten daß eine gewisse Ellipse tatsächlich, historisch, aus einem Kreis entstanden wäre* (Entwicklungshypothese) sondern nur um unser Auge für einen formalen Zusammenhang zu schärfen.
> Aber auch die Entwicklungshypothese kann ich als weiter nichts sehen als {die // eine} Einkleidung eines formalen Zusammenhangs. [Ms-110, 257.4 / *GB*]
>
> ---
>
> A hypothetical link is not meant to do anything except draw attention to the similarity, the connection, between the *facts*. As one might illustrate the internal relation of a circle to an ellipse by gradually transforming an ellipse into a circle, *but not in order to assert that a given ellipse in fact, historically came from a circle* (hypothesis of development) but only to sharpen our eye for a formal connection.

15 Translation from Majetschak, 2016 and *PI* §122, with minor additions.
16 See Schulte, 1990 for a detailed discussion of the influence of Goethe and Spengler on Wittgenstein's surveyable representation.

> But I can equally see the hypothesis of development as nothing but a way of expressing a formal connection. [*GB*[17]]

As the example of the circle and the ellipse makes clear, the idea of inventing hypothetical intermediate links plays an important role even in Ms-110, 6 years before Wittgenstein explicitly emphasises the aspect of the "Erfindens" / "inventing" by adding it to the Ts-239 version in 1937. We can use invented intermediate links as part of a surveyable representation without negatively impacting its value because the adequacy of intermediate links is measured by the "formal connection" that we want to focus on, which is part of our grammar and expressed by logical and not empirical propositions. Of course, this does not mean that any hypothetical intermediate link would do the job: To use a square as a hypothetical intermediate link would be useless if our goal were to shed light on the formal connection between a circle and an ellipse.

This brings us back to *imaginability* as the criterion for logical possibility and impossibility. As has been mentioned before, imaginability in this context must not be understood as depending on one's power of imagination, so that one person possesses enough phantasy to imagine something as logically possible while another does not, but rather as a reflection of what is included in or excluded from a particular language game. A surveyable representation can help in this regard by providing examples that we are otherwise unable to imagine purely due to a lack of imagination, but which we readily accept as imaginable once they are presented to us. In this way, a nearly circular ellipse as an intermediate link between an elongated ellipse and a circle can compensate for a lack of imagination or phantasy, whereas a square as an intermediate link is useless for that purpose.

However, this distinction between 'imaginability as phantasy' and logical imaginability is not always clear cut, which is why it should not be stretched beyond its limits, or else the resulting interpretation will misread Wittgenstein as putting forth a dogmatic criterion that supposedly holds in all cases. Especially in his later years, Wittgenstein frequently imagines tribes of people with very different forms of life from our own, where a lot of imaginary background is necessary to make their seemingly "mad" behaviour (*LFM XXI*, p. 202) understandable to us. An example that appears both in the *LFM* and *RFM I* is the tribe of people who buy and sell wood by the area, regardless of how high it is stacked, so that one can increase or decrease the price by stacking it higher or lower, thereby changing its area (*LFM XXI*, p. 202–03; Ms-117, 47.2–49.2; Ms-118, 34v.5–36v.1; Ts-221a/b, 172.5–174.3; Ts-222, 116.5–118.3 / *RFM I* §§147–52). At first, these people appear mad to us, but we could certainly imagine a form of life that makes their way of 'calculating' appear understandable to us, for example if there were such an abundance of wood that any method of

17 Translation from Majetschak, 2016, with minor additions.

calculating a price would be merely ceremonial and the people would not feel cheated if someone restacked the pile to get a 'better' price. Here, the example implicitly presupposes a whole way of life, which requires us to imagine more than just a new use of a concept that we are familiar with but had missed due to a lack of imagination, as in the case of the ellipse and the circle.

Coming back to the example of people trisecting routinely through approximation by repeated bisection, such an imaginary group of people could serve as a first step toward an intermediate link in a surveyable representation. From a dogmatic perspective, the impossibility of a trisection using only compass and straightedge (together with its corresponding proof) is akin to the discovery of a mathematical fact with ultraphysical implications for what we can or cannot do, with any alternative as merely a misunderstanding bordering on madness. The example of a group of people with a different form of life can form an intermediate link between these two extreme viewpoints, of fact on the one side and madness on the other, not by attacking the validity of the proof (an act which would fall under the auspices of mathematics, not philosophy), but rather by attacking the role and standing of the proof as an ultraphysical discovery. By imagining how people could be entirely unimpressed by the proof, we can decouple the mathematical aspect of the proof, which demonstrates a *logical* impossiblity, from its interpretation in 'prose' as a limitative result with ultraphysical consequences.

The standing and "civic status" of a contradiction is explicitly called out by Wittgenstein as "the philosophical problem" in *PI* §125, a remark which is all the more relevant in the current context because the diagonal arguments of the following chapters all proceed via a *reductio ad absurdum*, with a contradiction arising at the same crucial point from the diagonalised construction. The remark is worth quoting in full, as it mirrors the approach of the following chapters:

> a) Es ist nicht Sache der Philosophie, den Widerspruch durch eine mathematische, logisch-mathematische, Entdeckung zu {beseitigen // lösen}. Sondern den Zustand der Mathematik, der uns beunruhigt, den Zustand *vor* der {Lösung // Vermeidung} des Widerspruchs, übersehbar zu machen. (Und damit geht man nicht etwa einer Schwierigkeit aus dem Wege.)
>
> b) Die fundamentale Tatsache ist hier: daß wir Regeln, eine Technik, für ein Spiel festlegen, und daß es dann, wenn wir den Regeln folgen, nicht so geht, wie wir angenommen hatten. Daß wir uns also gleichsam in unseren eigenen Regeln verfangen.
>
> Dieses Verfangen in unseren Regeln ist, was wir verstehen, d.h. übersehen wollen.
>
> Es wirft ein Licht auf unsern Begriff des Meinens. Denn es kommt also in jenen Fällen anders, als wir es gemeint, vorausgesehen, hatten. Wir sagen eben, wenn, z.B., der Widerspruch auftritt: "So hab' ich's nicht gemeint."
>
> c) Die bürgerliche Stellung des Widerspruchs, oder seine Stellung in der bürgerlichen Welt: das ist das philosophische Problem. [Ms-130, 14.5 & Ms-130, 12.2–12.3 & Ms-130, 13.5; Ts-228, 160.4 & Ts-228, 159.2–159.3 & Ts-228,

159.7; Ts-230a/b/c, 36.4 & Ts-230a/b/c, 35.2–35.3 & Ts-230a/b/c, 35.7; Ts-227a/b, 88a.1 / *PU* §125]

It is not the business of philosophy to resolve a contradiction by means of a mathematical or logico-mathematical discovery, but to render surveyable the state of mathematics that troubles us – the state of affairs *before* the contradiction is resolved. (And in doing this one is not sidestepping a difficulty.)

Here the fundamental fact is that we lay down rules, a technique, for playing a game, and that then, when we follow the rules, things don't turn out as we had assumed. So that we are, as it were, entangled in our own rules.

This entanglement in our rules is what we want to understand: that is, to survey.

It throws light on our concept of meaning something. For in those cases, things turn out otherwise than we had meant, foreseen. That is just what we say when, for example, a contradiction appears: "That's not the way I meant it."

The civic status of a contradiction, or its status in civic life – that is the philosophical problem. [*PI* §125]

Here Wittgenstein emphasises once again that the task of philosophy must not be to "resolve a contradiction by means of a mathematical or logico-mathematical discovery", but rather to investigate the troubling state of the contradiction and make it "surveyable". A purely mathematical perspective can easily lead to the belief that a contradiction is always a sign of trouble, which *forces* us to abandon a particular path so that we *must* move into a different direction, but as Wittgenstein points out, a contradiction is not always a sign of trouble, it can be merely a reflection of running into something unexpected, so that we say: "That's not the way I meant it." In such a case, we will often discard an assumption (*reductio ad absurdum*, see also Ms-126, 124.2–125.2), but we also *could* conceivably accept the rules in question as contradictory and use them in practice. Whether we do one or the other depends on the "civic status" of the contradiction, in other words on its role and standing in our form of life. It can certainly be the case that we reject an assumption that leads to a contradiction, but not because something physical or ultraphysical would *force* us to do so, rather simply because *this is what we do*. As Wittgenstein notes in another context: "No, it is not true that it *must* — but it *does* follow: we *perform* this transition" (Ms-117, 1.1 [p. 13–14], Ts-221a/b, 147.2, Ts-222, 15.1 / *RFM I* §12, see Section 1.1).

Up until this point, surveyability has been treated in this introduction as a uniform concept across all of Wittgenstein's remarks. But is this really a justified assumption? It must be pointed out that Wittgenstein's concepts sometimes changed in meaning and significance over the years and that he frequently mentioned surveyability not only in his more meta-philosophical remarks in the *PI*, but also in the *RFM*. This raises the question of whether Wittgenstein's notion of surveyability in the *PI* (together with his meta-philosophical remarks in Ms-110 and later Ts-211/Ts-212/Ts-213) can be considered to be the same as in his more mathematical remarks in the *RFM*, especially in *RFM III*, where the surveyability of proofs is of central concern.

At the risk of giving a disappointing answer, neither 'yes' nor 'no' seem to capture the relationship between surveyability in mathematics and surveyability in philosophy: It would be hard to argue that the two notions fully coincide, but it is similarly improbable that they are completely distinct, either, because there is a strong family resemblance between surveyability in mathematics and surveyability in philosophy. This is partly explained by the observation that 'surveyability in mathematics' can be understood in two ways: It can refer to the use of and need for surveyability in proofs or calculations that are worked on by *mathematicians*, or it can refer to the use of and need for surveyability in a surveyable representation of mathematical concepts as investigated by *philosophers*. Briefly, in contrast to the surveyability in philosophy, which results from "a grouping of facts" (Ms-110, 256.6 / *GB I*) and of "intermediate links" (Ms-110, 257.3 / *GB I*), the surveyability of a proof depends on being able to reproduce it repeatedly without error (Ms-122, 5r.2 / *RFM III* §1) and use the proof as a paradigm (Ms-122, 34v.2 / *RFM III* §14). It is undeniable that there are striking differences in form and content between a mathematical proof and a surveyable representation of the concepts involved in such a proof, with Wittgenstein leaving the mathematical work with all of its proofs and calculations to the mathematicians.[18] However, there is a close and important connection between the surveyability of a proof in the mathematical sense and a surveyable representation in the philosophical sense, as Wittgenstein's investigation of proofs of arithmetic in the Russellian framework of the *Principia Mathematica* makes clear (see also Section 1.4): While an unsurveyably large 'proof' of the addition of large numbers in Russell's notation might be of little concern to a mathematician, who would see it as merely impractical but nevertheless foundational for our more informal way of adding numbers, the unsurveyability of such a 'proof' is absolutely crucial to Wittgenstein because he flips the relation of the 'formal' Russellian 'proof' and our 'informal' way of adding natural numbers on its head: We recognise the Russellian 'proof' as a proof of the addition only because we already have a concept of addition that stems from our 'informal' way of adding numbers, which gives us a

18 This explains why Mühlhölzer, 2010 considers *mechanical reproducibility* to be the distinguishing characteristic of surveyability in *RFM III*, in contrast to Wittgenstein's concept of surveyability in the *PI*. A proof that cannot be reliably reproduced is not a proof. At least in this strict form, it is not clear why or how the same would be true for a surveyable representation in the philosophical sense. This is certainly a very valuable observation, which clarifies the relationship between surveyability in mathematics and surveyability in philosophy, but does not appear to be sufficient on its own to draw a hard line between the two notions. Crucially, both notions of surveyability situate a concept in a variety of uses and applications, which explains why surveyable representations in philosophy can act as an antidote to unsurveyable proofs in mathematics.

standard of measurement for what is to be considered a correct addition, in contrast to a large Russellian 'proof', which is impossible to reproduce without frequently making mistakes. In this way, Russell's uniform logical framework actually depends on a variety of more specialised and less formally-logical language games. It is this mathematical variety that Wittgenstein wants to survey, with a surveyable representation (in the philosophical sense) acting as a clarification of the use of a surveyable proof, or the lack of use of an unsurveyable 'proof'. The mistake of Russell is not mathematical but philosophical, because he views his logical system as a way to generalise the more specialised mathematical systems in a foundational system that is supposedly independent from them, while failing to see that this 'foundation' cannot be used in the absence of the more specialised systems.

Surveyability in mathematics and surveyability in philosophy are thus closely linked: Mathematicians will often strive for proofs that are mechanically reproducible and thus naturally make their proofs surveyable. Sometimes, however, a 'proof' will seem surveyable (and thus appear as a proof) when it actually is not usable as such, with the appearance of surveyability being suggested purely through its analogy with a less formal but surveyable counterpart, such as in the case of Russell's proofs of addition and our usual way of adding numbers. The conceptual confusion arising from this mistake can be clarified with the help of a surveyable representation in philosophy, which groups and *describes* the different uses of a mathematical concept without attempting to *explain* it in general.

0.3 METHODOLOGY AND SCOPE

As already mentioned, a critical investigation of diagonal impossibility proofs in the tradition of Cantor, Gödel and Turing might at first glance be seen as an attempt to contest their validity. After all, what could be the point of critically examining well established proofs, if not to find faults in them? Quite often then, commentators have read Wittgenstein's remarks on Gödel or Cantor as an attempt to supposedly show how these proofs are false or trivial. It should be reiterated that this is decidedly *not* the intention of the following investigation. In the process, it shall hopefully become clear that such an interpretation is at odds with a charitable reading of Wittgenstein, which is to say a reading that takes seriously his stated intention of mathematical non-interference. A philosophical investigation in the sense of Wittgenstein will leave mathematical work to the mathematician and abstain from passing judgement on the validity of proofs. Instead, the focus will be on investigating how these proofs are *used*, not only inside their respective fields, but also 'at the edges' of their practical applications. Such an investigation must at times depart from the problems that Wittgenstein himself studied explicitly and instead enter uncharted territory, precisely because the object of study cannot be limited to Wittgenstein and his body of work, but must encompass our use of mathematical proofs and their application. Understanding or explaining Wittgenstein's philosophy is therefore explicitly not the only goal of the following investigation, even though an application of Wittgenstein's remarks in new contexts might in the end have the welcome side effect of illuminating some of these very remarks.

The following chapters are therefore not purely exegetical and will in fact transition from a more exegetical investigation of authors that Wittgenstein explicitly engaged with (Cantor and to a lesser degree Gödel) to a discussion of more implicit connections (between Wittgenstein and Turing) that draws on more heterogeneous sources in Wittgenstein's *Nachlass*. Despite this strong focus on Cantor, Gödel and Turing, the issues at stake will at all times be investigated from the perspective of Wittgenstein and presuppose his conception of philosophy and philosophical methods. Given that the views of the 'later Wittgenstein' shifted sometimes quite substantially over the period of his later life and that there is no single canonical source that would be applicable in all contexts, some clarifications are in order: All the texts by Wittgenstein discussed hereafter stem primarily from the years 1937 to 1944 and are thus written shortly after the "pre-war version" of the *Philosophical Investigations*. These mathematical writings by Wittgenstein either directly lead up to the *PI* or are roughly contemporary with it, which justifies a reading from the perspective of the Wittgenstein of the *PI*.

Some of the more 'general' concepts that will be relevant in the coming chapters have already been mentioned and include concepts that originate in earlier periods of Wittgenstein's writings but remain relevant (for example the distinction between *logical, physical and ultraphysical impossibility* as well as *surveyability*), others serve as guiding principles for Wittgenstein's philosophy of mathematics, even if they are not always explicitly called out (for example his *intention of non-interference* and the distinction between *empirical and logico-mathematical propositions*).

Although most of the tools and concepts will be drawn from Wittgenstein's *Nachlass* and Wittgenstein will be the most important primary source over the course of the whole dissertation, each of the three main chapters will nevertheless focus on the application of a particular diagonal argument and structure the chapter according to the diagonal proof, not always following the exact order of Wittgenstein's remarks. In light of the fact that Wittgenstein's conceptual investigations are often independent of each other and do not build up to a systematic treatment or a grand unified theory, the three main chapters similarly only presuppose the more general concepts of Wittgenstein's philosophy sketched out in this introduction and can thus be read in any order, with each chapter focusing on a particular diagonal argument. The different threads will then be brought together in the concluding chapter, which by its nature acts as a slightly more general outlook, with all the problems that such a general perspective might bring in the eyes of Wittgenstein.

Two appendices attempt to situate Wittgenstein's view of diagonal arguments in a larger context, with the more philosophical Appendix A focusing on a more contemporary application of the diagonal argument in *Algorithmic Information Theory*, and Appendix B serving as a philological high-level overview of the editorial decisions in the *Remarks on the Foundations of Mathematics*.

The resulting investigations in the three main chapters and the appendices will hopefully elucidate both the philosophical issues at stake in the diagonal proofs as well as the reasons for Wittgenstein's approach and interest in these matters.

All of the following chapters will draw heavily on different *Nachlass* documents and quote from the "linear" transcriptions of these documents, as provided by the Wittgenstein Archives Bergen, usually with minor typographic or grammatical alternatives omitted. If the remark in question has been published as part of the *Remarks on the Foundations of Mathematics* or in other 'works', the section number in the published version will additionally be mentioned. Apart from making it easier to trace these remarks back to the *Nachlass*, direct references to the *Nachlass* documents are crucial in the context of the *RFM*, as this 'work' has the questionable distinction of being one of the most heavily edited works that have been published by Wittgenstein's literary executors after his death (see Appendix B for more details on the editorial status of this work). Many of the editorial omissions and changes are quite minor and often reasonable, but especially in the context of *RFM II*, which discusses Cantor's diagonal argument, the editorial interventions sometimes paint a different picture than Wittgenstein's own documents, which justifies a close inspection of the *Nachlass* documents.

The remarks that will be discussed hereafter are drawn primarily from the following documents in Wittgenstein's *Nachlass*:

1. Chapter 1 will focus on the documents belonging to part II of the *RFM*, which include a clearly delineated part of Ms-117 (probably written between the beginning of October 1937 and the end of June 1938) and selections from Ms-121 (written between the end of April 1938 and the beginning of January 1939). Further, the remarks discussed will also include many unpublished remarks from Ms-121 as well as from the entirely unpublished pocket notebooks Ms-162a and Ms-162b (which start in January 1939). Ms-117 and Ms-121 contain the first and also final versions of the remarks in *RFM II*, in other words there are no (surviving) earlier pocket notebooks or later notebooks

or typescripts that would contain identical or textually similar versions of these remarks. Additionally, a short fragment of 6 loose sheets (Ms-178d), which does not appear to have been discussed in existing literature, will be briefly examined and a more precise dating proposed.

2. Chapter 2 revolves primarily around the remarks on Gödel in appendix III of part I of the *RFM*, compiled from Ts-221a/b, 246–55 and most likely typed in 1938 (with sources in Ms-118 dating back to 1937), as well as several remarks on Gödel in *RFM VII*, compiled from Ms-124, 82–95 and written between 27.6.1941 and 4.7.1941. A passage from Ms-121, 72–85 on Gödel, which was not published as part of the *RFM*, and the continuation of the remarks from *RFM VII* in Ms-163 will also be discussed. The remarks belonging to *RFM I App III* in Ts-221a/b and to *RFM VII* in Ms-124 appear in similar versions also in other documents, for Ts-221a/b there are related versions in Ms-118 and Ts-223, for Ms-124 in the notebook Ms-163.[19]

3. Chapter 3 is, as mentioned, less exegetical and not strictly focused on a single corpus of documents in Wittgenstein's *Nachlass*. The main remark on Turing, published as *RPP I* §1096, appears in Ms-135, 59v.2 as well as in the typescripts Ts-229, 448.1 and Ts-245, 319.3. Most other remarks are drawn from Ms-124 and were in many cases published in *RFM VII*, with other versions of the remarks appearing mostly in Ms-161, Ms-163 and in some cases in Ms-127.

Given that the edited version of a remark published by Wittgenstein's literary executors and its source in the *Nachlass* sometimes differ substantially (with additional variants in the source documents left out by the editors), a short note on the way remarks are cited is in order: In all instances where a *Nachlass* source exists (which is the case for all remarks in the published 'works', but not the lectures) the version cited is from the *Nachlass*, with all variants except minor typographic variations preserved. Remarks are not cited by page, but rather by unique remark identifier, which combines the page of the beginning of the remark with a number distinguishing between remarks on the same page (with "Ms-124, 138.3" corresponding to the third remark on page 138 of manuscript 124 in the *Nachlass*, according to the von Wright catalogue, see Von Wright, 1993). In the case of remarks extending over multiple pages, the remark identifier refers to the page of the beginning of the remark. Whenever the remark in question has been published in a 'work' compiled by Wittgenstein's

19 There are a few other references to Gödel in the *Nachlass*, but they are all made in passing and do not constitute a sustained treatment of the topic, these remarks will therefore not be discussed in this text: Ms-117, 147.1; Ms-117, 151.5; Ms-122, 28v.2; Ms-124, 115.1 and Ms-126, 131.2.

literary executors, the sigil of the work (which are listed at the end of this thesis) together with its section number is cited, for example "Ms-124, 138.3 / *BGM VII* §41". English translations for German remarks are given in a two-column layout, with translations taken from the published work if the remark has been published (in which case variants are usually not preserved), indicated by the sigil specifying the work in question ("*BGM VII* §41"), and otherwise translations being my own (in which case variants are usually preserved), indicated by the lack of a sigil or any other reference.

0.4 RELATED WORK

The status of Wittgenstein's philosophy of mathematics in the context of Wittgenstein's philosophy as a whole has traditionally been problematic, a situation that is also reflected in its reception in the literature: For one, there is no 'work' authored by Wittgenstein himself that deals specifically with mathematical topics as there is in the case of the *Tractatus* or the *PI*, which are both based on typescripts created by Wittgenstein. While the mathematical remarks of Wittgenstein's 'middle period' in the *Philosophical Remarks* all stem from the typescript Ts-209, there is no such single manuscript for Wittgenstein's later writings on mathematics in the late 1930s and early 1940s. Apart from part I, the different parts of the *RFM* are all compiled from different manuscripts (of varying quality and editorial status).

Consequently, Wittgenstein's remarks on the diagonal arguments of Cantor, Turing and Gödel have received relatively little attention compared to the *PI* or the very late writings such as *On Certainty*. Furthermore, especially the initial reaction was mostly dismissive, with many remarks considered as "a surprisingly insignificant product of a sparkling mind" (Kreisel, 1958, p. 158). To this date, there does not appear to be a single monographic treatment of Wittgenstein's remarks on the diagonal argument across all the documents spanning the *RFM*, only publications that consider Wittgenstein's remarks on one of the three diagonal arguments in isolation.

Wittgenstein's remarks on Cantor are perhaps not as scandalous as his infamous remarks on Gödel, but they have nevertheless elicited their faire share of criticism, mostly due to Wittgenstein's rather dogmatic rejection of set theory. Many commentators consider Wittgenstein to fall short of his stated ideal of non-interference and non-revisionism, e.g. Rodych, 2000 and Putnam, 2007, and read his position in *RFM II* to be revisionist and as advocating philosophical theses. There are recent attempts to rehabilitate Wittgenstein's seemingly mistaken remarks, e.g. Wheeler, 2021, where Wittgenstein's rejection of set theory is interpreted not as a wholesale dismissal of the applications of set theory inside mathematics, but rather as a rejection of the sense of concepts such as denumerability *prior* to a proof such as Cantor's. Such a reading manages to explain Wittgenstein's rejection from the admittedly radical perspective of Wittgenstein's larger philosophy and is quite compatible with the reading in this thesis.

Given the largely negative reception of Wittgenstein's remarks on Cantor's diagonal argument, it is not surprising that detailed exegetical readings of these remarks have been exceedingly rare. The only monographic treatment of *RFM II* appears to be Redecker, 2006, which unfortunately does not include a discussion of any of the unpublished remarks in Ms-121. While Redecker finds value in Wittgenstein's remarks, she nevertheless reads Wittgenstein as falling short of his ideal of non-interference in mathematical matters and to advocate theses (see e.g. Redecker, 2006, pp. 31–50 and especially Redecker, 2006, pp. 319–320). Similarly, Ramharter, 2018 holds the view that Wittgenstein's remarks in *RFM II* contain definite mathematical errors which are attributable to his lack of a proper mathematical background, even if many of the remarks offer interesting philosophical insights. More charitable readings, which also include a discussion of the unpublished remarks in Ms-121, Ms-162a and Ms-162b, are presented in the chapters of Mühlhölzer, 2020 and Floyd, 2020. Floyd's chapter takes a broader perspective and includes an examination of the remarks on Cantor in the context of Turing (which is relevant for Chapter 3), whereas Mühlhölzer presents a closer reading of Wittgenstein's remarks and offers a charitable interpretation that is closest to the reading in this thesis.

Most charitable interpretations take Wittgenstein to attack the prosaic interpretations of the proof, not the mathematical proof itself (the reading in the following chapters also falls into this category), while more critical interpretations read Wittgenstein as attacking the mathematical proof, either deliberately or by mistake. A third option, which is put forward in both Steiner, 2001, pp. 269–270 and Gefwert, 1998, p. 246, is to deny that Cantor's diagonal argument is a genuine mathematical proof in the eyes of Wittgenstein (which then turns his seemingly mathematical critique into an attack of a purely philosophical pseudo-proof). While such a reading might be interesting, it is challenging to see how this would resolve the accusation of interference and revisionism, as such an interpretation challenges the mathematical status of a universally accepted proof.

The remarks on Cantor's diagonal argument have sometimes been interpreted as an attack on the mathematical concept of infinity, with Wittgenstein advocating for a form of finitism (Marion, 1998) or even strict finitism (Dummett, 1970) and thereby taking a revisionist position in mathematics. Marion focuses on Wittgenstein's middle period, which certainly exhibits a stronger dogmatism and more pronounced finitistic tendencies, but Marion, 1998, pp. 193–202 explicitly connects this finitistic interpretation with Wittgenstein's remarks on Cantor in *RFM II*. Evidence against such a finitistic reading is provided by Frascolla, 2006, pp. 142–156, where Wittgenstein is instead read as "quasi-revisionary": not attacking or attempting to refute mathematical results, but working towards philosophical clarification that might lead mathematicians to lose interest and thereby give up certain branches of mathematics of their own accord (Frascolla, 2006, p. 160). Such a "quasi-revisionary" interpretation does not conflict with Wittgenstein's goal of non-interference and is entirely compatible with the following chapters.

Wittgenstein's remarks on Gödel's first incompleteness theorem have without a doubt generated the most interest out of all three diagonal arguments. The initial reviews of the published remarks in *RFM I; App. III* and *RFM VII* by Kreisel, 1958, Bernays, 1959 and Dummett, 1959 were scathing and for the next 30 years Wittgenstein's remarks on Gödel were seen as the mistaken attempt to refute a mathematical theorem that was beyond reproach.

This started to change with Shanker, 1988, who presented the first charitable reading of Wittgenstein's remarks and interpreted them as attacking Gödel's *interpretation* metaphysical of the theorem, not the mathematical theorem itself. Shanker's paper was followed a few years later by another equally charitable reading, Floyd, 1995, which presented Wittgenstein's remarks on Gödel in the context of remarks on Wittgenstein's favourite example of impossibility proofs: the impossibility of trisecting an angle using only compass and straightedge. In comparison to Shanker, Floyd presented a much closer reading of the remarks in *RFM I; App. III*. (These two charitable interpretations are closest in spirit to the reading presented in this thesis, with the notable exception of how Wittgenstein's remarks on inconsistency are interpreted.)

Rodych, 1999 and Rodych, 2002 presented a reading that was less charitable and saw definite mistakes to Wittgenstein's remarks, while nevertheless emphasising the value of the remarks in the context of Wittgenstein's philosophy as a whole. The papers thus exhibit an interesting balance and show how Wittgenstein's remarks can be rehabilitated even if one understands him as a dogmatic philosopher advocating for a particular position in logic. (Rodych's emphasis on the role of inconsistency and the conditional nature of Gödel's proof is entirely compatible with the interpretation in the following chapters, although his interpretation of Wittgenstein as advocating a dogmatic position is explicitly rejected.)

Floyd and Putnam, 2000 constitutes a seminal paper in the debate around Wittgenstein and Gödel, as it introduces what is now known as the "Floyd-Putnam Thesis". According to the thesis, Wittgenstein's "notorious paragraph" in *RFM I; App. III* §8 shows a remarkable insight on the part of Wittgenstein and is to be understood as a model-theoretic clarification of Gödel's theorem. The Floyd-Putnam Thesis has led to a lively debate, including rebuttals from Steiner, 2001, Bays, 2004, Bays, 2006, answers in Floyd and Putnam, 2006, Floyd and Putnam, 2008, as well as further rebuttals of the rebuttals (specifically of Steiner, 2001), e.g. Rodych, 2003 and Rodych, 2006. The debate is far from over, with papers as recent as Lajevardi, 2021. (For the present thesis, the Floyd-Putnam Thesis is only of minor importance, but nevertheless presents a very interesting insight into Gödel's theorem.)

In addition to the charitable readings by Shanker or Floyd, dismissive interpretations by Steiner or Bays, and readings that fall somewhere in between by Rodych, there is also another notable position that reads Wittgenstein as presenting a *paraconsistent* reading that explicitly accepts the diagonal conclusion as a contradiction, e.g. Priest, 2004 and Berto, 2009. These readings interpret Wittgenstein rather charitably, but nevertheless ascribe to him a particular logical position. As a result, they have often been dismissed by Wittgenstein researchers as attributing to Wittgenstein a view that goes against his intention of not advocating a particular position in logic (for example in the close reading of Kienzler and Grève, 2016), but offer the advantage of explaining some of Wittgenstein's more outrageous remarks on inconsistency in logic. (This thesis will read these particular remarks in a way that is often compatible with paraconsistent readings, while rejecting any attempt to treat this position dogmatically as the single and only correct position that one could take regarding the different diagonal arguments.)

Among Wittgenstein scholars, the least discussed of the three diagonal arguments investigated in this thesis is certainly Turing's diagonal argument. This is explained by the fact that direct references in the *Nachlass* to Turing are exceedingly rare and there are only three direct points of connection that could be discussed in the context of Wittgenstein:

First, one might focus on Turing's philosophy of mind, more specifically the idea of the "Turing test" for machines, and discuss the question of whether machines can be said to think from the general perspective of Wittgenstein's philosophy of psychology, without limiting the discussion to specific remarks on Turing (e.g. Trächtler, 2021, pp. 99–111). Since the focus of this thesis is Wittgenstein's philosophy of mathematics, such a perspective will not be discussed here in any detail (although some of the later sections in Chapter 3 border on issues normally classified as philosophy of mind).

Second, Turing's notion of computation, his concept of Turing machines and the Church-Turing Thesis can be investigated from the perspective of Wittgenstein's philosophy of mathematics in general and his remark on Turing's diagonal argument in particular (which is also what this thesis attempts to do). Publications on these topics in the Wittgenstein community have been rare, the first detailed discussions of Wittgenstein's remark on Turing's diagonal argument appear to be Shanker, 1987 and Shanker, 1998. The other notable exception is a series of papers by Juliet Floyd: Most directly applicable to Turing's diagonal argument are Floyd, 2012 and Floyd, 2019, other publications in this area include Floyd, 2016, Floyd, 2017, Floyd, 2018 and Floyd, 2020. These papers make a compelling argument for a deeper connection between the two thinkers than one might expect based on the rare textual references, Floyd's interpretation is largely compatible with the discussion of Turing in this thesis (though the following chapters explore a slightly different direction, primarily focused on the role of consistency in Turing's diagonal argument).

Third, the 'debate' between Wittgenstein and Turing on the importance of consistency in formal systems in the *LFM* has led to some discussion that is focused more on consistency and less on issues specific to Turing machines and computability, e.g. Wrigley, 1980, Matthíasson, 2021 and Persichetti, 2021. While the exchange between Turing and Wittgenstein will be discussed briefly (see Section 3.3), the views expressed by Wittgenstein in the *LFM* echo more general remarks on contradictions in the *Nachlass* and will be examined in this larger context.

1

CANTOR, NUMBERS AND ENUMERABILITY

> Unser Verdacht sollte immer rege sein, wenn ein Beweis mehr beweist, als seine Mittel ihm erlauben. Man könnte so etwas einen 'prahlerischen Beweis' nennen. [Ms-117, 109.3 / *BGM II* §21]

> Our suspicion ought always to be aroused when a proof proves more than its means allow it. Something of this sort might be called 'a puffed-up proof'. [*RFM II* §21]

At first glance, one of the most puzzling aspects of Wittgenstein's remarks on mathematics is certainly their tendency to seemingly criticise and even contradict pieces of mathematics that are widely considered beyond reproach. The remark quoted above, part of a series of remarks on Georg Cantor's diagonal method, is an illustrative example of the difficulties posed by Wittgenstein's philosophy of mathematics. How could a proof be said to prove more than its means allow?

The most natural interpretation might be that such a proof must be mathematically *wrong* and that Wittgenstein's aim is to demonstrate a fault in Cantor's diagonal method. Wittgenstein would then put forward *arguments* against a flawed piece of mathematics and his philosophy could consequently be judged on its mathematical merits. But while Wittgenstein was undoubtedly a brilliant philosopher, his mathematical erudition was certainly lacking compared to mathematicians of his time, which might call into question the significance of his writings on the philosophy of mathematics, at least in comparison to what are traditionally considered to be his 'main works'.[1] An interpretation

1 Georg Kreisel, a contemporary and friend of Wittgenstein, is an often cited example of this viewpoint: "Wittgenstein's significant contributions [...] concern very elementary computations" (Kreisel, 1958, p. 135), but his comments on 'higher mathematics' "are, for the most part, uninformed" (Kreisel, 1958, p. 136). Kreisel continues: "Wittgenstein's views on mathematical logic are not worth much because he knew very little and what he knew was confined to the Frege-Russell line of goods" (Kreisel, 1958, pp. 143–44). The review ends scathingly: "I did not enjoy reading the present book. Of course I do not know what I should have thought of it fifteen years ago; now it seems to me to be a surprisingly insignificant product of a sparkling mind" (Kreisel, 1958, p. 158). Quite tellingly, Kreisel holds the view that the characteristic traits of Wittgenstein's more popular writings hold no significance in the context of his writings on mathematics: "What has been described so far is not at all like the popular impression of 'Wittgenstein's philosophy' such as his *anti-metaphysics*, his panaceas of *rule of language* and *application*, his attitude to traditional schools of philosophy. Certainly, the book is not free from these traits of Wittgenstein's writings: but I do not believe they are of any significance, and, in point of fact, when he embarks on serious analysis [...], there is no trace of them" (Kreisel, 1958, pp. 136–37). One of the aims of this chapter is to emphasise that Wittgenstein's philosophy of mathematics appears insignificant only if it is read completely detached from his

along these lines, of Wittgenstein's mathematical writings as *criticising* concrete pieces of mathematics *on a mathematical level*, does pose a number of challenges, however: Not only would such a reading fail to engage with Wittgenstein's thoughts and instead reduce his philosophical discussion of these issues to a mere misunderstanding, such an interpretation would more importantly be incompatible with Wittgenstein's goal of not interfering in mathematical matters (*LFM I*, p. 13, Ts-227a, 89.2 / *PI* §124, Ms-124, 82.2–82.3 / *RFM VII* §19). The purpose of this chapter is to present an interpretation which does not exhibit these problems, but instead takes Wittgenstein at his word and considers the contexts that motivate his remarks, with the ultimate aim of showing the importance of Wittgenstein's reflections on Cantor's diagonal method for a philosophical understanding of its applications in foundational issues in mathematics and logic.

Before discussing Wittgenstein's remarks in detail, it might be helpful to sketch out in broad strokes the topic of this chapter and how it relates to some of the more general aspects of Wittgenstein's philosophy. Part II of Wittgenstein's *Remarks on the Foundations of Mathematics*, a collection of remarks that were heavily edited and drawn from a range of different documents in Wittgenstein's *Nachlass*, revolves around Cantor's diagonal argument and can at first glance appear to be concerned only with a very specific mathematical proof, namely of the uncountability of the real numbers. Briefly, countability in this context can be roughly and informally explained as follows: A set of objects is said to be countable (or "countably infinite") if it is possible to form a 1:1 correspondence with the set of the natural numbers so that each natural number maps to a single object, with no object left out in this mapping. For example, the set of integers (which includes natural numbers and their negative counterparts) is countable, because we can map the natural numbers to the integers in an alternating fashion, so that odd natural numbers get mapped to positive integers and even natural numbers to negative integers (0 to 0; 1 to 1; 2 to -1; 3 to 2; 4 to -2; etc.). Even though the set of integers is in some sense 'larger' than the set of natural numbers (because each natural number except 0 corresponds to two integers, a positive number and its negative counterpart), the two sets can be brought into the desired 1:1 correspondence. The same is true even for the set of rational numbers, a result that can appear counter-intuitive at first, because there are infinitely many rational numbers already between the first two natural numbers ($\frac{1}{2}$, $\frac{1}{3}$, $\frac{2}{3}$, $\frac{1}{4}$, $\frac{2}{4}$, $\frac{3}{4}$, etc.) and thus seemingly 'not enough' natural numbers for a 1:1 correspondence. But if we arrange all positive rational numbers two-dimensionally, with the numerator in one dimension and the denominator in the other, we

'general' philosophy, as Kreisel does, a view which is admittedly facilitated by the problematic editorial policy of Wittgenstein's literary executors and their decision to present his writings on mathematics as a sustained and isolated treatment of particular topics.

can easily bring all the positive rational numbers into a 1:1 correspondence with the natural numbers by starting in the 'corner' of the two-dimensional table and proceeding outward from there, skipping all the fractions that can be simplified (as shown in Table 1, with $\frac{2}{2}$ skipped). It is then trivial to extend this scheme from the positive rational numbers to all rational numbers, by including their negative counterparts in an alternating fashion, exactly as in the case of the integers mentioned above. The rational numbers are thus countable, they can be ordered in such a way that they are enumerable by the natural numbers.

	1	2	3	4	...
1	$\frac{1}{1}$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{4}$	...
2	$\frac{2}{1}$	$(\frac{2}{2})$	$\frac{2}{3}$	...	...
3	$\frac{3}{1}$	$\frac{3}{2}$	...	...	...
4	$\frac{4}{1}$	...	...	...	...
...	...	...	...	...	...

Table 1: Enumerating the Positive Rational Numbers

The same does not hold for the real numbers, however, as Cantor proved with his diagonal argument. The details of diagonalisation will be discussed later in this chapter, but the overall idea is simple enough: Imagine that we are faced with a list of 1000 numbers, each 1000 decimal places long. How can we construct a new number that is guaranteed to be different from all the numbers in the list? If we ensure that the newly constructed number differs from the first number in the first decimal place, from the second number in the second decimal place and so forth, the number thus constructed must differ from all the numbers in the list. If we write out the 1000 numbers vertically, with their decimal places arranged horizontally, we only need to look at the diagonal to construct a new number that differs from all the numbers in the list by adding or subtracting 1 from the decimal place on the diagonal (Table 2).

№1	$\boxed{0 \to 1}$	0	0	0	0	...
№2	4	$\boxed{1 \to 2}$	4	2	1	...
№3	7	3	$\boxed{2 \to 3}$	0	5	...
№4	0	0	0	$\boxed{0 \to 1}$	0	...
№5	2	3	6	0	$\boxed{6 \to 7}$	...
...	...	...	...	...	...	...

Table 2: Constructing a Number by Adding 1 to the Diagonal

Cantor's diagonal argument is a generalisation of this idea to the infinite set of real numbers: Let us focus on the real numbers between 0 and 1 and assume that there is a way to order all these numbers and write them out vertically one after another, each in decimal notation with their decimal expansion stretching out horizontally to infinity, so that this enumeration of real numbers would form a two-dimensional table of decimal places infinitely stretching out to the bottom and to the right. We can now construct a real number that is not contained in this infinite enumeration, by constructing a number that differs from each number in the table *in the diagonal*, in other words with the number differing from the first number in the first decimal place, from the second number in the second decimal place and so forth. Since the number differs from any number in the table in at least one decimal place, it cannot be contained in the table, contrary to the assumption that there is a way to give a countable ordering of all the real numbers. Faced with any such ordering of the real numbers, we can always construct a diagonalised number that will 'escape' from this ordering, the real numbers are therefore *uncountable*.

But why is Wittgenstein interested in this seemingly rather specialised piece of mathematics, so much in fact that he devotes the whole of part II of the "Remarks on the Foundations of Mathematics" (compiled by the literary executors from documents spanning the years 1937 to 1939) to this topic? Most of Wittgenstein's other writings on mathematics are much broader in scope and usually deal with concepts such as 'proof' and 'calculation' on a more abstract level, with concrete examples that are drawn almost exclusively from elementary mathematics and logic. Along with part II, the other notable exception is Wittgenstein's preoccupation with Gödel's incompleteness theorem (Part I, Appendix III in the *RFM*), another application of the diagonal method with close connections to Cantor's original diagonal argument. Are these two collections of remarks merely a misguided attempt to engage with concrete pieces of mathematics, when Wittgenstein's strengths really lie in his treatment of more abstract concepts? The editorial decisions of the literary executors, which present *RFM II* as a distinct treatment of a single topic with little connections to other remarks, certainly give rise to such an interpretation. Even then it is still possible to unearth interesting aspects in these writings, but they will inevitably appear to be far less important than most of Wittgenstein's remaining work.

In contrast to such a view, *RFM II* becomes much more illuminating once it is read against the backdrop of Wittgenstein's other writings on mathematics. To put it a bit crudely, Wittgenstein is not especially interested in the uncountability of the real numbers itself, but rather in the use (and abuse) of the sort of diagonal arguments that originate with Cantor and for which the proof of the uncountability of the real numbers is merely symptomatic. Seen in this light, *RFM II* is only the first (albeit the most extensive) examination of a number of more abstract issues with a multitude of connections to other aspect of Wittgenstein's writings, an examination that has as much to do with Gödel (Chapter 2) and with Turing (Chapter 3) as it has to do with Cantor. This is hard to see when Cantor's use of the diagonal argument is considered on its own, because it is much more purely 'intra-mathematical' than the other examples of diagonal arguments with their applications in logic and computability.

This viewpoint will be further supported by the rest of the chapter. For now, two introductory examples of the connections with other parts of Wittgenstein's writings must suffice:

1. While the rudimentary exposition of diagonalisation presented above proceeded from the finite (1000 numbers with 1000 decimal places) to the infinite case (the assumption of a countable ordering of the infinitely many real numbers, with infinitely many decimal places) without giving much thought to the question of whether this transition from the finite to the infinite is a innocuous as it seems, this is precisely the step that Wittgensteins wants to investigate. The preoccupation with the transition from a finite collection of concrete examples to the infinite case (with a *rule* that holds for *all* elements) is a theme that plays an important role in Wittgenstein's thought, reaching far beyond the confines of set theory and the real numbers. Accordingly, his interest in Cantor's diagonal argument should be read in the context of these larger issues.

2. It is important to note that the (countable) rational numbers are distinguished from the (uncountable) real numbers by the fact that the rational numbers form an explicit system (of fractional numbers, each with a numerator and a denominator), whereas the real numbers are much more heterogeneous, forming a *multitude of systems*[2], a "system of systems", as Wittgenstein calls it in one of the unpublished remarks belonging to the *RFM II* corpus. That the rational numbers all share the same structure, while the real numbers are a *family* of different structures (namely the rational numbers, numbers such as π, numbers defined through diagonalisation, etc.), is of extreme importance to Wittgenstein and very easy to overlook in Cantor's diagonal argument, which has the tendency to treat all numbers equally under the guise of extensionality. This aspect, of definitions of systems in different formalisms and languages and of questions of 'higher-order systems', links *RFM II* with Wittgenstein's remarks in the *PI* on family resemblances, "orthography" and

2 One might argue that the real numbers form a single clearly defined system, as they can all be constructed extensionally with the help of the Dedekind cut. But it is precisely this idea, that such an extensional definition works just as well as an intensional definition to define a system, which Wittgenstein wants to investigate.

"second-order philosophy" (*PI §121*), as well as his concept of a "surveyable representation" (*PI §122*).

Despite these links to other, less mathematical remarks in the *Nachlass*, the temptation to read Wittgenstein as a systematic philosopher (and consequently his writings on Cantor's diagonal argument only as a concrete example of a more general and abstract philosophical argument) must be resisted if one wants to do justice to Wittgenstein's conception of philosophy as a therapeutic activity, where concepts are usually considered against the backdrop of their use in language games. Presenting and discussing links to other parts of the *Nachlass* should not be understood as an attempt to abstract away all the 'messy' mathematical details of Cantor's diagonal argument. On the contrary, Wittgenstein's remarks are often subtle enough to require a discussion of what might at first appear to be mere details. Connecting the more mathematical investigation of Cantor's diagonal argument in *RFM II* with other less purely mathematical uses of the diagonal method can help to illuminate the context in which these concepts are used. This can then be a first step towards a "surveyable representation" in the sense of Wittgenstein, which *describes* the concepts that are involved in the different diagonal arguments instead of trying to *explain* them as part of a unified and systematic theory.

The approach of this chapter will therefore be primarily exegetical and focus on the context of Wittgenstein's remarks in the *Nachlass* documents Ms-117 (probably written between the beginning of October 1937 and the end of June 1938), Ms-121 (end of April 1938 until beginning of January 1939) and Ms-162a/Ms-162b (starting in January 1939), with the former two being the basis for *RFM II*. The first 22 sections of *RFM II* correspond almost exactly to a part of Ms-117 that is clearly delineated from the non-mathematical parts before and after it. The editorial work by the literary executors is more problematic for the remarks *RFM II* §§23–62, however, which are all selected from Ms-121, but omit many remarks which are related to Wittgenstein's discussion of the diagonal method.[3] The focus of the following discussion will thus be mostly on these later remarks in *RFM II* and subsequently consider remarks from Ms-162a and Ms-162b, which were not published in *RFM* and form Wittgenstein's last extensive discussion of Cantor and the diagonal method. Apart from these two pocket notebooks (of which only a selection of remarks will be discussed) and a larger passage on provability and Gödel in Ms-121, the present chapter aims to present a comprehensive discussion of both the published and unpublished remarks related to *RFM II*, sometimes trading in a detailed discussion of particular remarks for a more holistic perspective of Wittgenstein's aim in the remarks on Cantor.

3 For more details on the parts of Ms-117 and Ms-121, see Joachim Schulte's "Text genetic-philosophical note" in the wittgensteinsource.org metadata for Ms-117 and Ms-121.

The outline of the present chapter follows the more or less clearly delineated parts of the aforementioned documents: §§1–22 (Ms-117) of *RFM II* introduce Cantor's diagonal argument and will be discussed relatively quickly (Section 1.1). The next 5 sections of this chapter all discuss remarks from Ms-121, with the first of these focusing mostly on unpublished remarks that connect Wittgenstein's discussion of the diagonal method with remarks on higher-order systems (Section 1.2, *RFM II* §§23–34), the overarching theme of this chapter. The next section examines remarks from Ms-121 that investigate various concepts of numbers (Section 1.3, *RFM II* §§35–39), before looking at unpublished remarks on the notion of *surveyability* (Section 1.4), which could be read as an excursion from the remarks on Cantor, but will prove to be relevant for the larger picture of Wittgenstein's discussion of the diagonal argument. The next section proceeds relatively quickly and deals with most of the remaining remarks of *RFM II* (Section 1.5, *RFM II* §§40–57). The following section acts as a preliminary conclusion of the published remarks (Section 1.6, *RFM II* §§58–62) and connects the previous remarks on surveyability with Wittgenstein's critique of the misleading aspects of Cantor's concept of infinity. The last section contains a discussion of selected remarks in Ms-162a and the beginning of Ms-162b (Section 1.7), focusing on enumerability in formal systems, which brings the chapter back to the notion of a "system of systems" and the surveyability of these higher-order systems.

1.1 CANTOR'S DIAGONAL ARGUMENT

Wittgenstein's introduction of the diagonal method in Ms-117, 97.3 / §1 can seem somewhat unorthodox, as he does not consider the diagonalisation of all real numbers (to then show the uncountability of the real numbers), but merely the series of all square roots. The diagonalisation proceeds in the usual way: If we imagine the square roots arranged vertically in an infinite series, with their decimal expansions extending infinitely horizontally, we can construct a number that differs from the first square root in the first decimal place (by adding or subtracting 1 from the first decimal place), from the second square root in the second decimal place (by adding or subtracting 1 from the second decimal place) and so on. This newly constructed number will then differ from every square root in the series (at least in the decimal place on the 'diagonal' of each square root) and we have therefore proved that there must be a real number that is not a square root.

This result seems rather uninteresting, for two reasons: First of all, there are numerous other examples of real numbers that are not square roots, Wittgenstein himself gives "$\sqrt[3]{2}$" as an example, the diagonal method is thus not needed and arguably a rather complex tool

	$\sqrt{x_1}$	$\sqrt{x_2}$	$\sqrt{x_3}$	$\sqrt{x_4}$	$\sqrt{x_5}$	...
$\sqrt{1} = 1.$	$\boxed{0}$	0	0	0	0	...
$\sqrt{2} = 1.$	4	$\boxed{1}$	4	2	1	...
$\sqrt{3} = 1.$	7	3	$\boxed{2}$	0	5	...
$\sqrt{4} = 2.$	0	0	0	$\boxed{0}$	0	...
$\sqrt{5} = 2.$	2	3	6	0	$\boxed{6}$	...
...	...	...	...	...	...	...

Table 3: Square Roots and the Diagonal Method

for the job. Second, the interest in the diagonal method stems from showing the uncountability of the *real* numbers by starting from the assumption that the real numbers are countable (and can therefore be arranged in an infinite vertical series, with their decimal places extending horizontally) and showing that the diagonalisation can always be used to give a rule for constructing a number that is different from all numbers in the vertical at least in the decimal place on the diagonal, which means that this number *is not part of the countable series of real numbers* but *is a real number*, proving that the assumption leads to a contradiction and the real numbers therefore cannot be countable. Cantor's diagonal argument thus shows that there is always a real number that escapes from the (assumed to be countable) *real* numbers, a number that escapes *its own kind*, so to speak, whereas Wittgenstein's example merely shows that there is always a real number that escapes the square roots, a fact that was never in doubt.[4]

Wittgenstein's choice may be mathematically trivial[5], but is philosophically quite interesting, as it immediately sets the tone for the following investigation. The choice of the square roots as the introductory example is important exactly *because* the use of the diagonal method is impractical in this context. Similarly, in 1939, Wittgenstein begins his reflections on the diagonal method in Ms-162a, 20.2–21.1 by asking for the "ordinary" and "practical purpose" of diagonalisation, before considering only finite cases, at least at first. Ms-162a will be discussed in more detail below (see Section 1.7), but it is important to note that the question of 'practicality' lies at the root of Wittgenstein's remarks. In contrast to the case of the real numbers, where the diagonalised construction is the only obvious choice to 'escape' the

4 While Cantor's second proof of the uncountability of the real numbers proceeds non-constructively from the assumption that the real numbers are countable, it is also possible to proceed similarly but in a constructive fashion to show that for any ordering of the (countable) algebraic numbers, the diagonalised number cannot be part of this ordering and must thus be transcendental, as explained in Mühlhölzer, 2020, pp. 131–33.

5 Even "mathematically ridiculous" (Mühlhölzer, 2020, p. 134).

countable set and could, as a number, be considered 'as good as it gets', the example of the square roots pits a 'practical' number such as "$\sqrt[3]{2}$" (which is frequently used in practical calculations) against the diagonalised construction, which is a number constructed *only for the purpose of diagonalisation* and which does not have the same practical uses.

In Ms-117, 98.2–99.1 / §§2–3, Wittgenstein then goes on to discuss whether the diagonalised number can be considered a satisfying answer to a task of the form "Show me a number that is different from all of these numbers". An even simpler case that requires no diagonalisation is the task "Name a number that agrees with $\sqrt{2}$ in every other decimal place!"[6]: It can be answered either by a number which is used in practical calculations (similar to $\sqrt[3]{2}$) but just happens to match $\sqrt{2}$ in every other decimal place, or it can be answered by: "It is the number obtained according to the rule: develop $\sqrt{2}$ & add 1 or -1 to every second decimal place". The latter answer can be unsatisfying, because what the person formulating the task had in mind were *practical* numbers like $\sqrt[3]{2}$. The alternative seems rather artificial and we could say that it follows the letter but not the spirit of the task. Similarly, the diagonalisation could be met with the reaction "But that's not what I meant!", because the diagonalised number, while being a real number, is not a number such as $\sqrt[3]{2}$, which we might have had in mind as the 'kind' of number that we expected when we first formulated the task. We can make this more concrete by pointing out an aspect of diagonalisation not explicitly mentioned by Wittgenstein in his remark: The diagonalised number depends on the base of the numbering system used, as it depends on the ability to explicitly modify a particular decimal place by addition or subtraction. The diagonalised number in a decimal system will thus be different from the one in a binary system. This stands in stark contrast to our understanding of numbers such as π, for which the concrete base chosen for their expansion is merely a 'superficial' detail that does not detract from their 'essence' as numbers.[7]

This is also the reason why in the case of numbers constructed through diagonalisation, the *method* of calculating the number and the

6 As pointed out in Mühlhölzer, 2020, p. 134, the task as stated by Wittgenstein does not make much sense in the context of the diagonal method, since the purpose of the diagonalisation is to produce a number that *differs* from another number, not one that *matches* it. The task should thus be read either as Mühlhölzer proposes, "Name a number that does not agree with $\sqrt{2}$ precisely at every second decimal place", or could alternatively also be read as "Name a number that agrees with $\sqrt{2}$ *only* at every other decimal place", which is how Redecker, 2006, p. 33 interprets the task. Mühlhölzer interprets Wittgenstein's "1 or -1" as meaning that -1 is to be applied if the decimal place is a 9, which makes sense, but Wittgenstein's "or" could alternatively be read to suggest that there are many different ways of constructing such a number, either by adding 1 (modulo 10, thus wrapping 9 to 0) or subtracting 1 (and wrapping 0 to 9). In any case, these are minor points and Wittgenstein's intent remains sufficiently clear.

7 Cf. Mühlhölzer, 2020, p. 134.

result, the calculated number, could be said to be one and the same, because the method that calculates the decimal places and the resulting decimal expansion are 'as good as it gets', there are no other interesting mathematical properties (such as for $\sqrt{2}$) that could lead us to view a decimal expansion as 'merely' a surface expression (using a particular base) of a deeper mathematical form.[8] For diagonalised numbers the 'inessential' decimal expansion is all there is. Nevertheless, we can certainly use the numbers constructed in this way in our calculations and find out whether such a number is less than another number, for example, by comparing the decimal expansion of a diagonalised number with the decimal expansion of, say, $\sqrt{2}$. But in contrast to comparing $\sqrt{2}$ with $\sqrt{3}$, two numbers that use the same method of calculation, comparing such a square root with a diagonalised number comes down to comparing different methods of calculating numbers, as Wittgenstein notes in Ms-117, 99.3–100.2 / §§4–5. As a result, we enter murky conceptual waters: We might be lead to believe that the method of comparison must be clear, since what we want to compare are 'straightforward' numbers with decimal expansions, but simply viewing them as calculated results obscures the fact that these numbers are of a very different kind, as they are calculated using very different methods.

The variety of methods and ways of comparing numbers is easily ignored or dismissed as unimportant in the wake of the diagonal method, since diagonalisation shows a way in which the diagonalised number and the numbers in the infinite series that the diagonal is based upon are *similar*, at the cost of glossing over what makes them *different*. This situation is aggravated by speaking of the results of the diagonal method in ordinary language without keeping in mind how this result was *calculated* (Ms-117, 100.4 / §7). We then call the diagonalised number a "real number" and the real numbers as a whole "uncountable", but it is easy to overlook what made us call the diagonalised number a real number and how the impossibility of counting these numbers originates from the diagonalisation. Without the calculation, a statement such as "the real numbers are uncountable" can conjure up the image of the *discovery* of something 'greater' than the infinite, which fills us with awe.[9]

8 The equivalence of method and result is also echoed in Ts-222, 61.3 / *RFM* §8: "(I once wrote: "In mathematics process and result are equivalent.")", which the editors of the *RFM* trace back to *TLP* 6.1261. There, Wittgenstein speaks only of logic, but an earlier version of the remark in Ms-102, 75r.3 reads "In der Logik (Mathematik) sind Prozeß & Resultat gleichwertig. (Darum keine Überraschungen.)". A discussion of the relation between logic and mathematics in the Tractarian period would go beyond the scope of this thesis, but it should be pointed out that there is another relevant remark from the middle period (Ms-113, 84r.3), which reads "(Im Kalkül sind Prozeß & Resultat einander äquivalent.)".

9 Redecker, 2006, pp. 31–50, holds the view that Wittgenstein's aim is to put forward a *thesis*, namely that the number constructed via the diagonal method is *not* a real number. Ramharter, 2010, p. 298 explicitly avoids stating a thesis, but points out that

The danger of being mislead by the result of the diagonal method in the absence of the calculation itself is exhibited in §8, which is concise enough to be quoted in full:

> 'Ich will Dich eine Methode lehren wie Du in einer Entwicklung allen diesen Entwicklungen nach der Reihe *ausweichen* kannst.' So eine Methode ist das Diagonalverfahren. – "Also erzeugt sie eine Reihe, die von allen diesen verschieden ist." Ist das richtig? – Ja; wenn Du nämlich diese Worte auf diesen, oben beschriebenen Fall anwenden willst. [Ms-117, 101.2 / *BGM II* §8]

> "I want to shew you a method by which you can serially *avoid* all these developments." The diagonal procedure is such a method. – "So it produces a series that is different from all of these." Is that right? – Yes; if, that is, you want to apply these words to the described case.

That the diagonal method "dodges" all the decimal expansions it is being applied to is a very apt description of what happens on the diagonal. To say that it "therefore produces" a series that is "different from all of them" is an example of the misleading consequence of considering the diagonalised expansion in the absence of the diagonal method. Yes, we *may* want to use these words (and usually we will), but only *if we choose to apply these words* to such a case. The idea that it *must* follow (instead of simply saying that it follows) is what Wittgenstein is concerned with here (which becomes very clear in Ms-117, 102.2 / §10, where he explicitly emphasises the "Therefore"). Our notion that it *must* follow stems from the finite case, where it is easy to see that the diagonal method can 'wait out' the (finite) list of decimal expansions and *then* construct a new expansion that dodges all of them. It may be difficult for us to see how and why we would not choose the exact same words to describe even the infinite case, but Wittgenstein gives an example in Ms-117, 101.3 / §9 and Ms-117, 103.2 / §11 of someone not agreeing that the words used in the finite case are applicable to the infinite case. This person would frame the situation as follows: Faced with an infinite list of decimal expansions, the diagonal method could be used to construct a new decimal expansion that dodges every expansion in the *square*, but there will always be other decimal expansions occurring *later* in the infinite list (the lower part of the rectangle in Ms-117, 103.2 / §11 that is not part of the square with the diagonal) that the diagonal could not yet dodge. We can thus not be sure that the diagonal really is different even from

---

the objection by Wittgenstein in §§5–7 would, in the face of a more rigorous formulation of the diagonal argument, "make Analysis impossible altogether", a view echoed in Ramharter, 2018, pp. 130–31. The following interpretation stands in stark contrast to such a view and attempts to show that a more natural reading results if Wittgenstein's remarks are interpreted as showing the *conceptual variety* of numbers that is at play in the diagonal method, not an attempt to disagree with mathematicians on what constitutes a real number. Any reading that interprets Wittgenstein as criticising the diagonal method on mathematical grounds must deal with his claims of not putting forward theses (Ts-227a, 89.6 / *PI* §128) and of not interfering with mathematicians (Ts-227a, 89.2 / *PI* §124, but also Ms-124, 82.2–82.3 / *RFM VII* §19).

these later decimal expansions. Wittgenstein is of course aware that such a view would be a misunderstanding of the application of the diagonal method, which is why he immediately adds what we would reply to such a person:

> "Aber Deine Regel reicht doch schon in's Unendliche, also weißt Du doch schon genau daß die Diagonal-Reihe von jeder andern verschieden {sein wird // ist}!" [Ms-117, 101.3 / *BGM II* §9]

> "But your rule already reaches to infinity, so you already know quite precisely that the diagonal series will be different from any other!" – [*RFM II* §9]

This is exactly the reason why we say that the diagonal series *is* different from all the other series, but it is superfluous and misleading to say that it *must* follow or that it "therefore" follows from the diagonal argument, because there is nothing that we could appeal to in an attempt to convince such a person of our view. This shows a striking parallel to an earlier remark from Wittgenstein in Ms-117:

> "Aus 'alle', wenn es so gemeint ist, muß doch *das* folgen." – Wenn es *wie* gemeint ist? Überlege es Dir, wie meinst Du es? Da schwebt Dir etwa noch ein Bild vor – und mehr hast Du nicht. – Nein, es *muß* nicht – aber es *folgt*: Wir *vollziehen* diesen Übergang.
>
> Und wir sagen: Wenn das nicht folgt, dann waren es eben nicht *alle*! – – und das zeigt nur, wie wir mit Worten in so einer Situation reagieren. – [Ms-117, 1.1 (p. 13–14), Ts-221a/b, 147.2, Ts-222, 15.1 / *BGM I* §12]

> "From '*all*', if it is meant *like this*', *this* must surely follow!" – If it is meant like *what*? Consider how you mean it. Here perhaps a further picture comes to your mind – and that is all you have got. – No, it is not true that it *must* – but it *does* follow: we *perform* this transition.
>
> And we say: If this does not follow, then it simply wouldn't be *all* – and that only shews how we react with words in such a situation. [*RFM I* §12]

Either the other person plays our language game or they do not, but in the latter case we have to convince them through other means and cannot point to the "therefore" as if this would somehow make the argument more convincing.[10] This also explains the addition of the "ad inf." to the modified diagram in Ms-117, 103.2 / §11, compared to the diagram in §1: While for us this addition might seem superfluous and the two diagrams interchangeable, the person rejecting the applicability of the diagonal method for producing a series that is *different* from all the others would consider this addition far from superfluous and point to it as demonstrating that the finite case (which is captured by the first diagram) is not at all like the infinite case (which is captured by the second diagram). For them, the "ad inf." is exactly

10 Mühlhölzer, 2020, p. 163 mentions *Z* §134 as another illuminating remark in this context, which begins "Do not say "one cannot", but say instead: "it doesn't exist in this game".", which apart from Ts-233a, 28.2 also appears in Ms-116, 178.4, Ms-158, 5r.1 and Ts-228, 37.4 (the latter two are mistakenly cited as "MS 158, p. 57" and "TS 222").

the reason why the diagonal can never catch up with the lower part of the rectangle.

Of course we must be careful not to ascribe the view of such an imagined interlocutor to Wittgenstein himself. That he is in no way advocating for a dismissal of Cantor's diagonal argument is made clear in Ms-117, 102.2 / §10:

> Es heißt nichts zu sagen: "*Also* sind die X-Zahlen nicht abzählbar". Man könnte etwa sagen: Den Zahlbegriff X nenne ich unabzählbar, wenn festgesetzt ist, daß, welche der unter ihn fallenden Zahlen immer Du in eine Reihe bringst die Diagonalzahl dieser Reihe auch unter ihn {fällt. // fallen solle.} [Ms-117, 102.2 / *BGM II* §10]
>
> ---
>
> It means nothing to say: "*Therefore* the X numbers are not denumerable". One might say something like this: I call number-concept X non-denumerable if it has been stipulated that, whatever numbers falling under this concept you arrange in a series, the diagonal number of this series is also to fall under that concept. [*RFM II* §10]

The issue is to say "*Therefore* the X-numbers are uncountable", because the "Therefore" cuts off the conclusion from the diagonal method and makes it seem as if *by using the diagonal method as a tool* we had reached a discovery that can now be used independently from the diagonal proof itself. What Wittgenstein seems to be driving at is not that calling these numbers "uncountable" were somehow *wrong*, but that we *call* these numbers "uncountable" *if* "it has been stipulated that, whatever numbers falling under this concept you arrange in a series, the diagonal number of this series is also to fall under that concept", meaning that the concept of something being "uncountable" derives its meaning from the diagonal method and does not have a sense in the absence of it. Put more plainly, Wittgenstein wants to point out that being uncountable is a *conceptual stipulation*, an *invention*, not a *discovery* in the realm of the infinite.

To show that a concept exhibits a *logical* instead of a *physical* impossibility, Wittgenstein often asks how such a concept is *used*, which is exactly how he proceeds in Ms-117, 104.2–104.3 / §§12–13. To prove via the diagonal argument that it is impossible to count the irrational numbers[11] *could* convince a person looking for such an ordering to give up their attempt, which appears to be the only use of the diagonal argument. But such a use is very vague, we can only picture the person somehow working "idiotically" on the infinite series, but we have no clear picture *which* methods of calculation cannot be ordered in the same way as the natural numbers.[12] Of course we know

11 Mühlhölzer, 2020, p. 161 notes that a proof of the uncountability of the irrational numbers with the help of the diagonal method is not as straightforward as it seems and that, in contrast to the real numbers, the uncountability of the irrational numbers cannot be demonstrated *purely* by means of the diagonal method. This seems to have been an oversight on the part of Wittgenstein, but it does not negatively impact his line of reasoning, as we could usually read "real numbers" where he writes "irrational numbers" and still follow his philosophical argument.

12 Following Mühlhölzer, 2020, p. 170, the term "natural numbers" will sometimes be used when Wittgenstein is talking about "cardinal numbers", because "[w]hen he

which numbers we call irrational, but that does not mean that we have a clear picture of how an ordering of these methods of calculation could even be attempted, since the irrational numbers are produced through a variety of different methods, not through a single uniform method as in the case of the series of square roots. This is why Wittgenstein draws our attention to the "idiotic" aspect of such a picture, which is crucial for understanding his critique: In the case of the square roots, we can "idiotically" and mechanically[13] list the square roots one by one, because we have a uniform method that produces them, which is why we know what we mean when we talk about ordering them. We are then mislead to believe that because we know how an idiotic attempt to order square roots or natural numbers looks like we also know how such an idiotic attempt would look like for the irrational numbers, even if our picture might be a bit more 'vague'.

But what Wittgenstein wants to show is that this vague picture is not just an imprecise view that is then brought into sharp focus by Cantor's diagonal argument. Instead, Cantor's diagonal argument introduces this new picture in the first place, it does not provide us with a better lens which helps us to *discover* a mathematical fact, it *invents* the concept of (the impossibility of) an ordering of the irrational numbers. This is what is dangerous about Cantor's diagonal argument: Not the mathematical result in and of itself, but its tendency to make the "formation of a concept" look like a "fact of nature" Ms-117, 108.3 / §19. The reaction to the diagonal argument, if understood as a conceptual invention, *can* be to give up an attempt to find an ordering for the real numbers, but not because the result of the argument were a fact that cannot be argued with (like the speed of light in physics), rather because it convinces us to adopt this new concept in our language. But exactly because the notion of attempting to find an ordering of the real numbers is so vague, another picture could perhaps lead to a more suitable concept than Cantor's diagonal argument and cause the person to take up their efforts again (Ms-117, 104.3 / §13), depending on the use the person had in mind for their attempt to order these numbers.[14]

---

uses the term "cardinal number" here [Mühlhölzer is referring to §16], Wittgenstein means, as always, finite cardinal numbers, that is the natural numbers 0, 1, 2, etc."

13 The connection between "idiotic" and mechanical calculations is not made explicit in this remark, but can be seen in the only other remark where Wittgenstein speaks of humans as "calculating machines" which we can picture to be "completely idiotic" (Ms-126, 33.4), but also in remarks where Wittgenstein associates calculating "without thinking" ("gedankenlos" and "ohne nachzudenken") with the notion of mechanical calculation, most notably in Ms-124, 58.2 and in a remark from 1944 that occurs both in Ms-124, 164.5 and Ms-127, 118.2: "'Mechanisch', das heißt: ohne zu denken. Aber *ganz* ohne zu denken? Ohne *nachzudenken*." See Section 3.2 for a discussion of the remark in the context of Turing's diagonal argument.

14 It is unfortunate that Wittgenstein does not give an example of a picture that would lead someone to resume their ordering attempts, because it is quite hard to imagine

This is the big picture, which Wittgenstein tackles in more detail in Ms-117, 104.3–108.3 / §§13–19. In Ms-117, 105.1 / §14, Wittgenstein readily admits that in the case of the irrational numbers the diagonal method can be said to demonstrate that "*these* methods of calculation" cannot be ordered in a series. But we started with a vague picture of what the "these" refers to, because we know of no way to even attempt to *produce* a complete list of irrational numbers, we only know what we will call an irrational number once we have a candidate in front of us.[15] In the case of the rational numbers, this is different, here we have both a criterion that allows us to check whether a number is rational *and a method to produce all rational numbers*. When the diagonal argument *assumes* that an ordering of the irrational numbers exists, it assumes such a method, but what it is that we assume here exactly if we have no concept of such a method for the real numbers? When we say that the natural numbers can be ordered in a series, this conjures up the misleading picture of ordering a finite series, 'just' extended infinitely. But in contrast to the finite case, where we can

---

a suitable picture that is not completely absurd or misunderstands the most basic aspects of Cantor's argument. Mühlhölzer, 2020, p. 166 suggests as an example "the picture that the diagonal number of a given series of numbers may subsequently be put at the beginning of the series and so would now belong to this new series.", but perhaps an equally fitting (though still somewhat absurd) picture would be to put the diagonal number at the *end* of the series, especially in light of §9 and §11, which could convince someone that in this way the diagonalised number will never 'dodge itself' as it has to run through all other numbers in the series 'before' being able to dodge its own diagonal decimal place. A person convinced by this picture could then argue that since there will always be infinitely many other numbers before the diagonalised number, diagonalisation will never lead to a conflict and the diagonalised number can still be considered to be part of the series, simply infinitely far away at the end. Such a picture of putting a number at the 'end' of an infinite list of numbers could seem nonsensical, but it should be pointed out that it must not appear any more nonsensical than Cantor's own picture of ω, the lowest transfinite ordinal number, appearing 'after' the infinite list of finite ordinal numbers. Again, the aim here is not to suggest that Cantor's proof were somehow wrong, but merely to point out that our pictures, for example of the transfinite, might not appear any more convincing or less absurd to a person than the picture of putting the diagonalised number 'at the end' is for us. We are justified in our use of pictures by the mathematical techniques that give them meaning, but we are often mislead to think that these pictures are the only 'natural' or 'reasonable' pictures that anyone could take up, if they only understood the proof properly. The alternative picture is certainly absurd to us, but it could be convincing in certain contexts and might perhaps help to explain Wittgenstein's aim in §9 and §11.

15 At first glance, we might think that a computational procedure could supply exactly the kind of precision that our vague concept of the real numbers seems to be lacking. Although a computational procedure can indeed be used to recursively list all computable numbers and these numbers will thus be enumerable (Turing, 1936, p. 68), they are not enumerable "by finite means" (Turing, 1936, p. 72) and there will always be "(in the ordinary sense) definable numbers" that are not computable, which Turing shows on the basis of his application of the diagonal argument (Turing, 1936, pp. 78–79). As a result, computational procedures cannot sharpen the vague concept of "these" numbers, but only replace it with an entirely new concept that differs considerably from our concept of the real numbers. Additionally, computable numbers raise new philosophical questions, which are considered in more detail in Chapter 3.

explain our concept of ordering a series by pointing to the concrete elements purely extensionally, the ability to order an infinite list of natural numbers depends on a very specific method and the concept of ordering these numbers is only used *analogously* to the concept in the finite case. To ask whether it is possible to order the real numbers thus comes down to asking whether we can do something with the real numbers that we would consider to be *analogous* to the case of the natural numbers, but *before* Cantor's argument we have no concept of what we could call analogous for "'all real numbers'":

> Wenn man also fragt: "Kann man die Reellen Zahlen in eine Reihe ordnen?" So könnte die gewissenhafte Antwort sein: "Ich kann mir vorläufig gar nichts Genaues darunter vorstellen". – "Aber Du kannst doch z.B. die Wurzeln & die algebraischen Zahlen in eine Reihe ordnen; also verstehst Du doch den Ausdruck!" – Richtiger gesagt ich *habe* hier gewisse analoge Gebilde, die ich mit dem gemeinsamen Namen "Reihen" benenne. Aber ich habe noch keine sichere Brücke von diesen Fällen zu dem 'aller reellen Zahlen'. Ich habe auch keine allgemeine Methode um zu versuchen ob sich die oder die Menge 'in eine Reihe ordnen läßt'. [Ms-117, 105.3 / *BGM II* §16]
>
> ---
>
> Asked: "Can the real numbers be ordered in a series?" the conscientious answer might be: "For the time being I can't form any precise idea of that". – "But you can order the roots and the algebraic numbers for example in a series; so you surely understand the expression!" – To put it better, I *have got* certain analogous formations, which I call by the common name 'series'. But so far I haven't any certain bridge from these cases to that of 'all real numbers'. Nor have I any general method of trying whether such-and-such a set 'can be ordered in a series'. [*RFM II* §16]

As a result of Cantor's diagonal argument, we can (and usually will) take up his concept of "uncountable" and use it in our language, but it is important to note that this 'concept formation' does not give us a clear picture of what it was that we assumed in Cantor's argument when we assumed "this" way of ordering ("dieses Ordnen hier") to be possible. Cantor can give us a *new* concept, but he cannot illuminate a *vague* concept that was unclear to us before. Once again, Wittgenstein's aim is to speak of Cantor's diagonal method as an invention instead of a discovery. Cantor does not and cannot discover what it is that is not possible and we can always answer: "I don't know – to repeat – what it is that *can't be done* here" (Ms-117, 105.3 / §16). The task of a philosopher is then neither to invent a new concept nor discover what we 'really' meant (which would amount to a philosophically misguided attempt to search for the essence of a vague concept), but only to show that we had no clear concept of such an ordering to begin with. Interpreted more modestly, Cantor's diagonal method can clarify a difference in our use of the concept of the real numbers in comparison to square roots:

> Die Wurzeln nennen wir "reelle Zahlen" & die Diagonalzahl, die aus den Wurzeln gebildet ist *auch*. Und ähnlich mit allen Reihen reeller Zahlen. Daher hat es keinen Sinn von einer "Reihe *aller* reellen Zahlen" zu reden, weil

> man ja auch die Diagonalzahl {der // jeder} Reihe eine "reelle Zahl" nennt. [Ms-117, 105.3 / *BGM II* §16]

> Such a difference as e.g. this: roots are called "real numbers", *and so too* is the diagonal number formed from the roots. And similarly for all series of real numbers. For this reason it makes no sense to talk about a "series of all real numbers", just because the diagonal number for each series is also called a "real number". [*RFM II* §16]

Seen in such a light, the diagonal method immediately appears much more down to earth, more "homespun" / "hausbacken" (Ts-213, 412r.2, Ms-126, 58.4 / *RFM V* §9). In Ms-117, 107.2–108.2 / §17–18, Wittgenstein imagines that the method could even have been taught to school children long before set theory had been invented, to generate a number that is different from all the numbers in a long but finite list. The diagonal method itself could then have a very practical use in our day-to-day activities.

This leads Wittgenstein to Ms-117, 108.3 / §19 and the danger of mistaking the "formation of a concept" for a "fact of nature", already briefly discussed above. He then proposes a more "modest" phrasing of the result of Cantor's diagonal argument:

> Bescheiden {heißt // lautet} der Satz: "Wenn man etwas eine Reihe reeller Zahlen nennt, so heißt die Entwicklung des Diagonalverfahrens auch eine 'reelle Zahl' {& zwar eine die 'von allen Gliedern der Reihe verschieden' {sei // ist}. // & zwar sagt man, sie sei von allen Gliedern der Reihe verschieden.} [Ms-117, 109.2 / *BGM II* §20]

> The following sentence sounds sober: "If something is called a series of real numbers, then the expansion given by the diagonal procedure is also called a 'real number', and is moreover said to be different from all members of the series". [*RFM II* §20]

Such a wording expresses what we *call* a real number and what we *call* "different from all members of the series" instead of talking about infinite sets *being* 'greater' than other infinite sets and of something as *being* 'uncountable', as if that notion had a sense even in the absence of the diagonal method. This is then what is "boastful" (Ms-117, 109.3 / §21, translated by Anscombe as "puffed-up") in the context of Cantor's diagonal argument: Not the mathematical result in itself but the tendency to view it as a discovery of infinities that are greater than other infinities.

A boastful expression of Cantor's result (over-)emphasises the *similarity* of natural and real numbers, by suggesting that we could *compare* the size of the set of real numbers with the natural numbers, when in fact we have no *uniform method of comparison* that could be applied to both. A modest expression of the mathematical result emphasises exactly the opposite, namely that our concept of real numbers differs in an important aspect from the concept of natural numbers, an aspect which we can easily overlook if we are mislead by the vague picture of an analogy borrowed from the finite case. Wittgenstein ends his remarks in Ms-117 with a beautifully succinct expression of this misleading tendency:

> Wenn gesagt würde: "Die Überlegung über das Diagonalverfahren zeigt Euch, daß der *Begriff* 'reelle Zahl' viel weniger Analogie mit dem Begriff Kardinalzahl hat, als man, durch gewisse Analogien verführt, zu glauben geneigt ist" so hätte das einen guten & ehrlichen Sinn. Es geschieht aber gerade das *Gegenteil*: indem die 'Menge' der reellen Zahlen angeblich der Größe nach mit der der Kardinalzahlen verglichen wird. Die Artverschiedenheit der beiden Konzeptionen wird durch eine schiefe Ausdrucksweise als Verschiedenheit der Ausdehnung dargestellt. Ich glaube & hoffe eine künftige Generation wird über diesen Hokus Pokus lachen. [Ms-117, 109.4 / *BGM II* §22]

> If it were said: "Consideration of the diagonal procedure shews you that the *concept* 'real number' has much less analogy with the concept 'cardinal number' than we, being misled by certain analogies, are inclined to believe", that would have a good and honest sense. But just the *opposite* happens: one pretends to compare the 'set' of real numbers in magnitude with that of cardinal numbers. The difference in kind between the two conceptions is represented by a skew form of expression, as difference of extension. I believe, and hope, that a future generation will laugh at this hocus pocus. [*RFM II* §22]

How is the last statement of this remark to be interpreted, the belief that "a future generation will laugh at this hocus pocus"? It is all too easy to read it as an indictment of set theory as a whole, but this would amount to an interference in mathematical matters. Of course it is possible that in 1937, when the remarks of Ms-117 were written, the aim of non-interference was not yet a central tenet of Wittgenstein's philosophy of mathematics, or that, alternatively, he simply 'slipped' and could not help but insert this dig at mathematical practice. A more charitable interpretation, however, is that what a future generation will laugh at is not set theory or Cantor's diagonal argument, not any piece of mathematics in fact, but only the misleading interpretations of the diagonal method in situations where they easily seep into our day-to-day language and extra-mathematical practice. Seen from the perspective of mathematicians, Wittgenstein's talk of "hocus pocus" must seem absurd or just plain wrong, even today, because set theory is clearly an important part of mathematics.[16] But

16 Such a view is for example expressed in Putnam, 2007, but is echoed in many other discussions of *RFM II* and Wittgenstein's philosophy of mathematics as a whole. As Wheeler, 2021, p. 2 notes: "Wittgenstein's biggest issues with Cantor's diagonal procedure lies in the fact that he (Cantor) attempted to use a concept (nondenumerability) which he took to already have a sense *before* his development of the diagonal procedure" and consequently "Putnam claims that Wittgenstein's issue with Cantor is that he uses a concept that is indeterminate; but he (Putnam) does not recognise the distinction between 'determined before the diagonal proof' and 'determined after (and *by way of*) the diagonal proof.'" As Wheeler points out, the fact that mathematicians *have* found a use for set theory or even that we *could* find a use for it does in no way contradict Wittgenstein's apparently dogmatic rejection of it, because Wittgenstein is concerned with the appearance of sense that is conjured up by Cantor's proof alone, by its misleading picture (its interpretation, not the mathematical proof itself). Wheeler, 2021, p. 8: "This is not to say that these pictures or expressions are ones that are *necessarily* useless, but rather, ones that we *presently* have no use for." This also applies to Wittgeinstein's view on inconsistency (see Chapter 2 and Chap-

Wittgenstein does not seem to be talking to the average mathematician, at least not to one that agrees with the dominant conception of mathematics as practiced today. Instead, he seems to be envisioning a future generation that has lost interest in Cantor's diagonal argument or views it as nothing more than a rather obvious but ultimately pointless exercise. Such a view is much more radical than it might first appear und would require a fundamental change in our way of life. It is unlikely, then, that Wittgenstein was under the illusion of being able to convince his contemporaries or even present-day mathematicians.[17]

1.2 A SYSTEM OF SYSTEMS

In May of 1938, about half a year after the remarks in Ms-117, Wittgenstein returns to Cantor's diagonal argument in Ms-121. The first published remark of that period in *RFM II* serves as a good example of the 'cultural' distance between Wittgenstein and mathematicians making use of the diagonal method:

> Die Krankheit einer Zeit heilt sich durch {eine // die} Veränderung in der Lebensweise der Menschen & die Krankheit der philosophischen Probleme könnte nur durch eine veränderte Denkweise & Lebensweise geheilt werden nicht durch eine Medizin die ein Einzelner erfand.
>
> Denke, daß der Gebrauch des Wagens gewisse Krankheiten hervorruft oder begünstigt & die Menschheit von dieser Krankheit geplagt wird, bis sie sich, aus irgendwelchen Ursachen, als Resultat irgendeiner Entwickelung, das Fahren wieder abgewöhnt. [Ms-121, 27r.4 / *BGM II* §23]
>
> ---
>
> The sickness of a time is cured by an alteration in the mode of life of human beings, and it was possible for the sickness of philosophical problems to get cured only through a changed mode of thought and of life, not through a medicine invented by an individual.
>
> Think of the use of the motor-car producing or encouraging certain sicknesses, and mankind being plagued by such sickness until, from some cause

---

ter 3), because Wittgenstein views undecidable or contradictory propositions only as *presently* useless, not *necessarily* useless (and we might find a use for an inconsistent formal system).

17 Mühlhölzer, 2020, pp. 177–79 presents a more detailed discussion of the "hocus pocus" sentence and interprets it as a "vehement rejection of set theory". But then it was clear even in 1938, following a wager between Weyl and Pólya, that "the "I believe" in his hocus-pocus statement is wrong". Mühlhölzer is surely correct that if the statement is read as a rejection of set theory, the statement has to be considered wrong when judged by present-day mathematical practice, but this commits Wittgenstein to a rather extreme position that is at odds with his quite nuanced focus on the *conceptual* interpretations surrounding Cantor's diagonal argument, not mathematical practice itself. As Mühlhölzer points out, Wittgenstein "de-dramatizes" uncountability and "is looking at mathematical situations in precisely such a sober way, far from all verbal mystifications, concentrated only on the actual mechanisms used in the proofs." If this captures the spirit of the hocus-pocus statement (and it is certainly compatible with Wittgenstein's remarks leading up to that statement) the hocus-pocus statement can and should be read as "primarily emphasizing this sort of distinction", *without rejecting set theory*.

> or other, as the result of some development or other, it abandons the habit of driving. [*RFM II* §23]

This can be seen as a more detailed explanation of what Wittgenstein has in mind when he thinks of a "future generation": It amounts to a fundamental change in their "mode of life", their "Lebensweise", comparable to a whole society giving up their custom of driving cars, with all the related changes in their way of life that it entails.[18] It is clear that such a change cannot be triggered by the (philosophical) "medicine" of a single person. The remark is a sign of Wittgenstein's keen awareness of how little inroads his philosophy would be able to make in mathematics departments and the activities of professional mathematicians.

But while this remark might read like an unbroken continuation of his thoughts from Ms-117, it is far from the only choice with which the editors of *RFM II* could have continued after §22. Apart from several remarks on the use of the word "{infinite // transfinite}" and infinity in connection with cardinal numbers at the beginning of Ms-121 (the 8 remarks from Ms-121, 1r.1–2v.2, which are, however, all marked with the curved "S" expressing Wittgenstein's dissatisfaction with the quality of these remarks), there is also a remark comparing "to want the impossible" with the game of "Daumenfangen" (Ms-121, 26r.2), a short remark on provability (Ms-121, 26v.2), and most notably the following enigmatic remarks that immediately precede the remark selected as §23 and were written three days prior to it:

18 There are obvious similarities between "mode of life" / "Lebensweise" and "form of life" / "Lebensform", raising the question of how these two terms differ. Here and in the following chapters, "Lebensform" will be interpreted not as a technical term on the same level as e.g. "language game", but merely as one of several terms that emphasise related aspects, with "pattern of life" / "Lebensmuster", "stencil of life" / "Lebensschablone" and "mode of life" / "Lebensweise" being other examples. Majetschak, 2010 shows convincingly that Wittgenstein uses the term "Lebensform" sparingly and that, more importantly, most of the occurrences of the term in the *Nachlass* clash with the standard interpretation of "Lebensform" as a cultural embedding, but can be explained by interpreting "forms of life" as the more down-to-earth expression "patterns of life". In the following text, "Lebensform" will thus be interpreted as a non-technical term that can often be used interchangeably with "Lebensmuster" (*contra* Moyal-Sharrock, 2015, see also the further exploration of this topic in Majetschak, 2020). For the sort of embedding that is usually associated with "Lebensform" the term "life" / "Leben" will be used. But in contrast to a fully synonymous reading of "Lebensform" and "Lebensmuster", the different terms will be read here as a continuum, with "Leben" and "Lebensmuster" on opposite ends of the spectrum, the former emphasising the cultural embedding and the latter emphasising a particular constellation of expressions and acts that are part of our life. A term such as "mode of life" / "Lebensweise" is then closer to "Leben" than to "Lebensmuster", whereas "form(s) of life" / "Lebensform(en)" move into the direction of "Lebensmuster", especially if used as a plural. Such a reading has the advantage of being able to explain the rare occurrences of "Lebensform" in works such as *On Certainty*, while still allowing for a use compatible with the standard interpretation (such as in the *Brown Book*). In particular, "Lebensform" can be distinguished from "Lebensformen" by interpreting them as two points (very close together) on a crowded spectrum of related terms.

> Das Vergnügen, das wir an einem aufgeblasenen Gummiballon haben. Wir sind nicht gewöhnt mit Körpern zu hantieren, die so groß im Verhältnis zu ihrem Gewicht sind.
>
> Es hilft wenn man sagt: der Beweis des Fermatschen Satzes ist nicht zu entdecken, sondern zu *erfinden*.
>
> 'Ein "System aller Systeme" ist ein Widerspruch.'
> Wie läßt sich dieser Satz anwenden? [Ms-121, 26v.4–27r.3]
>
> ---
>
> The pleasure we get from an inflated rubber balloon. We are not used to handling bodies that are so big in relation to their weight.
>
> It helps to say: the proof of Fermat's theorem is not to be discovered, but to be *invented*.
>
> 'A "system of all systems" is a contradiction.'
> How can this sentence be applied?

The editors of the *RFM* can hardly be faulted for not including the rather obscure first two remarks. But as we will see in the other remarks on Cantor's diagonal method, Wittgenstein returns frequently to the notion of a "system of systems", which first appears in Ms-121, 27r.3. The above three remarks, though hardly self-explanatory, serve as guiding thread for the rest of his reflections and can help to explain his interest in Cantor's diagonal argument. In contrast to Ms-117, which could be read mostly as an investigation of the concept of "uncountable" numbers, Wittgenstein's interest in Ms-121 is broader and focuses on the use of Cantor's diagonal method as a tool to escape a "system" and talk of a "system of all systems", which intimately connects Wittgenstein's remarks in Ms-121 with his reflections on Russell, Gödel and the idea of provability in a system.

The following interpretation will be supported by a more detailed look at the later remarks in Ms-121, but could be summarised in the light of the above three remarks as follows:

1. Our fascination with Cantor's diagonal method is in need of a careful philosophical investigation. We are fascinated because something huge (the 'greater than infinite') is apparently produced by something so small (the simple method of diagonalisation), similar to an inflated rubber balloon, which is "so large in relation to its weight".

2. A new proof of something that is as of yet undecided does not shed light on an 'unexplored corner' inside the existing system, it instead creates a *new system* that shows some analogies to the old system but is not always directly comparable to it. Cantor's diagonal argument does not shed light on a vague concept of ordering the real numbers, it invents a new concept.

3. In the wake of Cantor's diagonal method, it can be tempting to think that it makes sense to speak of a "system of all systems"

> and show that such a system leads to a contradiction, similar to how the assumption of an (ordered) system of the (variety of) systems of real numbers leads to a contradiction. But to understand what such a statement means, we need to look at its use, which requires us to look at how we speak of a "system of all systems" if we want to understand the diagonal method.

Immediately after Ms-121, 27r.4 / §23, Wittgenstein's reflections revolve around the "greatest cardinal number" and he starts by considering the following task (in a remark not published in *RFM II*):

> "Nenn' mir eine Zahl, die größer ist, als die Zahl aller ganzen Zahlen!" – {diese // Diese} Aufgabe hat den Charakter einer mathematischen Scherzfrage. [Ms-121, 27v.2]
>
> ---
>
> "Name me a number that is greater than the number of all integers!" - this task has the character of a mathematical trick question.

A mathematician could answer such a question quite seriously: $\aleph_0$ is the size of the set of the natural numbers, it could therefore be considered the "number of all natural numbers". The task then is to answer with a cardinal number greater than $\aleph_0$, with $\aleph_1$ (the smallest cardinal number greater than $\aleph_0$) being the obvious choice. Wittgenstein is of course aware of this, his remark explicitly wants to leave this mathematical perspective behind (at least for the duration of the argument) and thus mirrors the "But that's not what I meant!" sentence in Ms-117, 99.1 / §3. Cantor's diagonal method with its cardinal numbers beyond $\aleph_0$ can give us the impression that there *is* something greater than the "number of all natural numbers", that there *must be* a number that fits such a description, not just that there is something that we *call* greater than infinite in a very restricted language game. How we use these cardinal numbers outside this mathematical language game is however largely undetermined: We cannot use the proposition "$\aleph_1$ is greater than $\aleph_0$" in the same way that we can use the sentence "6 is greater than 5", because the "... is greater than ..." belongs to very different language games in these two cases. In the latter case, we can use it in a language game where we compare the number of apples for example, something that only works for finitely many apples. Seen from the perspective of these practical uses of "... is greater than ...", the task mentioned by Wittgenstein is a riddle: It can be answered through the use of cardinal numbers, but the answer will have the appearance of a joke, because it gives a useless answer where the question appeared to ask for a useful one. The analogy between the two different uses of "... is greater than ..." only takes us some distance before the paths eventually diverge. It would be hardly surprising if someone reacted by saying "But that's not what I meant!" when faced with $\aleph_1$ as the answer to the above question, or alternatively accepts it, but only as the answer to a riddle without any practical uses.

The use of such a mathematical answer is also the topic of the next remark, where Wittgenstein considers the task "Give me a number between $\frac{1}{n}$ and $\frac{1}{m}$" and notes that the "usefulness [of this exercise] lies in the fact that there is a *system* of such {tasks // problems} here" (Ms-121, 27v.3). We can imagine that such an exercise is quite useful to introduce students to basic concepts of calculations involving fractions, a student could for example solve such an exercise by giving $\frac{m+1}{m \times n}$ as an answer and thereby demonstrate (or fail to demonstrate) a certain understanding of the simplification and comparison of fractions. Crucially, the usefulness of the exercise depends on the fact that there is a whole *system* of calculations with rational numbers that is used in practice.

Alternatively, we can imagine a student who, after having seen Cantor's diagonal argument, decides to solve the exercise by developing the decimal expansions of $\frac{1}{n}$ and $\frac{1}{m}$ one by one in parallel and by then comparing the decimal places at each step. A number between $\frac{1}{n}$ and $\frac{1}{m}$ can then be constructed by carefully choosing decimal places so that the newly constructed number lies between these two numbers. The specifics do not matter here, but it is easy to see that if the two numbers are unequal there will be a point in the development of the expansions where a decimal place can be chosen for the 'in-between-number' so that the remaining decimal places cease to matter and the constructed number must lie between $\frac{1}{n}$ and $\frac{1}{m}$. This approach could be considered a valid solution, but it would completely miss the point of the exercise, since the result of the construction is not connected to the system of rational numbers through concepts such as the simplification of a fraction, at least not without the help of the concept of decimal expansions of rational numbers.

Similarly, it is unclear how the concept of cardinal numbers can be usefully applied *outside* of mathematics, as Wittgenstein notes in Ms-121, 28r.2. We have a practical extra-mathematical use for the concept of natural numbers, we can use them to count apples etc., but as Wittgenstein points out, it is far from clear why we use the word "number" in the case of cardinal numbers if we do not know which role they could play outside of mathematical propositions.

Similarly, it is unclear what the practical use could be for a proposition such "There is no greatest cardinal number.", which Wittgenstein considers in Ms-121, 28v.2 / §24. In contrast to a proposition such as "$25 \times 25 = 625$", which could be a useful answer to a practical calculation, the proposition that there is no greatest cardinal number does not serve such a purpose, but instead only appears to reflect an aspect of our concept of cardinal numbers.[19]

19 As Felix Mühlhölzer has helpfully pointed out, Wittgenstein uses the term "cardinal number" only in the sense of "natural number", the proposition "There is no greatest cardinal number" should consequently be read as "There is no greatest natural number." The former proposition follows immediately from Cantor's theorem, which states that for any set A, the power set of A has a strictly greater cardinality

The fact that we even ask whether there is a greatest cardinal number is notable, as it suggests that the answer is not entirely obvious (Ms-121, 28v.3 / §25). Wittgenstein draws our attention to an important difference between the questions "Is there a greatest cardinal number" and "Is there a greatest number of apples": In the case of numbers of apples, it seems nonsensical to even ask the question in such an abstract way, but in the case of the greatest cardinal numbers the question is sensible enough to deserve a proof and we might have been unsure or doubtful of the answer before seeing it proved. Leaving the proof aside for a minute, the difference is even more basic: We would never state "there is no greatest number of apples" as a *theorem*, we simply *act* this way. For every imaginable number of apples, we can always imagine a situation where we have one more apple. We can do so because the certainty that there is no greatest number of apples shows itself in our practical use of the words "greatest number", it is part of our way of counting and comparing different numbers, which is an important part of many of our forms of life. But no such corresponding (extra-mathematical) use exists for "greatest cardinal number".[20]

To say that there is no greatest cardinal number can conjure up the extensional image of an infinite row of numbers with no end in sight. A better picture would be to speak of a "permission" or "licence": If we give someone a licence to produce an unlimited number of a particular item, we do not mean that this person will produce a number of items that is infinite, but rather that their ability to keep producing these items is not limited by the permission or licence (Ms-

than A. As a result, a greater cardinal number can always be produced by considering the power set of a given set and then the power set of this power set, and so forth. The proof of Cantor's theorem is quite simple and employs at its heart also a diagonal argument (Priest, 1995, pp. 131–32), which opens up the proof to the kind of investigation that Wittgenstein is engaging in. His emphasis on the uselessness of such a proposition would then all the more apply if "There is no greatest cardinal number" is read as referring to transfinite numbers, even if it is more likely that Wittgenstein is thinking of the natural numbers here.

20 It is tempting to dismiss this conceptual distinction between "greatest number of apples" and "greatest cardinal number" by pointing out that our practice of counting apples can be captured quite easily by a simple axiomatisation of the natural numbers without any appeal to actual practice. In other words, it does not matter that we count *apples*, because we can give a formal definition of what it means to *count in general*, which can then be extended to and compared with the cardinal numbers. For example, using the well known Peano axioms we can define natural numbers as the constant 0 together with a successor function S, so that 1 is defined to be $S(0)$, 2 is defined to be $S(S(0))$ and so on. But such an objection would fundamentally misunderstand that these seemingly innocuous definitions are a central point of Wittgenstein's critique and that our practical arithmetic cannot simply be reduced to these logical axioms, not because these axioms were somehow wrong, but rather because these definitions depend on and only make sense in the context of our ordinary practice of counting and calculating. This topic lies at the heart of *RFM III*, but also appears in unpublished remarks between his published *RFM II* remarks in Ms-121, for example Ms-121, 57v.1 (see Section 1.4).

121, 29r.2 / §26). The permission / licence is unlimited, it has no end.[21]

We *allow* ourselves to say that there are ever greater cardinal numbers because we use them in this way. We do not *set* an end to our ability to play these language games, but it is misleading to say that the cardinal numbers *have* no end. The fact that we do not set an end to our ability to play language games with cardinal numbers is a grammatical proposition of a very particular kind:

> [...] Es wäre also wieder ein grammatischer Satz, aber von *ganz* anderer Art als "$25 \times 25 = 625$". Er wäre aber von großer Bedeutung, wenn der Schüler etwa geneigt wäre (vielleicht weil er einer ganz andern Kultur erzogen worden wäre) ein definitives Ende dieser Reihe von Sprachspielen zu erwarten. [Ms-121, 29r.3 / *BGM II* §27]

> [...] So it would again be a grammatical proposition, but of an *entirely* different kind from '$25 \times 25 = 625$'. It would however be of great importance if the pupil were, say, inclined to expect a definitive end to this series of language-games (perhaps because he had been brought up in a different culture). [Ms-121, 29r.3 / *RFM II* §27]

It is of an entirely different kind because the 'embedding' of this proposition in our life is entirely different, as the language game of cardinal numbers is used in much more abstract contexts than arithmetic with its myriad of practical extra-mathematical uses. It would thus be imaginable for a pupil from a different culture to expect a definitive end to this series, but much harder to imagine a pupil brought up with an entirely different understanding of arithmetic.

---

21 In the first edition of *RFM*, Anscombe translates Wittgenstein's "Von einer *Erlaubnis* sagen wir, sie habe kein Ende." rather awkwardly as "We say of a *licence* that it does not terminate.", which was changed in the revised 1978 edition to read: "We say of a *permission* that it has no end". The translation in the first edition can easily be (mis-)read to mean that the "licence" has no temporal end, in other words that it is an 'irrevocable' licence, which does not seem to be Wittgenstein's point here. Rather, it is a licence that does not limit the licensee's 'rights' and can be used to produce as many cardinal numbers as one wishes. This does not mean that it has to be 'eternally' valid, in fact it might be revoked if the form of life changes. But even such an alternative wording that tries to counterbalance Anscombe's translation is still misleading, because the proposition "there is no greatest cardinal number" is a mathematical proposition and thus *not temporal*, not because it were somehow eternal, but because time does not enter into it. Anscombe's translation in the first edition emphasises this temporal aspect much more than the original German and can thus be quite confusing. It has to be kept in mind that the licence is unrestricted or unlimited in its use, but not necessarily irrevocable. Anscombe's translation is all the more puzzling in light of Ms-121, 63r.4 / §45, where Wittgenstein clearly spells out these notions: "Von einer Technik zu sagen, sie sei unbegrenzt, heißt *nicht*, sie laufe ohne aufzuhören weiter – *wachse* ins Ungemessene; sondern, es fehle ihr die Institution des Endes, sie sei nicht abgeschlossen." Anscombe translates it as: "To say that a technique is unlimited does *not* mean that it goes on without ever stopping – that it increases immeasurably; but that it lacks the institution of the end, that it is not finished off." This translation is much better, but also shows that Anscombe's "does not terminate" is much closer to "goes on without ever stopping" than to "unlimited" and thus evokes exactly the opposite of what is intended by Wittgenstein.

In the next remark, unpublished in *RFM II*, Wittgenstein explicitly connects the notion of a "system of all systems", which first appeared in Ms-121, 27r.3, with the proposition that there is no greatest cardinal number:

> Wie ist es nun mit dem Satz, daß es kein System aller Systeme gibt, der dem Satz, daß es keine größte Kardinalzahl gibt, in gewisser Weise ähnlich ist? [Ms-121, 29v.1]
>
> ---
>
> Now what about the proposition that there is no system of all systems, which is in some ways similar to the proposition that there is no greatest cardinal number?

The proposition "There is no system of all systems." (or, correspondingly in Ms-121, 27r.3, "A 'system of all systems' is a contradiction.") is similar to the proposition "There is no greatest cardinal number." in so far as both state the inexistence of an *end* in a 'hierarchy' of language games, but not (if we keep in mind the picture of a "permission" / "licence") because there were something like a 'collection' of all systems or all cardinal numbers and this collection is simply too large to be 'captured' by the system of all systems. Such a too-large-to-be-captured picture would once again lead us to believe that *we know what we mean* when we talk about a "system of all systems" or a "greatest cardinal number", that we have a *clear concept* of what such a system would be like, similar to a blueprint of a construction that we are as of yet unable to construct. But instead "we have no idea how such a system of all real numbers could look like" (Ms-121, 29v.3).

This conceptual confusion is the target of Wittgenstein's critique, not our use of cardinal or real numbers. Consequently, he is not attacking set theory itself, but only our "Betrachtungsart der Mengenlehre":[22]

> Es ist in der Betrachtungsart der Mengenlehre etwas von {der // einer} primitiven Denkweise {eines wilden Volksstammes // wilder Völkerschaften}. Ich meine: ich könnte mir denken, daß ein solcher die Mathematik eines zivilisierten Volkes {erlernt // aufgegriffen} & {ihr nun eine // ihr diese} barbarische Deutung gegeben hätte. [Ms-121, 29v.2]
>
> ---
>
> There is something of the primitive way of thinking of a wild tribe in the way of looking at set theory. I mean: I could imagine such a person {learning

22 This contrasts with Steiner's view, who reads Wittgenstein as attacking set theory as a form of metaphysics in superficially mathematical clothing: Steiner, 2001, p. 270: "In short, then, attacking Cantor did not seem like attacking any branch of mathematics, since he viewed set theory as academic philosophy in sheep's clothing: a pseudoexplanatory 'theory' having no redeeming applications to anything other than metaphysics. To assault set theory, therefore, was not to revise mathematics"; Steiner, 2001, p. 271: "Wittgenstein regarded set theory as not fully mathematics." As Wittgenstein's remarks in Ms-121 make clear, it is unlikely that Wittgenstein held such an extreme view. The more charitable interpretation would be that Wittgenstein did see the mathematical content of set theory (and did not intend to attack this part), but attempted to investigate the *tendency* of set theory to lead us to philosophically nonsensical *interpretations* of these perfectly valid mathematical results.

> // taking up} the mathematics of a civilised people & would now have given it {a // this} barbaric interpretation.

The German "Betrachtungsart der Mengenlehre" can be read in two ways: It can either mean our perspective on set theory, in other words our way of looking at set theory, or it can alternatively be read as the viewpoint which we adopt when we look at something under the lens of set theory. The latter interpretation might seem more natural, but the remark makes it clear that the former is also intended: Someone with the "primitive way of thinking" of a wild tribe could learn the mathematics of a more civilised people and then give it a "barbaric interpretation". The barbaric, set-theoretic interpretation is thus independent of the mathematical system itself, which can be learned (and presumably applied) in a more civilised manner and is only later interpreted barbarically. This is an important point, because Wittgenstein's remarks can often appear as a sweeping dismissal of set theory, while his point of interest is actually much more specific and concerns the "barbaric" *interpretations* of such a theory.

We can now see more clearly why Wittgenstein is interested in the concept of a greatest cardinal number in the first place: His remarks on Cantor's diagonal argument and the idea of a greatest cardinal number are only partly directed at the concept of the real numbers, but crucially also at the idea of a "system of all systems" with all its outgrowth in logic and the foundations of mathematics. It is no coincidence that Wittgenstein's reflections in Ms-121 on Cantor's diagonal method flow seamlessly into remarks on provability in Russell's logic and (implicitly) also the use of the diagonal method by Gödel. If these seemingly unrelated remarks are left out, as happened in *RFM II*, Wittgenstein's critique of Cantor's diagonal method must appear much more myopic than it actually is, because he appears to be attacking normal mathematical practice that is practically universally considered beyond reproach.

But what if someone said: "There *must* be a system of all systems!" (Ms-121, 30r.2)? This could be the position of a logician attempting to formalise mathematics, who believes that a suitable logical system could form the foundation of all specialised mathematical systems. Accordingly, Wittgenstein considers questions concerning provability and truth in Russell's logic in the following remarks, which might at first appear to mark a departure from his thoughts on the diagonal method, but are still linked to it through the overarching theme of the system of all systems.

Leaving aside the remark Ms-121, 30r.3 (marked with a curved "S" expressing Wittgenstein's dissatisfaction), the central issue is expressed in the next two remarks:

> Beweisbarkeit ist eine *'interne Relation'* des Satzes zu den Axiomen.
>
> Soll ich nun sagen: der Beweis von p ist ein Beweis {dieses Satzes // der Wahrheit dieses Satzes} & seiner Beweisbarkeit? [Ms-121, 30v.2–30v.3]

> Provability is an *'internal relation'* of the proposition to the axioms.
>
> Shall I now say: the proof of p is a proof {of this proposition // of the truth of this proposition} *&* of its provability?

Wittgenstein is concerned here with the move from provability to truth: A proof of p could be said to be a proof of both of these aspects, truth *and* provability, because it explicitly asserts the proposition p at the end of a chain of steps of inference, it *says that* p, but it also implicitly demonstrates the provability of p through the construction of the proof. The provability of p is thus an *internal* relation, which *shows* itself as a "geometric property" of the "structure of the proposition" (Ms-121, 30v.4). Viewed through this lens, we could further choose to adopt the wording "Construct the proposition ..." instead of "Prove the proposition ..." (Ms-121, 31r.2). But then, even though we have only substituted an expression, these different wordings emphasise a different attitude towards truth: If we say that we *prove* a proposition, we think of this proposition as being true, the truth of the proposition thus seems to follow from the proof of it. But if we say that we *construct* a proposition, we can certainly construct a false proposition just as well as a true one. It can seem as if Russell's logic said that something 'really is true' if it can be proved, that a Russellian proof is more than 'just' a construction, but this is exactly the misleading idea that Wittgenstein wants to shed light on:

> Aber sagt die Russellsche Logik nicht daß etwas wahr ist, wenn es *so* konstruierbar ist? Sie sagt gar nichts darüber, sie konstruiert diese Sätze & weitere Sätze mittels ihnen.
>
> "Aber die Logik behauptet diese Sätze doch?" – Nein, sie konstruiert ihre Behauptungen. [Ms-121, 31v.2]

> But doesn't Russellian logic say that something is true if it is constructible *in this way*? It says nothing at all about it, it constructs these propositions & further propositions by means of them.
>
> "But logic does assert these propositions?" - No, it constructs their assertions.

The step from provability to truth is not something that Russell's logic can ever make on its own, the proof can only lead to a proposition, but it cannot say "it is provable, *therefore* it is true". It makes no sense to say "therefore it follows", it *just follows* and to add any "therefore" to it does not lead any closer to truth than without it. This shows striking parallels to Wittgenstein's remarks on Cantor's diagonal method in Ms-117 and his attack on the "therefore" (See Section 1.1) and makes it clear once more that his reflections on provability in Ms-121 simply change the object of study, but are still part of the same (or at least a closely related) investigation.

Of course we could object that what can be constructed in Russell's logic is precisely what is true and that this differentiates Russell's logic from less interesting logical systems. Wittgenstein considers this objection in the next remark and points out that even if we were to say this, the more interesting question is how this *shows* itself. It is certainly correct that "what is constructible is considered to be true" is a rule in our course of action in Russell's logic (Ms-121, 31v.3), but this rule cannot be justified logically 'inside' the logical system, it is justified by our use in other proofs:

> Nicht: "Was bewiesen ist, ist wahr", sondern: was bewiesen ist, wird zu weiteren Beweisen verwendet! [Ms-121, 32r.2]
>
> ---
>
> Not: "What is proved is true", but: what is proved is used for further proofs!

It might seem as if a logical system that acted as the foundation for all of mathematics, in other words a system of all systems, could once and for all settle what is true by giving us a certainty of the form "What is proved is true". But this would amount to attempting to say what needs to show itself. In the end, the truth of a proposition and as a consequence the relation between provability in a logical system and the truth of propositions proved in such a system will depend on how the truth of such propositions shows itself in the use of these propositions.

Wittgenstein then shifts his focus slightly and considers proofs by induction, starting in Ms-121, 32r.3. Only those remarks that show some connections to his remarks on the diagonal method and the "system of all systems" will be discussed here. At first glance, it might seem as if Cantor's diagonal proof and inductive proofs had little in common, but Wittgenstein is interested in the shift from the finite to the infinite, which is obviously inductive: Cantor's diagonal proof is convincing precisely by stating that for *every finite* 'sub-table' of rows and columns, the diagonalised number will be different from all rows, *therefore* the infinite diagonal will be different from all rows of numbers. (It is exactly the innocuous-looking "u.s.f. ad inf." that Wittgenstein will examine in the next remark.)[23]

Second, Wittgenstein seems interested in induction as an example of the kind of laws of inference that we use without hesitation in mathematics but which are general enough that, if formalised, would belong to the rules of a system of all systems, which gives us systematic rules that apply in all systems. Inference by induction seems to be

---

23 For a discussion of Wittgenstein's remarks on induction that focuses on his middle period, where the majority of his remarks on this topic can be found, see Ramharter, 2014. Ramharter interprets Wittgenstein as being at least partly revisionist, which may be unavoidable for Wittgenstein's more dogmatic middle period, but is not the reading employed here for his later remarks, which could be said to be "merely a dispute over words" (cf. Ramharter, 2014, p. 199). Remarks on induction appear much more rarely in the years after 1937, apart from the short discussion in Ms-121 they are mostly restricted to the later part of Ms-117 and the last third of Ms-122, which form the basis of *RFM III* and mainly revolve around Wittgenstein's concept of "surveyability" in mathematics. The connection with this concept is explored in more detail below (Section 1.4 and Section 1.6), it is certainly no coincidence that in Wittgenstein's later philosophy these concepts often go hand in hand.

"a matter of logic" (Ms-121, 32r.3). We accept the validity of this law of inference *instinctively* by "intuition" (Ms-121, 32v.3) and a system of all systems would presumably allow us to replace this informal intuition with formal laws.[24]

But at the root of such a notion lies the implicit assumption that inference by induction is a clear-cut concept that we can simply apply in a system where previously no inference by induction was known or used. We thus think of a proof by induction as an abbreviation, as 'just' a shortcut that essentially leaves the logical system as it is, similar to how the "ad inf." is 'just' an abbreviation in the picture of Cantor's diagonal argument (Ms-117, 103.2 / §11, see Section 1.1), which Wittgenstein wants to warn against:

> "u.s.f. ad inf." ist keine abgekürzte Schreibweise.
>
> Wenn man den Induktionsbeweis als eine Abkürzung auffaßt, dann ist er eine Abkürzung die gleichsam durch einen neuen Raum führt; als kürzte man den Weg von hier nach Wien dadurch ab, daß man *durch* die Erde statt auf ihrer Oberfläche fährt. [Ms-121, 33r.4–33r.5]
>
> ---
>
> "a.s.f. ad inf." is not an abbreviated spelling.
>
> If one regards the proof of induction as a shortcut, then it is a shortcut which leads, as it were, through a new space; as if one shortened the way from here to Vienna by driving *through* the earth instead of on its surface.

Such a shortcut tunnelling through the earth is not just a faster way to reach the same destination, it is a *new* construction, we introduce "a new technique into logic" (Ms-121, 33v.2). Slightly exaggerating, we could say that the apparent 'harmlessness' of a mere 'abbreviation' such as "ad inf." is the central issue that motivates Wittgenstein's remarks on Cantor's diagonal argument. This is the part of the argument that does not arouse any suspicion, it seems absolutely clear, which is exactly why Wittgenstein is interested in it. It seems as if we already knew what the ordered series of all real numbers would look like, if only vaguely, because we know what the "ad inf." means in our picture of such an ordering that we then use for the diagonalisation. Similarly, it seems as if we already knew how and when inference by induction can be applied even in the infinite case, because we have a clear picture of how we count finite objects and assume that induction is a clear generalisation of the concepts involved in counting and proceeding from one item to the next. But as Wittgenstein notes in Ms-127, 23.3 / *RFM V* §37, "the irrational numbers are – so to speak – special cases." The 'general' technique of induction, the 'general' "u.s.f. ad inf." is a *new invention* in a particular logical system, the logical system in which we reason by proof of induction is not the same logical system as the one without this proof technique (as if the proof by induction were some kind of innocuous 'tool' that

24 See Section 3.5 for a discussion of the concept of intuition in relation to computing machines, which also touches on the issue of replacing informal intuition with mechanical rule following.

we could either apply or leave in the toolbox without any impact on the 'system itself').

But since there is no system of systems in the background that could justify our application of laws of inference, the application of something like proof by induction needs to be justified in the special case, by the use of the specialised system. Wittgenstein thus flips the traditional view, that a concrete system such as arithmetic rests on a more abstract and foundational system of formal logic, on its head: A concrete proposition like $2 \times 2 = 4$ can be considered to rest on the foundations of the whole system of arithmetic and consequently on a whole system of axioms and rules of inference, but we could just as well say that this whole system rests on the useful application of the concrete proposition. The multiplication thus proves something in the "geometry of the numerals" (Ms-121, 34r.2), which is why the practical applications are paramount for a philosophical investigation of something like Cantor's diagonal argument:

> Es ist also wichtig zu fragen: Wie kann der Satz, daß die Rationalzahlen sich in eine Reihe ordnen lassen, praktisch angewandt werden? [Ms-121, 35v.2]
>
> ---
>
> So it is important to ask: How can the proposition that the rational numbers can be ordered in a series be applied practically?

Here we can imagine many possibilities, for example as part of a description of an automated machine that sequentially runs through all the rational numbers and tests them for a particular property. But in the case of the irrational numbers, we had no practical use for such an ordering in mind and *then* Cantor came along and showed us the futility of such an endeavour. Instead of originating in practical requirements, Cantor's proof of the impossibility of such an ordering for the real numbers closed off this avenue before a practical, extra-mathematical use was even considered. It gives us a *method* to "destroy" any imaginable ordering of the irrational numbers (Ms-121, 35v.3 / §28).[25]

We might think that by giving us the possibility to construct a diagonalised number for any ordering of the irrational numbers Cantor's diagonal method shows us an irrational number that differs from all the numbers in the (ordering) system. This is certainly true for the algebraic numbers, where we can use the diagonal method to construct

25 Wittgenstein uses "(zer)stören", with the "(zer)" inexplicably dropped from *RFM* and the "stören" translated by Anscombe as "upsetting any order". Mühlhölzer, 2020, p. 180 (footnote), notes the inadequacy of Anscombe's translation and uses "to disrupt" instead, while following Anscombe in emphasising the "stören" more than the "zerstören". But a disrupted ordering could still be considered an ordering, merely slightly deficient, whereas Cantor's proof convinces us to completely give up the idea that there is or could be anything that we would call an ordering of the real numbers, it *destroys* the idea of such an ordering. Mühlhölzer, 2020, p. 180: "[O]ne might wish rather to say that this method does not "disrupt" any order but shows that *there isn't* a way to order all the real numbers and that, therefore, they *cannot* be ordered" (emphasis in the original).

a transcendental number. But in the case of the irrational numbers, the diagonal method is purely 'destructive', it *destroys* an possible ordering, but it does not *construct* a concrete irrational number, because it depends on a system that is immediately rejected at the conclusion of the diagonal argument. Instead of showing us an irrational number that is different from all the numbers in the system, the diagonal method gives meaning to the concept of a number that is different from all the numbers in the system:

> Das Cantorsche Diagonalverfahren zeigt uns nicht eine Irrationalzahl die vor allen im System verschieden ist, aber es gibt dem mathematischen Satz Sinn die Zahl so & so sei von allen des Systems verschieden. Cantor könnte sagen: Du kannst *dadurch* beweisen, daß eine Zahl von allen des Systems verschieden ist, daß Du beweist daß sie in der ersten Stelle von der ersten Zahl, in der zweiten Stelle von der zweiten Stelle von der zweiten Zahl u.s.f. verschieden ist.
>
> Cantor sagt etwas über die Multiplizität des Begriffs "{Reelle Zahl // Entwicklung}, verschieden von allen eines Systems." [Ms-121, 36r.2 / *BGM II* §29]
>
> ---
>
> Cantor's diagonal procedure does not shew us an irrational number different from all in the system, but it gives sense to the mathematical proposition that the number so-and-so is different from all those of the system. Cantor would say: You can prove that a number is different from all the numbers in the system *by* proving that it differs in its first place from its first number and in its second place from its second number and so on.
>
> Cantor is saying something about the multiplicity of the concept "Real number different from all the ones of a system". [*RFM II* §29]

By considering numbers in systems of extensions and showing how, extensionally, we can methodically dodge the decimal places of the extensions in the system, Cantor gives meaning to the concept of extensions that are different from all other extensions in the system. This can appear to present us with a clear and systematic picture of extensions, but in fact "the grammar of the word "extension" is not yet determined" (Ms-121, 36v.2 / §30).[26] We have no concept of the totality of all irrational or real numbers, exactly because irrational numbers are, intensionally seen, "special cases" (Ms-127, 23.3 / *RFM V* §37, quoted above) and Cantor achieves more conceptual clarity only by "proposing" to call an extension different from the others in a system if it is "diagonally different" (Ms-121, 36v.3 / §31).[27] It might seem as if the extensional view were the full picture, but by

---

26 Translation from Mühlhölzer, 2020, p. 187, who convincingly argues for the translation of the German "Extension" as "extension" against Anscombe's translation as "expansion". It also contains a much more detailed interpretation of Wittgenstein's "non-extensionalist" stance in *RFM II* as well as *RFM V* than can be touched upon here.

27 As Frascolla, 2006, p. 159 puts it: "The task of a creative mathematician like Cantor is not to establish an indisputable truth, but to induce us to enrich in a certain way our linguistic apparatus. [...] The ability of the creative mathematician is shown by his proposing a sign construction (proof) which produces, in all those who have received a certain training, the willingness to introduce a certain new conceptual tool."

sidestepping the variety of intensional language games in favour of a uniform view focused solely on the decimal expansions of a system of extensions Cantor's diagonal argument gives up the ability to clarify our use of the word "extension". This is not surprising, since such a clarification of grammar is usually not a focus of interest for mathematicians and requires a philosophical investigation of our use of the word.

What Cantor gives us is a "task": "Find a number whose expansion is diagonally different from those in this system" (Ms-121, 37r.2 / §32), not a number itself. We might think that this distinction is either unimportant or that it collapses in the case of the diagonal number, but for Wittgenstein the misleading aspect of Cantor's diagonal argument is precisely that it tends to blur this very conceptual distinction. This becomes clear in the next four remarks between §32 and §33, which were unfortunately not published in *RFM*, but which form a very succinct summary of the issue that motivated Wittgenstein's interest in Cantor's diagonal argument since the very first remarks in Ms-117.[28] In the first of the unpublished remarks (Ms-121, 37r.3), Wittgenstein draws our attention to the fact that the *decimal expansion* of π is usually of no interest to us in mathematics, it is secondary to the use of π as a number. What then follows is a very succinct summary of this distinction and is worth quoting in full:

> Wenn wir ein System von *Regeln* der Entwicklung haben, können wir eine Regel {bilden // geben}, so daß ihre Entwicklung Schritt für Schritt von denen des Systems verschieden ist.
>
> {Es ist nun ein großer Unterschied // Aber hier ist ein Unterschied}, ob die Regel von den *Entwickelungen* ausgehend durch ihre Änderung die neue Entwickelung hervorbringt, oder ob sie einen andern Ausgangspunkt hat aber ein Beweis dafür existiert, daß ihre Entwickelung Schritt für Schritt von denen des Systems verschieden ist. [Ms-121, 37r.4]
>
> ---
>
> If we have a system of *rules* of expansion, we can {form // give} a rule so that its expansion, step by step, is different from those of the system.
>
> {Now // But here,} there is a great difference, whether the rule, starting from the *expansions*, by its change brings forth the new expansion, or whether it has a different basis but there is a proof that its expansion is, step by step, different from those of the system.

The diagonalised number clearly corresponds to the former case. There is no "other origin" for the diagonalised number and *then additionally* a proof that it is different from all the other numbers in the system, it is *solely* based on the diagonal modification of decimal expansions. The diagonal "sleight of hand" ("Taschenspielerkunststück"[29])

28 See Mühlhölzer, 2020, pp. 135–37, which not only discusses the importance of the unpublished remarks in this context but also highlights the inadequacy of Anscombe's translation of §32 ("Thus it can be *set* as a question" for "Es gibt also eine *Aufgabe*").

29 This expression is used by Wittgenstein in multiple remarks, most notably in Ms-122, 117v.1 concerning the use of Russell's logic as an "abbreviation", which shows parallels to the "u.s.f. ad inf." in the diagonal method, but also as "Taschenspielertrick" in Ms-159, 23r.1. Another important use is Ts-227a, 189.4 / *PI* §308 and its variants

makes us overlook this important difference, however. This does not usually happen in our practical use of numbers, for example when a textbook (such as Hardy's) gives examples of irrational numbers, where the examples consist only of numbers of the second kind, for which the decimal expansion is not an inherent part of their definition (Ms-121, 37v.2).

It is the difference between a rule that *"results in"* a different expansion (as a side effect) and one that *"produces"* a different expansion (as its only purpose). In a way, a diagonally produced number lacks the "essence" of what we usually call a number, because it is as if there is no number that the decimal expansion belongs to (Ms-121, 38r.2).

The next two remarks are published in *RFM* again, and here Wittgenstein returns quite explicitly to the notion of a system of all systems, an "Über-System" for the irrational numbers. While at first glance these remarks might seem to express little more than his first remarks in Ms-117, they actually build heavily on the previous remarks in Ms-121 with its focus on "systems" and introduce an important new aspect, the comparison of an element *in* the system with the *system as a whole*. It has by now become clear that Wittgenstein considers the irrational numbers as being far from uniform (in contrast to the rational numbers), they consist of *"diverse systems"* in the number line:

> Man könnte sagen: Außer den rationalen Punkten befinden sich auf der Zahlenlinie *diverse Systeme* irrationaler Punkte.
>
> Es gibt kein System der Irrationalzahlen – aber auch kein Über-System, keine 'Menge der irrationalen Zahlen' von einer Unendlichkeit höherer Ordnung. [Ms-121, 38v.1 / *BGM II* §33]

> It might be said: Besides the rational points there are *diverse systems* of irrational points to be found in the number line.
>
> There is no system of irrational numbers – but also no super-system, no 'set of irrational numbers' of higher-order infinity. [*RFM II* §33]

The first paragraph is clear enough, but what can we make of second part of the remark? That Wittgenstein does not dispute the concept of the set of real numbers, nor the uncountability of this set, has become clear on a number of occasions, though he certainly tries to clarify the expressions surrounding these proofs so that they appear less like mathematical discoveries and more like conceptual stipulations. His rejection of a system of irrational numbers and of an "Über-System" of higher-order infinity should thus not be read as a dogmatic thesis[30], but only as a description of our grammar: We simply do not use irrational numbers in a way that presupposes a system or could

---

(with "Taschenspielerkunststück" replaced with "Taschenspielerstück") in Ms-116, 335.1, Ts-228, 182.5 and Ts-230a/b/c, 94.3.

30 *Contra* Mühlhölzer, 2020, p. 189, who holds the view that "[i]t is not only too dogmatic, but also smacks of "ontology", which is far removed from Wittgenstein's intent in his later philosophy."

make use of such a system, nor do we have any concept of an "Über-System" that we apply in our use. Cantor's diagonal argument, on the other hand, by making it seem as if there were such an "Über-System" of a higher-order infinity, disregards and misunderstands the grammar of "irrational number". Or rather, Cantor's diagonal argument *changes* our grammar by introducing a new concept but leads us to believe that the diagonal method merely presents us with a precise expression of an already existing concept. This is made even more clear in the next remark, which acts as the central junction point of the reflections in Ms-121 up until now:

> Cantor definiert eine *Verschiedenheit höherer Ordnung* nämlich eine 'Verschiedenheit' einer Entwicklung von einem *System* von Entwicklungen. Man kann diese Erklärung so benützen, daß man zeigt daß eine Zahl in diesem Sinne von einem System von Zahlen verschieden ist: sagen wir π von dem System der algebraischen Zahlen. Aber wir können nicht gut sagen, die Regel, die Stellen in der Diagonale so & so zu verändern, sei {dadurch // nun} als von den Regeln des Systems verschieden {bewiesen // demonstriert}, weil diese Regel selbst 'höherer Ordnung' ist denn sie *handelt* von der Veränderung eines Systems von Regeln & daher ist es von vornherein nicht klar, in welchem Fall wir die Entwickelung so einer *Regel* von allen Entwicklungen des Systems verschieden erklären wollen. [Ms-121, 38v.2 / *BGM II* §34]

> Cantor defines a *difference of higher order*, that is to say a difference of an expansion from a *system* of expansions. This definition can be used so as to shew that a number is in this sense different from a system of numbers: let us say π from the system of algebraic numbers. But we cannot very well say that the rule of altering the places in the diagonal in such-and-such a way is as such proved different from the rules of the system, because this rule is itself of 'higher order'; for it *treats of* the alteration of a system of rules, and for that reason it is not clear in advance in which cases we shall be willing to declare the expansion of *such a* rule different from all the expansions of the system. [*RFM II* §34]

According to Wittgenstein, Cantor's diagonal argument can be accused of applying double standards: Going in to the proof, we have a clear picture only of what it means to compare rules that describe decimal expansions with *each other*, but then Cantor compares a rule inside the system with the *whole system* while appearing to use only the normal standard of comparison. Instead of comparing only rules of the form that we already know, the diagonal method in fact *creates* a new kind of "higher-order" rule, it presupposes the whole system and *changes* the system of rules. It is not a discovery in the realm of the infinite, but rather a conceptual invention.

1.3 What counts as a number?

That Wittgenstein is not entirely satisfied with the quality of the remarks in Ms-121 becomes evident in the 11 remarks leading up to the next remark published in *RFM II*, where he is constantly revisiting and reworking the problematic aspects of treating decimal expansions as numbers. Wittgenstein laments that showing where an absurdity lies is much harder than merely noticing that a mathematical interpretation is absurd in the first place (Ms-121, 39r.3) and then compares multiple rules that explicitly change the decimal expansions of π with π itself, for example:

> Wie wäre es mit diesem Satz: Es gibt eine Zahl, die an jeder Stelle von π verschieden ist. Nämlich die Regel, jede Stelle von π in irgend einer Weise zu verändern –? [Ms-121, 40v.3]
>
> ---
>
> How about this sentence: There is a number that is different in every place from π. Namely, the rule to change each place of π in some way -?

The importance of this remark becomes clear if it is read in the context of Ms-117, 98.2–99.1 / §§2–3, where Wittgenstein brought up the task "Name a number that agrees with $\sqrt{2}$ [only] in every other decimal place" and considered why a diagonalised number would not be a satisfying answer but might elicit a reaction of the form "But that's not what I meant!". Now, in Ms-121, Wittgenstein arrives at the most absurd way of 'answering' such a question, which is to answer it by merely restating the task as a rule. It is immediately obvious that such an 'answer' is shallow and 'contains' no 'surplus meaning' so to speak, its meaning depends entirely on the question and is trivially derived from it. It thus forms the trivial end of the scale of possible answers, with something like "π" being on the other end of the scale. But then why do we accept the diagonal method as a valid answer? If it lies somewhere on the scale between a meaningful answer such a "π" and the trivial reformulation of the task, are we really certain that it is closer to a meaningful answer than to a trivial one? Where and why do we draw the line?

It might be objected that such a trivial restatement of the question as an answer is only possible if we are looking for a number that is different from a *single* other number, but not applicable in the case of an infinite list of numbers, because Cantor's insight lies in the application of the modification of decimal places to a whole system of expansions, with the actual mechanism of modification (subtraction or addition of 1, for example) being secondary and trivial. This is of course correct, but if we consider the mechanism of changing the decimal places unimportant and trivial, why not the whole application of the diagonal method? Of course in our culture the diagonal method is associated with important proofs and thus enjoys a certain standing, but as Wittgenstein mentioned, we might also imagine that the diagonal method had originally been taught to school children. If we had been accustomed to the diagonal method from an early age, we might reject the notion that something as trivial as a diagonally different number could be considered a 'proper' or 'real' number. At the crucial point of Cantor's diagonal argument, such a person might then reply: "Yes, I agree that at present we have no way of ordering the real numbers in a way that would correspond to the ordering of the rational numbers, but the diagonal 'number' just constructed is not really a number and thus not a proof that an ordering of the real numbers *cannot* exist. Who knows, we might invent something in the future that we would call such an ordering but even then your diagonal method will do no harm, because it can never construct a number 'on its own'. (And if it by chance constructs a 'proper' number such as π then it does not matter because the number will already be in the list.)" We would certainly say that such a person does not understand our mathematics or has a completely different conception of it, but does that mean that their conception of mathematics must be flawed? What Wittgenstein seems to be driving at is not that Cantor's diagonal argument should be abandoned, but that we *could imagine* a way of doing mathematics that remains completely unimpressed with Cantor's diagonal argument and any talk of the transfinite or infinities greater than other infinities. These people would simply reject any methods that "tamper with the extension":

> Warum sollten wir nicht sagen: die Regel, die Diagonale zu verändern, sei mit den Regeln des Systems *unvergleichbar*?
>
> "tamper with the extension" [Ms-121, 41r.3]

> Why should we not say: the rule of changing the diagonal is *incomparable* with the rules of the system?
>
> "tamper with the extension"

That Cantor's diagonal argument considers the "higher-order" diagonal rule, which depends on and compares its expansion with a whole *system* of decimal expansions, as comparable to the other decimal expansions is a conceptual *decision*. We allow such a tampering with the extension in our mathematical calculus, but it is by no means immediately obvious why we choose to do that. Cantor's proof, with its extensionalist language, has the tendency to obscure the differences between the different conceptions of numbers at play here. Wittgenstein takes one more shot at this sleight of hand:

> Ich verstehe, daß man von zwei arithmetischen Regeln sagt, sie seien verschieden wenn die eine an der ersten Stelle eine andre Ziffer ergibt, als die andere – – – aber kann man auch sagen, die Regel, die Entwicklung von χ hinzuschreiben, aber die erste Stelle zu verändern, sei von χ verschieden, *da* die Entwicklungen an der ersten Stelle nicht übereinstimmen?? [Ms-121, 41r.4]

> I understand that one says of two arithmetical rules that they are different if one produces a different digit in the first place than the other - - but can one also say that the rule of writing down the expansion of χ, but changing the first place, is different from χ, *because* the expansions in the first place do not agree??

This is a beautiful clarification of the different concepts of numbers at play in Cantor's proof. What counts as a number might seem absolutely clear and of course it *is* clear in most situations. We can say that two numbers are different *if* they show a difference in their decimal expansions, because the decimal expansions are only the 'inessential surface' of numbers. Wittgenstein does not want to deny this fact, nor our practice of using comparisons of decimal expansions in practice to decide the equality of numbers. But Cantor's diagonal argument plays a new kind of language game, it suddenly compares numbers that are only different *because* they *are* different decimal expansions. An inessential surface feature has become the essence. We thus need to carefully revisit even concepts that we have until now taken for granted or risk being mislead by a "boastful" interpretation.[31]

This boastful aspect of Cantor's proof is also discussed in a short fragment of loose pages, Ms-178c,[32] where Wittgenstein explicitly calls out the different number concepts at play in Cantor's diagonal argument. This is what makes Cantor's diagonal argument boastful, it promises more than it actually delivers:

> Das Bild der Cantorschen Überlegungen ist ungemein irreführend. Es zeigt uns nämlich Extensionen & Zahlzeichen, die doch nicht als Zahlzeichen zu benützen sind.
>
> ["Aber ist das nicht im Beweis vermieden: – – –?"] So daß es nicht klar ist, ob wir den Extensionen eine neue Extension einfügen (wie es ja aussieht) {oder ein neues Gesetz aufzeigen. // oder {den Gesetzen // dem System der Gesetze} ein neues Gesetz beifügen.}
>
> Wie kommt es, daß dies kleine Stück {Mathematik // Rechnung} so {Großes // viel} leistet? – {Weil es etwas verspricht, nicht ein Versprechen hält. //

31 Wittgenstein's remark in Ms-121 shows a striking resemblance to a much earlier remark that first appeared as Ms-107, 63.6 in 1929 and was also included in two typescripts from 1930 (Ts-208, 78r.8; Ts-209, 102.1):

> Man kann nicht sagen: zwei reelle Zahlen sind identisch wenn sie in allen Stellen übereinstimmen. Man kann nicht sagen: sie sind verschieden, wenn sie an einer Stelle ihrer Entwicklung nicht übereinstimmen. Man kann ebensowenig sagen, die eine sei größer als die andere, wenn {ihre // die} erste {nicht übereinstimmende // unpaarige} Stelle größer sei als die entsprechende der anderen.

> One cannot say: two real numbers are identical if they agree in all places. One cannot say: they are different if they do not agree at one point of their expansion. Nor can one say that one is greater than the other if the first digit is greater than the corresponding digit of the other.

At first, this earlier remark seems to stand in direct opposition to Wittgenstein's later view in Ms-121, but on closer inspection of the context in Ms-107 the remarks show an important continuity in Wittgenstein's thinking from 1929 to 1938. In his earlier remarks, Wittgenstein is occupied with numbers that are different from other real numbers only through an explicit modification of their decimal expansions (much like the diagonalised number in Ms-121). The comparability of *these* 'numbers' is what Wittgenstein wants to deny, because in his view real numbers are more than just their extensional aspect ("A real number *yields* extensions, it is not an extension.", Ms-107, 62.7), with its echo in the "essence" of numbers in Ms-121, 38r.2, quoted above. But even back in 1929 he does not deny that numbers such as π and e can be compared by comparing their decimal expansions (Ms-107, 64.2). The difference between 1929 and 1938 is less a total rejection of earlier viewpoints and more a clarification of the conceptual confusion in a short remark that strikes more precisely at the heart of these issues while at the same time being much less dogmatic.

32 Ms-178c was most likely written after the remarks in Ms-121, as one of the remarks shows textual similarities with the other document (Ms-178c, 3.2 and Ms-121, 90v.2), with the version in Ms-178c apparently written later.

> Weil seine Leistung ist, zu versprechen; nicht: Versprechen zu halten.} [Ms-178c, 1.1–1.2]
>
> ---
>
> The picture of Cantor's reflections is extremely misleading. It shows us extensions & number signs, which however are not to be used as number signs.
> ["But is this not prevented in the proof: - - -?"] So that it is not clear whether we are adding a new extension to the extensions (as it seems) {or showing a new law. // or {attaching a new law to the laws // to the system of laws}}
>
> How is it that this little piece of mathematics does {such a big thing // so much}? - {Because it promises something, not because it keeps a promise. // Because its achievement is to promise, not to keep a promise.}

The next remark in Ms-121 is published in *RFM II* and forms both a return to Wittgenstein's earlier thoughts on the greatest cardinal number and a continuation of the idea that we *choose* to adopt the new number concept exhibited by the diagonalised number in Cantor's proof. That these considerations "may lead us to say" that there is no greatest cardinal number, that "we can *let* the considerations lead us there"[33], that "we can say *this* and give *this* as our reason" (Ms-121, 41v.2 / §35) is not to be taken lightly. Wittgenstein is very careful not to paint a picture of mathematics as an arbitrary formal game, to the contrary: The fact that "$2 \times 2 = 4$" is 'only' a rule does not mean that we can simply swap the language game of arithmetic with another made up game, because "$2 \times 2 = 4$" is *anchored* in our whole fabric of life with its myriad of related language games in a way that another arbitrary game is not. But can the same be said about the conceptual decision at the heart of Cantor's diagonal argument, the decision to count a 'pure' decimal expansion (without any other intensional number rule behind it) as a number? Wittgenstein raises this question in the second part of the remark:

> Aber wenn wir es nun sagen – was ist weiter damit anzufangen? In welcher {Praxis // Anwendung} ist dieser Satz *verankert*? Er ist vorläufig ein Stück {mathematischen Gerüsts // mathematischer Architektur}, das in der Luft hängt, so aussieht als wäre es, sagen wir, ein Architrav, aber von nichts getragen wird & nichts {trägt // tragen kann}. [Ms-121, 41v.2 / *BGM II* §35]
>
> ---
>
> But if we do say it – what are we to do next? In what practice is this proposition anchored? It is for the time being a piece of mathematical architecture which hangs in the air, and looks as if it were, let us say, an architrave, but not supported by anything and supporting nothing. [*RFM II* §35]

Is this not exactly the kind of dogmatism that Wittgenstein is supposedly trying to dispel? What can we make of such a blanket criticism, targeted at a beautifully simple mathematical method with a

---

33 Anscombe translates this quite misleadingly as "we can *make* the considerations lead us to that", putting the focus on our action instead of reading it as letting ourselves be lead by something else.

variety of uses in different mathematical fields?[34] It might seem as if Wittgenstein went too far here, but this remark actually follows as a rather natural continuation of his previous (and sadly unpublished) remarks. Importantly, Wittgenstein says that *"for the time being"* the proposition that there is no greatest cardinal number is "not supported by anything and supporting nothing". The critique is not targeted at mathematical practice and even less at practical uses of such a proposition (and how could philosophy ever be certain that no such use might be found in the future), but only at our acceptance of such a propositions *solely based* on Cantor's diagonal argument. As Wittgenstein remarks in Ms-121, 42r.3 / §37, the proposition "supports as much as the grounds that support it do", but Cantor's proof alone does not give us sufficient "grounds" to accept or reject the conceptual decision of what to count as a number. It is a valid and perhaps even interesting invention, but one for which we have not yet found a practical use.

This is further spelled out in the next three remarks, the first of which is unpublished in *RFM II*. Wittgenstein compares the proposition that there is no greatest cardinal number with the proposition "$25 \times 25 = 625$" *in isolation*, that is to say as a single proposition without the context of the whole technique of arithmetic around it. We might think that even in isolation, such a proposition is at least the "rudiment of a mathematical technique", so that it represents "a small selection of the great truth of the whole system" (Ms-121, 42v.1). Wittgenstein does not explicitly speak of discovery and invention in this remark, but it is clear that this notion of a small slice of a larger truth commits us to a picture of mathematical discovery and convinces us to look for the "whole system" that is yet to be discovered.

Here we are mislead by the ostensible analogy between a finite series of numbers (which we can count and for which we have a whole grammar of talking about the size of a series) and an infinite series whose size can be $\aleph_0$. But according to Wittgenstein, a "series in the mathematical sense is a {method of construction for series of linguistic expressions // series of possibilities of linguistic constructions}"

34 Ramharter, 2018, p. 136 offers a harsh interim conclusion: "[Wittgenstein] knew neither the mathematical considerations nor the theological-philosophical motives that led Cantor to the development of the cardinal numbers" (my translation). But Priest, 1995, pp. 127–138 cites a number of examples of Cantor's own philosophical interpretation of the "transfinite" and "absolute infinity" that can be considered prime examples of the kind of mathematical over-interpretation and resulting philosophical confusion that Wittgenstein is attacking here. Whether or not Wittgenstein knew these "theological-philosophical motives" is of secondary importance, what matters is that he is certainly not attacking a straw man, but an actual source of conceptual confusion. (Whether mathematicians nowadays still hold these views is of course another matter, but even here Ramharter seems to be focused primarily on the 'working mathematician' and less on philosophical issues in foundational areas of mathematics or logic, where Wittgenstein's critique is still relevant, though sometimes more polemical than strictly necessary.)

(Ms-121, 43r.1 / §38)[35], for which the 'size' $\aleph_0$ is only "a *kind of* number", but without the grammar that characterises our use in the finite case. We can certainly form the expression "class of all classes which are equinumerous with the class 'infinite series'", but we have no use for it. As Wittgenstein notes at the end of the remark, it is "not: yet to be discovered, but: still to be *invented*". This is beautifully illustrated by the next remark (Ms-121, 44v.2 / §39), where Wittgenstein compares the situation with a game that shows superficial similarities to chess, with a "playing-board divided into squares" and with "pieces likes chess pieces on it". Someone could then explain:

> "{Das ist der // Diese Figur ist der} *'König'*, das sind die *'Ritter'*, das die *'Bürger'*. – Mehr wissen wir von dem Spiel noch nicht; aber das ist immerhin etwas. – Und mehr wird vielleicht noch entdeckt werden." [Ms-121, 44v.2 / *BGM II* §39]

> "This piece is the '*King*', these are the '*Knights*', these the '*Commoners*'. – So far that's all we know about the game; but that's always something. – And perhaps more will be discovered. [*RFM II* §39]

It is immediately obvious why such an explanation sounds quite strange. A game is defined by its rules, but it does not make sense to say that we will *discover* more rules if the rules of the game are not yet fully specified. We are mislead by the similarities with the other, fully specified game of chess and think that we could 'discover' more rules in our new game by looking at the existing game of chess. But these existing rules cannot give us the justification or the "grounds" for new rules in our new game. They can certainly inspire us to give the new game some rules with certain analogies to chess, but these new rules will be an *invention* nonetheless. The same is true for Cantor's diagonal argument, even though the lofty mathematical language makes it much less obvious.

The remarks on Cantor's diagonal argument then trail off and instead Wittgenstein focuses mostly on other topics for the next 15 pages, with a break of around two months in between (from July to September of 1938).

1.4 SURVEYABILITY, RUSSELL AND NUMBERS

During his 'excursion' between §39 and §40, Wittgenstein returns briefly to Cantor in Ms-121, 49v.3 before writing about "unfolding" certain properties of a number of marbles Ms-121, 51r.1–55r.4. This is an example that Wittgenstein already used in Ts-222, 30.3 / *RFM I* §36, where 100 marbles are arranged in 10 rows of 10 marbles.[36]

---

35 *RFM II* contains only the first variant. Furthermore, Anscombe puts the first "series" in single quotes, even though there is no indication for this in the original manuscript.

36 Most of the remarks between Ms-121, 51r.2 and Ms-121, 53r.2 occur also at the end of Ts-221a/b in Ts-221a/b, 267.1–268.5, but were not included in Ts-222, which forms

Such an arrangement accomplishes two things: First, it "unfolds" the mathematical property that 100 marbles are 10×10 marbles and can thus act as a proof of the proposition "$10 \times 10 = 100$". This observation is closely connected with Wittgenstein's distinction between experiment and calculation, which is more thoroughly explored in Ts-221a/b, Ts-222 / *RFM I*, but is of secondary importance for the present discussion. Second, it gives us a more *surveyable* or *perspicuous* representation of the 100 marbles, because the arrangement of 10 by 10 allows us to count the whole collection much more quickly. If the marbles are sufficiently uniform and the arrangement tidy enough, we can see at a glance that the arrangement forms a square, we thus need to count only the number of marbles in a row, square it and arrive at the result of the whole calculation.

The arrangement could be said to be a new *notation* for the number of marbles, which instead of showing the number as a unary row of 100 single marbles ("| | | | | | | | | | ...") shows them as a square of 10 marbles ("$||||||||||^2$"). By "unfolding" the marbles in such a way, first as an experiment and then later as a calculation, as a *rule*, we have made the first step towards our (even more surveyable) decimal notation, where unary rows of single strokes up to "| | | | | | | | |" are replaced by 0 to 9, so that "| | | | | | | | | | × | | | | | | | | | |" is instead written as "10×10".

The next remarks are unique to Ms-121, but continue the line of thought. As Wittgenstein mentions in Ms-121, 55v.2, his aim is to compare the use of counting in mathematics with the use outside of mathematics. In purely mathematical situations, especially when considering foundational questions, we can often be lead to believe that the particular manner of counting were unimportant, so that the difference between unary counting and decimal counting would be only an inessential matter of definition. This is the case for Russell's *Principia Mathematica* (Ms-121, 55v.3), a system in which calculations such as $2 + 3 = 5$ are possible but become so unwieldy that they resemble a unary calculation such as | | + | | | = | | | | |. As a consequence, additions and multiplications in *Principia Mathematica* are *in theory* calculable for any natural number, but become too error-prone *in practice* for any except very small numbers. From the perspective of a logician, this practical difference is an unimportant detail and a system such as Russell's *Principia Mathematica* shows us the foundation of our techniques of arithmetic. A logical system thus provides us with the *certainty* that 12×12 *must* equal 144, it gives us confidence that our arithmetic is correct. But for Wittgenstein, the inverse is true: A proof can only act as a proof if it is "surveyable" ("über-

the basis of *RFM I*. Their inclusion in Ts-221a/b underscores their close connection with the beginning of Ms-162a, which can be assumed to have been written around the same time as the remarks in Ms-121: Ms-162a, 1.1–8.4 correspond to the remarks Ts-221a/b, 268.6–271.2, in other words exactly to the remarks in Ts-221a/b that follow the remarks in Ms-121.

sichtlich" / "übersehbar"), because we must be able to use a proof as a paradigmatic picture which we can "reproduce with certainty":

> [...] {Zum Beweis gehört, daß seine Vorgänge übersichtlich sind // Die Vorgänge eines mathematischen Beweises müssen übersehbar sein}, d.h. wir müssen im Stande sein, ihn mit Sicherheit immer wieder richtig reproduzieren zu können. (Was ist das Kriterium dieser Sicherheit?) [Ms-121, 55v.3]

> [...] {For a proof it is essential that its processes are surveyable // The processes of a mathematical proof must be surveyable}, i.e. we must be able to reproduce it correctly again and again with certainty. (What is the criterion of this certainty?)

In practice, we cannot reproduce large numbers written in unary notation without making a lot of mistakes, nor can we use them in arithmetical proofs without being unsure that we have not miscounted such a number at some step along the way. As a consequence, we cannot *calculate* with unary numbers once they exceed a certain size, in the same way that we could not play chess if we could not distinguish the chess pieces. But even this way of phrasing the situation is dangerously misleading, because it still appears as if the unary numbers were still *numbers* in the same way that decimal numbers are numbers, merely more cumbersome, so that the 'mathematical essence' of "| | | | |" and "5" remains the same. But as Wittgenstein remarks in the context of the chess piece example, a game with uniform pieces *ceases to be chess* and its pieces *cease to be chess pieces*, because only a game with distinguishable pieces can be called chess, *is* chess (Ms-121, 56v.2).

This sheds light on the philosophically misleading role of definitions, even (or especially) if they are used to 'merely' introduce a more convenient notation. A calculation such as $129 \times 336 = 43344$ does not have a surveyable representation in Russell's *Principia Mathematica* and thus does not follow from a proposition expressed in such a system. To say that "it follows via suitable definitions" is of course true, but (contrary to what Russell's *Principia Mathematica* might suggest) entirely meaningless, as there is nothing forcing us to adopt these particular definitions over any others (Ms-121, 56v.2).

The justification of the proposition $12 \times 12 = 144$ cannot be stated and proved by Russell's system, instead it shows itself in our techniques of addition and multiplication, which are embedded in and given meaning by a whole form of life. Arithmetic using decimal notation is not secondary to the more foundational Russellian logic, but the other way around, because we would have no concept of arithmetic if not for our more 'primitive' non-logical notation.[37] This flips

---

37 This is also emphasised by Marion, 1998, pp. 232–233:

> This is why Wittgenstein claimed that there is a danger in looking 'at the shortened procedure as a pale shadow of the unshortened one' (RFM iii, § 19). On the contrary, the shortened procedure (decimal notation) tells us the outcome of the unshortened procedure (stroke notation). [...] It is important in this context to understand that Wittgenstein's argument rests

on its head the idea of a 'primitive' culture that has not yet developed mathematics rigorously enough to be able to calculate with certainty, so that their calculation of 10×10 sometimes results in 99, sometimes in 100 and sometimes in 101, but where "all three are considered correct", an example that Wittgenstein gives in Ms-121, 57v.2, albeit without the link to the notion of a 'primitive' form of mathematics as proposed here.[38] Such an uncertain form of arithmetic, however, is precisely the arithmetic of a supposedly certain and rigorous system such as *Principia Mathematica*, which we can only rescue from its primitive state by discarding it as our standard of measurement in favour of our traditional decimal arithmetic. In other words, if someone proved to us repeatedly and without any discernible mistake that in Russell's system 10×10 resulted in 99, we would nevertheless consider the proof to be flawed, because we know that 10×10 *must* equal 100, thanks to our knowledge of arithmetic. In case of a conflict between Russell's logic and our traditional arithmetic, the latter will be the judge of what is considered correct, which shows the absurdity of believing that the logical calculus could justify arithmetic.

This is why it is highly misleading to say of a definition that it "only abbreviates" an expression. Instead, we should think of it as "introducing a new calculus", as Wittgenstein suggests in Ms-121, 57r.4. Of course Russell's *Principia Mathematica* is only an extreme example, the same philosophical observation applies even to basic Peano arithmetic, numbers defined in terms of a single base number 0 and a successor operation, or the equivalent example that Wittgenstein gives in Ms-121, 57v.1, "$1, 1+1, 1+(1+1), 1+(1+(1+1))$ etc". This is why to the philosopher, a mathematical definition cannot be an afterthought, but rather the start of a detailed philosophical investigation that examines the "*role*" of the definition (Ms-121, 59v.2–59v.3), which requires asking "a hundred more questions" to resolve or dis-

not just on the fact that the change in complexity implied by the change of notation is to be taken into account, but also on the fact that the introduction of a new notation usually means more than the introduction of a method of abbreviation. Wittgenstein was pointing out against Russell that the positional notation is not dependent on the stroke notation but has a life of its own.

38 Even though Wittgenstein does not explicitly use the word "primitive" in this context, the connection between Russell's unsurveyable logic and mathematics of primitive tribes is not as far-fetched as it might at first seem: That Wittgenstein compares another foundational system, set theory, with a "primitive way of thinking" (Ms-121, 29v.2) was already discussed above (see Section 1.2). Additionally, there is a close connection between Wittgenstein's idea of a "surveyable representation" / "perspicuous representation" ("Übersichtliche Darstellung") and his *Remarks on Frazer's Golden Bough* (see Majetschak, 2012). In the same way that Frazer's supposedly more 'civilised' "English parson" (Ts-211, 316.7 / *GB* p. 125) actually has a much more 'primitive' conception of the rituals of 'primitive' people than these people themselves, the supposedly more advanced logical calculi are actually much more primitive than their less formalised counterparts. For a more detailed exploration of 'primitive' forms of mathematics and the connection with Frazer, Spengler and "Übersicht" see Brusotti, 2014, pp. 28, 42–57.

solve a single one (Ms-121, 59v.4). Once we adopt the view that a definition does not abbreviate, but instead introduces a new calculus, it becomes obvious that a 'foundational' calculus can only be a very small and specialised part of mathematics, not the foundation for all of mathematics, which is only a secondary abbreviation for it:

> (Die Mathematik ist aber nicht symbolische Logik; sondern diese ein kleiner Teil der Mathematik. Der Teil, der, durch ein Mißverständnis, (die) 'Grundlage der Mathematik' zu sein schien.) [Ms-121, 60r.1]
>
> ---
>
> (Mathematics, however, is not symbolic logic; but rather the latter a small part of mathematics. The part which, through a misunderstanding, appeared to be (the) 'foundation of mathematics').

Incidentally, this remark as well as Ms-121, 59v.4 are nearly identical to two remarks in the fragment Ms-178d, consisting of 6 loose sheets.[39] There, Wittgenstein discusses self-referencing propositions and the role of contradictions in logic, with the link to Ms-121 not only established through the textual similarity of the remarks but also through the idea that in the case of 'abbreviating' definitions and self-referencing paradoxes, were are not *forced* to choose a particular path (Ms-121, 56v.2), contrary to what mathematical orthodoxy would suggest (see Section 3.4 for a discussion of other remarks in Ms-178d). Additionally, Ms-178d, 5.1 mentions the sleight of hand that is part of a "Taschenspielerkunststück" (see Section 1.2), echoing the "boastful" interpretation of Cantor's proof.

Although Wittgenstein's investigation of Cantor in Ms-121 can at times appear as a meandering exploration of unrelated topics, these excursions often turn out to be part of a larger investigation treating Cantor's diagonal argument not as its main object of study, but merely as a symptomatic case of conceptual confusion caused by a lack of "surveyability", resulting in conceptual confusions and the inability to clearly see the role of definitions. In the case of Cantor's diagonal argument, the "u.s.f. ad inf." is one of these seemingly innocuous definitions, but for other uses of the diagonal method the philosophical investigation has to branch off in other directions.

1.5 A GENERAL FORM OF COMPARISON

On 25 December 1938, three months after the preceding remark in Ms-121 and nearly half a year after his last explicit remark on Cantor's

---

39 Ms-178d is dated as "1.1.1940? – 31.12.1940?" in the Wittgenstein Archives Bergen *Nachlass* metadata. This dating is questionable, as the remark Ms-178d, 4.2 contains more variants than the one in Ms-121 (with only the last variant of Ms-178d appearing in the version of Ms-121) and is much more 'draft-like' than the entry in the latter notebook. The same is true for the corresponding remarks Ms-178d, 4.1 and Ms-121, 59v.4, with the last variants from Ms-178d used in Ms-121 and the version in Ms-121 additionally containing the insertion "noch" above "hundert", which is not present in Ms-178d, suggesting that Ms-178d was in fact written *before* Ms-121, 59v.4, in other words before 5 September 1938 and not in 1940.

diagonal argument, Wittgenstein returns to questions loosely related to Cantor's diagonal argument and its treatment of the infinite. The topic now occupying Wittgenstein is the proposition that fractions cannot be ordered according to their size. Of course they can be ordered in other ways and thus brought into a 1:1 correspondence with the natural numbers (in other words they are countably infinite), but in contrast to the natural numbers the fractions cannot be countably ordered according to size, since between any two rational numbers there are always infinitely more rational numbers. It is this talk of 'infinitely many' rational numbers 'that lie between' any two rational numbers that interests Wittgenstein, as it seems to hint at one of the "mysteries of the mathematical world" (Ms-121, 60r.2 / §40).

That Wittgenstein sees this 'mysterious' infinity as being related to the infinity of the cardinal numbers (and thus his earlier remarks with more explicit connections to Cantor's diagonal method) is made evident in a remark not published in *RFM II*, Ms-121, 60v.4, where he revisits the infinite series of cardinal numbers and the astonishment that their infinity brings with it. Upon hearing that there is no greatest cardinal number, that their series is endless, we picture this infinite series as something that is "monstrously long", even "more than monstrously long". But as soon as we change the aspect and speak of a "technique" without a designated end, the astonishment disappears:

> Daß dagegen die Technik des Bildens von Kardinalzahlen (etwa durch Addition von 1) kein Ende hat, daß in ihr kein Ende vorgesehen ist, {ist ein sehr leicht {verständlicher Satz // verständliches Sätzchen} & nichts daran, worüber wir staunen würden. // ist ein ganz einfaches & leicht verständliches Sätzchen.} Niemand wäre versucht die Technik des Zählens oder des Multiplizierens im unbegrenzten Zahlenraum eine "unendlich lange Technik" zu nennen. [Ms-121, 60v.4]

> That, on the other hand, the technique of forming cardinal numbers (for instance by adding 1) has no end, that in it no end is envisaged, {is a very easily understandable sentence & there is nothing about it that would amaze us. // is a very simple & easily understandable sentence.} No one would be tempted to call the technique of counting or multiplying in the unlimited number space an "infinitely long technique".

Similarly, the proposition that fractions cannot be ordered according to their size conjures up an image of an endless series of trees with new trees shooting up between every two trees and so forth without end, an image that "can make our head spin" (Ms-121, 61v.1 / §42). The astonishment brought up by such a picture disappears once we change the "(re-)presentation of facts" ("Darstellung des Sachverhalts") / search for a new "(form of) (re-)presentation" ("Darstellung(sweise)") / fall back on the "((re-)presentation) of the technique of calculating fractions" ("(Darstellung) der Technik")[40] That there is no

---

40 The version in *RFM II* omits all these variants of "Darstellung". Perhaps not too much should be read into these variants, but Wittgenstein's struggle to find new

"next greatest fraction" is entirely unsurprising once we realise that we teach a "technique of continuous interpolation" as part of our technique of calculating fractions and would thus never *"want to call"* any fraction the "next greatest" (Ms-121, 63r.2–63r.3 / §§43–44).

The next remark is a central remark that should be read together with §26 (see footnote in Section 1.2) and summarises what Wittgenstein sees as the problematic interpretation of the diagonal method:

> Von einer Technik zu sagen, sie sei unbegrenzt, heißt *nicht*, sie laufe ohne aufzuhören weiter – *wachse* ins Ungemessene; sondern, es fehle ihr die Institution des Endes, sie sei nicht abgeschlossen. Wie man von einem Satz sagen {könnte // kann}, es mangle ihm der Abschluß, wenn der Schlußpunkt fehlt oder von einem Spielfeld es sei {nicht begrenzt // unbegrenzt}, {wenn ihm die Regeln des Spiels keine gezogene Grenze vorschreiben. // wenn die Spielregeln keine Begrenzung – etwa durch einen Strich – vorschreiben.} [Ms-121, 63r.4 / *BGM II* §45]

> To say that a technique is unlimited does *not* mean that it goes on without ever stopping – that it increases immeasurably; but that it lacks the institution of the end, that it is not finished off. As one may say of a sentence that it is not finished off if it has no period. Or of a playing-field that is unlimited, when the rules of the game do not prescribe any boundaries – say by means of a line. [*RFM II* §45]

We can easily picture a large number of activities without an "institution of the end", but we are not astonished by their incompleteness, nor do we call these activities "infinite". That certain *techniques* give us the ability to go on indefinitely is hardly surprising, but we would usually not describe them in terms of an infinite series. It is clear that this critique applies to much more than just the restricted example of the ordering of fractions, it is instead a more general examination of our interpretation of the diagonal method. This method is a technique that can be used to indefinitely produce diagonalised elements which we then call different from all the elements in the system.

The misleading aspect of the diagonal method as producing an infinite series instead of as a technique without end (infinitely vs. indefinitely) is the tendency of the former interpretation to suggest an analogy where none exists. It is a *"new form of expression"*, which we however try to describe "by means of the old expressions" (Ms-121, 63v.2 / §46).

If we say "A fraction has not a next biggest fraction but a cardinal number has a next biggest cardinal number" (Ms-121, 64r.2 / §47), we tend to think that there *is something* analogous to the next biggest cardinal number even in the system of fractions, only that it does not exist (but that we have a clear picture of what it is that cannot exist). A better picture, however, is a comparison of two different games: There is nothing in our game of fractions that we *call* "the next biggest fraction" (Ms-121, 64r.3 / §48), we have not included

---

forms of representation that leave all facts as they are while resolving our conceptual confusion suggests a certain affinity with his conception of philosophy as clarifying through a "surveyable representation" (*PI* §122, cf. also *PI* §92).

any rule in this system that we could call analogous to the system of cardinal numbers.

We might be tempted to think that there is a general notion of "next biggest element" in a series, a notion that is independent of any particular system, so that we could then say "there is no next biggest element in the system of fractions" and "there is a next biggest element in the system of cardinal numbers". These two propositions might indeed be used in certain contexts, but what Wittgenstein seems to be arguing for is twofold: First, the use of these propositions is much more restricted than it first appears, because they are not general mathematical observations, as we might be inclined to believe based on Cantor's diagonal argument, but only useful as propositions after we have introduced a "new *concept*" that allows us to compare and distinguish different types of games (Ms-121, 64v.3 / §50). Second, there is no general, system-independent 'vantage point' from which we could gain an understanding of a general concept such as "next biggest element", which becomes clear once we consider the piece of the King in chess: We can certainly say "In draughts there isn't a King", but the King as a concept exists only in the context of chess. In the context of draughts we have no idea what could be meant by "King", we can only say that there is nothing that we could *call* a King in draughts. This distinction is quite delicate and hard to clothe into words. As Wittgenstein remarks in another context: "These things are finer spun than crude hands have any inkling of." (First appearance in 1939 in Ms-162b, 25v.2, also appears in 1944 in Ms-127, 110.3 and Ms-124, 161.2, which is published as *RFM VII* §57). This fine distinction is beautifully expressed in an unpublished remark that follows shortly after §50:

> 'Wenn Einer Dich fragt: "welches ist der nächst größere Bruch?", antworte ihm: "so etwas gibt's nicht"! (N.B. "So etwas gibt's nicht" – nicht: "es gibt keinen nächst größeren Bruch".) [Ms-121, 65v.3]
>
> ---
>
> If someone asks you: "what is the next biggest fraction?", answer him: "there is no such thing"! (N.B. "There is no such thing" - not: "there is no next biggest fraction").

It is unfortunate that the editors of the *RFM* chose not to publish the above or any of the other 10 remarks between §50 and §51, because there Wittgenstein repeatedly imagines practical uses for the expression "the next biggest fraction", mostly in the context of teaching someone how to calculate with fractions (and interpolate between them). The remarks are a beautiful example of Wittgenstein's way of philosophising: Exactly because the distinction between "there is no such thing" and "there is no next biggest fraction" is so delicate, it is important to look at the practical grammar of these expressions. Otherwise, this distinction risks being blurred, or worse, collapses completely under the pressure of generality.

Saying that "there is no next biggest number" does not mean that there is such a thing in general, but that it does not exist in this particular game (at least not in the absence of an extra concept that allows us to compare different systems with each other, which something like the diagonal method does not give us on its own). It says that there is no such concept *in the context of this game*, nothing that we would call the "next biggest number", "here, in this game":

> 'In dieser Technik gibt es also keine Verwendung für den Ausdruck {des 'nächst größeren Reihengliedes' // der 'nächst größeren Zahl'} oder: 'Was wolltest Du hier die 'nächstgrößere Zahl' nennen?' Wir werden sagen: es gibt hier keine. Hier, in diesem Spiel. [Ms-121, 66v.3]
>
> ---
>
> 'So in this technique there is no use for the expression {'next biggest series member' // 'next biggest number'} or: 'What did you want to call the 'next biggest number' here?' We will say: there is none here. Here, in this game.

It is this "here" that Cantor's diagonal argument obscures by seemingly being able to say something about any possible form of game, even games that have not been invented. But the use of a proposition such as "there is no next biggest fraction" is to show us where an analogy breaks down between different systems, not to test whether the propositions holds in two different systems, which would emphasise their similarity and comparability. In the context of fractions, the proposition "there is a next biggest fraction" is not wrong, but simply nonsense. As a consequence, "what is the next biggest fraction" is already nonsensical as a question. It cannot be answered by saying "there is no next biggest fraction", unless we mean it in the deflationary sense of "there is nothing that we would *call* the next biggest fraction":

> 'Frag also nicht, durch die Analogie mit den Kardinalzahlen {verführt // verleitet}: "welches ist der nächstgrößere Bruch"!' Dies hat offenbar Sinn. [Ms-121, 67r.1]
>
> ---
>
> 'So don't ask, {seduced // misled} by the analogy with the cardinal numbers: "which is the next biggest fraction"!' This apparently makes sense.

This makes it clear why it is misleading to treat these propositions as a "fact of nature" (Ms-117, 108.3 / §19). The existence or inexistence of a "next biggest fraction" is not a fact in the mathematical realm, eternally 'true' and independent from the rest of the system of fractions, but only meaningful against the backdrop of this technique. It is not a "fact of nature" for fractions, only a fact of the "nature of their use":

> 'Die Brüche lassen sich nicht ihrer Größe nach in eine Reihe ordnen' – {aber nicht ihrer Natur nach, sondern den Regeln nach, & der Natur ihrer Verwendung {nach // gemäß}. // {Aber // aber} es liegt nicht in ihrer Natur, sondern in den Regeln & in der Natur ihrer Verwendung.} [Ms-121, 67r.2]
>
> ---
>
> 'Fractions cannot be arranged in a series according to their size' - {but not according to their nature, rather according to the rules, & according to the nature of their use. // But it is not due to their nature, rather due to the rules & the nature of their use.}

But even if we accept this reasoning and see an expression such as "the next biggest number" as a case that depends on the special meaning given to it inside the system of fractions, what about the fact that there is a way to order all fractions in an infinite series, just not according to their size? Once we realise that we can order fractions by seeing them as pairs of numbers arranged in a table (with the fraction $\frac{1}{2}$ as the cell in the first column and the second row), we can order them by walking through the table in an ever increasing triangle: $\frac{1}{1}$, $\frac{1}{2}$, $\frac{2}{1}$, $\frac{1}{3}$, $\frac{2}{2}$, $\frac{3}{1}$, etc. Have we then not learned a general technique that we can apply to a whole range of systems? Does it not make sense to ask the very general question which systems can be ordered in such a way, with the rational numbers being a system that can be ordered, while for the real numbers this is impossible (as a consequence of Cantor's diagonal argument)?

Wittgenstein considers these questions in §§51–57 (Ms-121, 67r.3–69v.2). Someone who learns about a way to order fractions in the way describe above unquestionably learns something, for example to give each fraction a number (in the series) and to determine the number of a given fraction (Ms-121, 67r.3/ §51). The method described above is a general *technique* that can be applied not just to fractions, but to any pair of numbers and even more generally to any tuple of numbers. It seems as if we had not only learned a special technique in the context of fractions, but more generally about the *existence* of such a technique (Ms-121, 67v.2 / §52).

As a result, it can seem as if it must have sense to speak of the applicability of the technique itself, as a concept that stands on its own. Yes, we have learned that the concept is applicable to the system of fractions, but more generally we have learned that the application of this concept is possible. We might then think of "... can be ordered in an infinite series" as a general predicate that we can use to test any imaginable system, and as a consequence distinguish systems that can be ordered from those that cannot, as a systematic way of classifying systems.

To warn against such a view, Wittgenstein considers the simpler example of multiplication. It would be strange to say that by teaching someone to multiply this person has not only learned the technique of multiplication, but also that "it is possible to multiply". It is possible to use the concept of multiplication more generally and apply it to different systems and even explain one in terms of each other, such as when we teach the multiplication of fractions in analogy to the multiplication of natural numbers. (Ms-121, 68r.2 / §54) But even this is misleading, because it is not a *general* application of a general concept (which would belong to the "system of all systems" that Wittgenstein mentioned earlier in Ms-121), but only an *analogous* application that is justified by and dependent on a concrete concept of comparison between these two systems.

We might be tempted to think that this ‘defect’ of the concept “... can be ordered in an infinite series” is a fault of our imprecise language, that we could create and apply a truly general concept if we only had a more general language in which to express the concept. As Wittgenstein says, our ordinary expressions are merely an “approximate description of the technique one is teaching, say a not unsuitable *title*, a heading to this chapter” (Ms-121, 68v.1 / §55). But the view that we could go beyond this merely superficial way of describing a technique to unambiguously capture the essence of a concept, by employing a more rigorous formalisation along the lines of Frege and Russell, is exactly what Wittgenstein wants to deny. Yes, it is possible to formalise and prove a proposition such as “multiplying is possible”, but it does not get us any closer to the ‘essence’ of the concept than the ‘imprecise’ picture we had before: The use of such a proposition is not any clearer than before and it never will be, if the use of the ‘imprecise’ picture was restricted to being a “memorable picture” in our process of teaching (Ms-121, 68v.2 / §56).

The right way to avoid and clarify philosophical confusions is not to formalise the imprecise picture until all the impreciseness has been extinguished, but rather to see the picture as a picture and focus on the “interest of this calculation in its application” (Ms-121, 69v.2 / §57), ignoring the chapter “titles”.[41]

In the next 7 remarks, none of which were published in *RFM II*, Wittgenstein returns explicitly to Cantor’s diagonal argument. In the first of these remarks, he once again considers the difference between saying that a number *is* different from others and saying that we *call* a number different from others. In the first case, the resulting number is embedded in a calculus, the proposition that the number is different thus has a *use*. But in the second case, the resulting number (such as the result of the diagonalisation) is not used in any calculus (Ms-121, 69v.3). In other words, only in the second case are we free to say that we *call* it a number because we could also imagine the alternative of not calling it a number, whereas in the first case we say that it *is* a number because we would otherwise have to give up a multitude of interconnected language games that use such a number.

The next remark continues this line of thought. Wittgenstein readily admits that we learn a new rule for constructing a decimal expansion, but what does Cantor’s proof actually *prove*? We could imagine the diagonal method to be the answer to a riddle, as “mathematical fireworks”. But does it not have to be more than that? Does it not *force* us to say that the diagonalised number *is* a number? As Wittgenstein points out, here we are thinking about the *finite* case:

> [...] Hast Du mir eine von allen diesen Zahlen verschiedene Zahl gezeigt? Du hast mir etwas gezeigt was ich ({vielleicht // etwa}) geneigt bin eine

41 This explicit reference to §55 is evident in the unpublished variants of Ms-121, 69v.2: “Trenne Dich [...] von [...] {diesen Phantasien // phantasieanregenden Titeln // von diesen Titeln}”

> {solche // von allen diesen verschiedene neue} Zahl zu *nennen*. [...] Ich möchte das rechtfertigen indem ich sage: Es ist eben hier *alles* anders, ich bin nicht mehr – wie im {andern // endlichen} Fall – *gezwungen* dies so zu nennen. Aber hier ist doch nur ein *Gradunterschied*! Du {könntest doch auch im andern Fall sagen, Du seist nicht gezwungen // kannst ja eben von jedem neuen Fall sagen hier gelte die alte Regel nicht mehr}. *Jeden* mathematischen Unterschied kannst Du Unterschied der *Art* nennen!
>
> Du kannst überall (oder nirgends) eine scharfe Biegung sehen. Gewiß; aber auf diesen Gradunterschied muß man aufmerksam {machen // sein}. Denn {durch diese Gradunterschiede // über diese Stufenleiter // auf dieser Stufenleiter} geht, was jeder einen Beweis nennt, in etwas über, was niemand mehr einen Beweis nennen würde. Wenn Du Dir des Unterschieds bewußt wirst, redest Du nun noch so wie früher? [Ms-121, 70r.2]

> [...] Have you shown me a number different from all these numbers? You have shown me something which I am perhaps inclined to call {such a number // a different number from all of these}. [...] I would like to justify this by saying: Here *everything* is different, I am no longer – as in the {other // finite} case – *forced* to call it that. But here we have only a *difference of degree*! {You could also say in the other case that you are not forced // can just say of every new case that the old rule no longer applies here}. You can call *every* mathematical difference a difference of *kind*!
>
> You can see a sharp bend everywhere (or nowhere). Certainly; but one must draw attention to this difference of degree. Because {through these differences of degree // over this ladder // on this ladder} what everybody calls a proof turns into something which nobody would call a proof anymore. When you become aware of the difference, do you still speak as before?

The above remark acts as a focal point of the previous reflections in Ms-121 and explains quite concisely why Wittgenstein wants to distinguish between the finite and the infinite case: In the finite case, the result of the diagonalisation is a number with a finite number of decimal places, in other words a *fraction*. This is why we say that it *is* a number, the alternative would be to give up our whole concept of fractions as number and all the language games that are linked to it. But in the infinite case "everything is different", we are not *"forced"* to call the diagonalised number a number, because it can be clearly distinguished from what we usually call a number: As Wittgenstein mentioned before, the diagonalised number is "essentially" a decimal expansion, whereas for our 'usual' numbers a decimal expansion is only an "inessential" aspect.

But then why not object that this line of reasoning applies to any step of following a rule, not just the step from finite to infinite case in Cantor's diagonal argument? After all, nothing *forces* us to follow rules the way we do. Wittgenstein seems to sidestep the issue slightly: Just because we *can* question every step in a mathematical argument does not mean that it is *useful* or of interest to question it. Wittgenstein's investigation of Cantor's diagonal argument is not a trivial form of scepticism, but a very targeted examination of a few central points in the proof. Whether we see the proof in a new light and consider some of these differences to be more than gradual is up to us. This will decide whether we change our manner of speaking or not, but it is not Wittgenstein's aim to point out 'defects' in the proof, because there is obviously nothing wrong with it, only to show us differences in use that could perhaps allow us to see otherwise neglected aspects.

By pointing out that the infinite diagonalised number is not embedded in a calculus in the same way as in the finite case, Wittgenstein has already mostly dealt with the objection. Cantor's diagonal argument cannot "force" us to call the infinite diagonalised number a number, but not due to some trivial scepticism, rather because the diagonalised number does not have the same 'standing' in our language as other more traditional numbers: In the words of the later writings, the diagonalised number is not as fixed, it is not a "hinge" in the same way as other numbers are, because it lacks the surrounding language games that restrict its degrees of freedom.[42] To deny that something like 2 or 12.5 is a number would be to give up a whole form of life, but to deny that the diagonalised number is a number is to give up only a comparatively small number of mathematical language games.[43]

Cantor thus leaves us with a choice: *If* we call diagonalisation a way to construct a new real number, *then* we will give up speaking of a system of all real numbers (Ms-121, 71r.2). But exactly because it is a choice, it is "highly misleading" to say that the real numbers "cannot" be ordered in a series, in analogy to the contrary proposition for rational numbers. This would suggest that the proposition were applicable to the real numbers but simply has a different truth value than in the case of rational numbers. Instead, our choice has made the proposition inapplicable, except as a non-mathematical "chapter title", a "mock facade".

It is these non-mathematical sentences, these mock facades that Wittgenstein wants to attack, or rather, not so much the use of mock facades in mathematics, but cases in which we mistake these mock facades for facades with mathematical substance behind them (Ms-121, 71v.2–71v.3). In these cases we risk using mock facades *instead* of mathematical calculi, we think of them as mathematics, when they are really only "prose" (which is the term that Wittgenstein employs

42 Cf. Ramharter, 2019 and the distinction between a mathematical proposition such as "$25 \times 25 = 625$" on the one hand, which is a grammatical proposition and a hinge proposition, and a mathematical proposition such as "There is no greatest cardinal number" on the other hand, which is also a grammatical proposition but not a hinge propositions.

43 To be fair, even this 'sacrifice' is already quite substantial and it is understandable that most mathematicians do not want to follow Wittgenstein here. But it should be pointed out that Wittgenstein is not advocating for a denial of Cantorian infinity, he merely wants to show that these concepts are much less firmly embedded in our way of life than Cantor's diagonal argument suggests and that we *could imagine* a different form of life, where Cantor's proof is viewed in a completely different light. It should also be kept in mind that Wittgenstein's critique of Cantor is not a critique of Cantor alone, but includes the *applications* of the diagonal method, for example by Gödel, in the 'foundations of mathematics', a field that Wittgenstein regards with suspicion.

in Ms-127, 185.2 / *RFM V* §46 and Ms-124, 138.3 / *RFM VII* §41, both from 1944). Wittgenstein gives an example of such a mock facade:

> Zu sagen "man kann sie nicht in ein System ordnen, weil ihrer mehr sind als in einem System Platz haben" ist greulicher Unsinn.
>
> Die Frage ist ja doch: wer sind die *sie* die ich nicht in ein System ordnen kann? Ist es denn nicht so daß mir der Cantorsche Beweis einen andern Sinn von "sie" zeigt? Wir haben hier eine andre Art von Begriff, eine neue Verwendungsart für ein Begriffswort. [Ms-121, 72r.1–72r.2]
>
> ---
>
> To say "you can't put them in a system because there are more of them than there is room for in a system" is atrocious nonsense.
>
> The question is: who are the *they* that I cannot arrange in a system? Is it not the case that Cantor's proof shows me another sense of "they"? We have here a different kind of concept, a new mode of use for a conceptual word.

Cantor's diagonal argument does not show us a number in the familiar sense that is different from all the others, the proof does not *discover* a concrete number. There are no "they" / "them", there is no collection of numbers like 2, 12.5 or π that resist our attempt to order them. Instead, the diagonal argument *invents* a new concept of "they" / "them", a new use of "they" / "them", not a collection of elements that are simply too numerous to be ordered.

Wittgenstein then separates the next remark with a horizontal line, what follows afterwards mostly revolves around contradictions in logic and Gödel's incompleteness theorem.[44] These remarks diverge too far from Wittgenstein's direct discussion of Cantor's diagonal argument to be discussed here at length, but the separating line should not be interpreted as the beginning of an entirely new topic: As the unpublished remarks after §57 have shown, Wittgenstein wraps up his line of thought by going back to points that were first brought up earlier in Ms-121 and Ms-117, albeit with more clarity than in his first attempts. He can now investigate another aspect of the diagonal method: its *application* in Gödel's incompleteness theorem. The discussion of contradictions and Gödel is not unrelated to the rest of Ms-121, merely important enough to deserve its own investigation. The remarks on Russell and Gödel in Ms-121 do not stand on their own, but are symptomatic cases of "mock facades" that appear as solid buildings, solid enough, in fact, to act as foundations of mathematics. It is understandable that the editors of *RFM II* chose to focus on the more explicit remarks on Cantor's diagonal argument, but a faithful reading of Wittgenstein must try to take into account the

44 While these remarks seem to treat a distinct issue, it would be mistake to read Wittgenstein's remarks 'on' contradictions in different parts of the *Nachlass* as a sustained and uniform treatment of a single topic. Instead, Wittgenstein's discussions of contradictions are as varied as the use (and distrust) of contradictions in mathematics and logic. Some of the different aspects of these remarks on contradictions are discussed in Ramharter, 2010, with Cantor's diagonal argument as the "guideline".

motivation for his engagement with this particular proof or risk misunderstanding his remarks as an overreaching attack of sound mathematical practice.

Apart from the general observation that Gödel's incompleteness theorem is a limitative result demonstrating the logical impossibility of a (consistent) "system of all systems" (see Section 1.2), there are a number of more direct links with explicit remarks on Cantor that can serve as a supporting argument for reading Wittgenstein's remarks on Cantor in connection with his interest in Gödel and Russell's *Principia Mathematica*. For example, the third remark after the separating line reflects on the view of propositions in *Principia Mathematica* as a *list*, which links Wittgenstein's discussion of contradictions, inconsistency and completeness with questions of enumerability and countability:[45]

> Man könnte die Principia Mathematica auffassen, nicht als fortlaufende Mitteilung, sondern als *Liste*, als Katalog, von Sätzen gewisser Form (mit beigefügten Analysen dieser Formen). [Ms-121, 72v.2]
>
> ---
>
> One could regard the Principia Mathematica not as a continuous message, but as a *list*, a catalogue, of propositions of a certain form (with attached analyses of these forms).

Furthermore, the following remarks make it clear that Wittgenstein is not interested in contradictions *in general*, but wants to focus on very specific occurrences of contradictions in situations like those produced by the diagonal method (where the assumption that all elements lie inside the system leads to a contradiction with the diagonalised element, that lies outside the system). It is only in these very specific situations that a contradiction can potentially be entirely harmless, because there is no (practical, extra-mathematical) use for contradictions produced through diagonalisation:

> Wenn Einer {dort einen Widerspruch // einen Widerspruch dort} findet, oder erzeugt, wo für die satzartigen {Gebilde // Zeichenverbindungen} die {einander widersprechen // den Widerspruch bilden}, keinerlei Verwendung vorgesehen ist, dann ist gegen diesen Widerspruch vorerst nichts einzuwenden. [Ms-121, 73v.2]
>
> ---
>
> If one finds, or creates, a contradiction where no use is intended for the proposition-like {constructions // sign combinations} which {contradict each other // form the contradiction}, then there is nothing to object to this contradiction for the time being.

Wittgenstein immediately makes it clear that he is not advocating for logical trivialism (so that a single contradiction in a system would entail every imaginable proposition), but rather suggests a more nuanced, even *paraconsistent* view for these specific situations:

---

45 An even more explicit connection in that regard are the two remarks Ms-162a, 68.2–69.2, which were written in January 1939, shortly after the last remarks in Ms-121 and should be seen as a continuation of that document. In Ms-162a, Wittgenstein discusses diagonalisation and then remarks: "Dann sind aber auch die Sätze der Arithmetik nicht abzählbar. Dagegen sind aber die in Russells System beweisbaren Sätze abzählbar."

> "Aber aus einem Widerspruch folgt ja *jeder* Satz! Was würde dann aus der Logik?"
> Nun so folgere nichts aus einem Widerspruch! [Ms-121, 74r.3]

> "But from a contradiction follows *any* proposition! What would then become of logic?"
> Well, then don't infer anything from a contradiction!

As Wittgenstein made clear in the context of Cantor's diagonal argument, we have a certain leeway in calling or not calling the diagonalised result a number, because its use is not as 'fixed' as in the case of the natural numbers that are embedded in a multitude of language games and hold a prominent position in our forms of life. The same applies to contradictions that are similarly 'artificial':

> Erinnere Dich *hier* Deiner Freiheit, möchte ich sagen, zu gehen, wie Du willst.
>
> Und heißt das nicht: Verstehe, was Dich sonst gebunden hat & daß Du also hier frei bist? [Ms-121, 75r.3–75r.4]

> Remember *here* your freedom, I want to say, to go as you will.
>
> And doesn't that mean: Understand what has otherwise bound you & that you are therefore free here?

We are free to choose how to proceed, for two reasons (which are closely linked in the case of diagonalised contradictions): First, the contradiction gives us two options, p and $\neg p$. Second, we are free to choose which option to take because the contradiction itself has no use, it is not embedded in our way of life. The second reason is crucial and applies only because the contradiction does not have "any kind of work to accomplish":

> Nicht {das // dies} ist {ein Unglück // perniziös}: einen Widerspruch zu erzeugen {in der Region, in der // dort, wo} weder der widerspruchsfreie noch der widerspruchsvolle Satz {eine // irgend welche} Arbeit zu leisten hat; wohl aber das: nicht zu wissen, {wo man in diese Region eingetreten ist // wie man dorthin gekommen ist // wo man in diese Region gekommen ist wo der Widerspruch nicht mehr schadet}. [Ms-121, 74v.2 / *BGM IV* §60]

> The pernicious thing is not: to produce a contradiction in the region which neither the consistent nor the contradictory proposition has any kind of work to accomplish; no, what *is* pernicious is: not to know how one reached the place where contradiction no longer does any harm. [*RFM IV* §60]

Lastly, the passage on contradictions and Gödel also revisits Wittgenstein's earlier remarks on inductive proof (see Section 1.2, especially the remarks Ms-121, 31v.2 and Ms-121, 32r.3):

> Nennen wir die Russellschen Beweise 'Konstruktionen von Sätzen' – was ist aber dann ein Induktionsbeweis? Er kann doch *als Konstruktion* nicht mit den andern verglichen werden. [Ms-121, 77v.2]

> If we call Russellian proofs 'constructions of propositions' - what then is a proof by induction? After all, it cannot be compared with the others *as a construction*.

Wittgenstein continues his remarks on contradictions and Gödel for 10 more pages, before returning more directly to Cantor in Ms-121, 86r.1, without separating this 'return' by a line or even implicitly through a chronological break, however.

1.6 BEYOND A SYSTEM OF OPERATIONS

Wittgenstein begins his return to an explicit discussion of Cantor's diagonal argument with an interpretation of Cantor's proof that sidesteps any misleading talk of uncountable sets being more numerous or in some way bigger than the infinity of the natural numbers. Instead, we could say that the proof shows us "that one has no conception of a system of infinite decimal fractions", contrary to what the "similarity of their notation with that of the cardinal numbers" might suggest (Ms-121, 86r.1). Such an interpretation emphasises the *difference* between the two cases: the systematic case of the cardinal numbers stands thus in stark contrast to the inherently unsystematic case of the real numbers, instead of talking about two supposedly systematic cases by comparing their 'number of numbers'. In the case of the real numbers, the "number of all real numbers" is only a "metaphor", which does not mean that the metaphor were wrong or worthless, only that the use of this picture is yet to be established by the calculus standing behind it (Ms-121, 86r.2). It matters not only that such a picture *can* be used, but *how*, as Wittgenstein explains a few remarks later:

> Laß {uns // mich} hinter die Kulissen dieser Definition schauen! (Ich will mich dann ruhig wieder in den Zuschauerraum setzen.) Die Frage scheint irrelevant – aber warst Du wirklich ganz ahnungslos, als Du sie gabst, hast Du sie nicht im Hinblick auf eine bestimmte Anwendung gegeben? Nun, es macht ja nichts, wenn es so ist. Nur *schillert* ({Frege // Freges Ausdruck}) Deine Definition: man kann sie einmal als unangreifbare, weil willkürliche, Festsetzung der Bezeichnung verstehen & {zugleich aber wieder // dann wieder} als *Satz* über die Natur der Zahlen.
>
> "Aber was kann man mehr von einer Konvention des Ausdrucks wollen, als daß sie sich hinterher als äußerst brauchbar erweist?!"
>
> Aber {da // hier} ist es eben schwer, {daß man sich & dem Andern kein x für ein u vormacht // sich & dem Andern kein x für ein u vorzumachen}: {denn ist {sie // die Definition} nun brauchbar, {*weil* // *indem*} sie unsrer Phantasie Nahrung gibt, oder in anderer Weise? // denn besteht nun die Brauchbarkeit dieser Definition darin, daß sie unserer Phantasie durch das Bild, {was // welches} sie einführt, allerlei Nahrung gibt; oder besteht sie in etwas anderm?} [Ms-121, 86v.3]
>
> ---
>
> Let {us // me} look behind the scenes of this definition! (I will then sit quietly back down in the audience again.) The question seems irrelevant - but were you really quite clueless when you gave it, did you not give it with a particular application in mind? Well, it doesn't have to matter if it happens to be so. Except that your definition *shimmers* ({Frege // Frege's expression}): one can understand it on one hand as an unassailable, because arbitrary, stipulation of the denotation & {but at the same time again // then again} as *proposition* about the nature of numbers.

> "But what more can one want from a convention of expression than that it should afterwards prove to be extremely serviceable!"
>
> But {there // here} it is just difficult that one does not lead oneself & the other up the garden path: {After all, is {it // the definition} useful *because* it gives food to our imagination, or in some other way? // For does the usefulness of this definition consist in the fact that it gives all kinds of food to our imagination by the image which it introduces; or does it consist in something else?}

As Wittgenstein mentions in the third paragraph, we could certainly call a definition useful even if it does nothing more than feed our imagination, in the same way that even a mock facade can be useful by being aesthetically pleasing. But as the first paragraph points out, the question that motivates a diagonal proof such as Cantor's (for example: "Can we sequentially number all the elements in the collection X?") matters, because it hints at the practical applications that gave rise to it and made us search for an answer. The definition that we adopt as a result of the diagonal proof (the "uncountable number of all real numbers", for example) seems to be entirely independent of our original motivation, however, as it depends solely on the mathematical proof. As a consequence, the definition appears to be either an entirely arbitrary stipulation (the formalist view) or a deep discovery about the nature of numbers (the platonist view).

Wittgenstein has made it repeatedly clear that these extreme viewpoints are a consequence of our failure to "look behind the scenes" (an image used again a few days later in Ms-162a, 87.1 and further explored in Ms-162a, 92.1), a result of our neglect of the real mathematical building behind the mock facade and our myopic focus on the prose instead of the mathematical substance. He now adds another picture, that of the suburb consisting entirely of mock doorways and mock windows:

> Ein Tor ist etwas durch das Haus, was dahinter steht, ein Fenster durch den Raum in den es Licht läßt. Denke Dir eine Stadt mit Häusern, Straßen & Gärten & eine ihrer Vorstädte bestünde aus Toren ohne Häusern, Fenstern in Mauern ohne Zimmer dahinter, Gartenzäune die keinen Garten umgeben, Gaslaternen, die mit keinem Gaswerk in Verbindung stehen. [Ms-121, 87v.2]

> A gate is something by virtue of the house that stands behind it, a window by virtue of the room into which it admits light. Think of a city with houses, streets & gardens & one of its suburbs would consist of gates without houses, windows in walls without rooms behind them, garden fences surrounding no garden, gas lamps not connected to any gasworks.

Not only is it a beautiful image, it also emphasises the parallels to other parts of the *Nachlass* (see the other occurrence of "Vorstädte" in Ms-142, 12.2, from where it is transferred to Ts-220, 10.2, Ts-239, 9.4 and finally Ts-227a, 15.3 / *PI* §18) and the close connection of the idea of mock facades and prose in mathematics with the idea of imagining a form of life.

The remarks that follow were all published in *RFM II* and form the concluding sections §§58–62. The first three of these deal with the use of the word "infinite" in mathematics and revisit ideas similar to the mock facades and the suburb. As Wittgenstein emphasises in Ms-121, 87v.3 / §58, the use of the word "infinite" in mathematics is not dangerous in and of itself, but it becomes misleading when it gives the calculus its meaning instead of vice versa.

That there is nothing infinite to be found 'inside' the calculus (Ms-121, 88v.2 / §60), as if we had peaked into a box and found nothing extraordinary inside, only finite rules, is of course the finitist view of mathematics. Wittgenstein does not want to advocate for finitism, which would amount to an interference in our language and the exclusion of a possibly useful concept. Instead, he wants to lead the investigation back to the "everyday employment of the word "infinite"", so that we look at how it is used "in connexion with these mathematical calculi": not in a mathematical vacuum, but embedded in our whole way of life.[46]

The last two remarks of *RFM II* are some of the most holistic and meta-philosophical in all of Wittgenstein's remarks on Cantor, it is thus understandable that they were chosen as the end of the published work. In the first remark (Ms-121, 89r.2 / §61), Wittgenstein draws a parallel between "finitism and behaviourism", which "both deny the existence of something" (the infinite and the inner, respectively), a remark that is mirrored in *LFM XII*, p. 111 ("Finitism and behaviourism are as alike as two eggs"). More interestingly, however, it shows a striking parallel to the very last remark of *Philosophy of Psychology – A Fragment* (previously published under the title *Philosophical Investigations – Part II*), §xiv (Ms-144, 40r.2, with a nearly identical variant in Ms-138, 12a.2):

46 Wittgenstein's emphasis on surveyable proofs (which become unsurveyable if they contain too many elements to be reproduced without errors) has sometimes been interpreted as an irrevocably finitist, even strictly finitist, position. But while it is true that Wittgenstein wants to deny that we could *directly* use a system such as Peano arithmetic to calculate with large numbers, this does not imply that we cannot have a concept of peano-unsurveyably large numbers or a concept of infinity that has mathematical sense, nor does it preclude the possibility of speaking of the infinity of the natural numbers without running into philosophical confusions. In all of those instances, the concepts are given a meaning in other languages games, our mistake is simply that we are misled by the analogy to Peano arithmetic in the small and finite cases, believing that the other language games could fall away and the concepts be reduced to Peano arithmetic (and other foundational systems). As Frascolla, 2006, p. 145 correctly points out while discussing surveyability in the context of Wittgenstein's perceived revisionism:

> Wittgenstein moves some criticisms against the platonistic interpretation of the true import of Cantor's proof; nevertheless they do not originate in any way from a presupposed identification of legitimate mathematics with finitist mathematics and, even less so, from the violation, by Cantor's proof, of the requirements imposed by strict finitism. Once the appropriate clarifications have been made about what, in his opinion, it really demonstrates, Cantor's proof is more than good enough for Wittgenstein, in spite of the certainly non-finite nature of the "objects" it deals with.

> Es ist für die Mathematik eine Untersuchung möglich ganz analog unsrer Untersuchung der Psychologie. Sie ist ebensowenig eine *mathematische*, wie die andre eine psychologische. In ihr wird *nicht* gerechnet, sie ist also, z.B., nicht Logistik. Sie könnte den Namen einer Untersuchung der 'Grundlagen der Mathematik' verdienen. [Ms-144, 40r.2 / *PPF* §xiv]

> An investigation entirely analogous to our investigation of psychology is possible also for mathematics. It is just as little a *mathematical* investigation as ours is a psychological one. It will *not* contain calculations, so it is not, for example, formal logic. It might deserve the name of an investigation of the 'foundations of mathematics'. [*PPF* §xiv]

As Wittgenstein mentions in *LFM XII*, the position of the "opponents" are equally absurd (although this way of phrasing it is rather dogmatic). In the case of psychology, it is the mentalist view that privileges the inner above all else, or more precisely the "pneumatic conception of thought" (Ts-227a, 84.2 / *PI* §109)[47]. In the case of mathematics, it is the platonist conception that views the uncountability of the real numbers as a *discovery* in the ideal realm of mathematics, Cantor's paradise of set theory (which Wittgenstein explicitly mentions in Ms-144, 39v.5 / *PPF* §xiv).

The parallels to his remark in *PPF* §xiv continue into the last remark of *RFM II* (Ms-121, 89r.3 / §62): Wittgenstein's philosophical investigation of Cantor's diagonal argument must not interfere in mathematical matters, it is not the task of the philosopher to "shew that calculations are wrong, but to subject the *interest* of calculations to a test", not even to say: "That is absurd" (in contrast to his more dogmatic statement in *LFM XII*). Instead, the aim of a philosophical investigation is to "survey" ("übersehen") the "justification of an expression" by clarifying its use.[48]

---

47 See Trächtler, 2021 for a detailed discussion of the "pneumatic conception" in Wittgenstein's later philosophy. Majetschak, 2020 provides an overview of the "misleading parallel between outer and inner" and also includes an examination of the role of "Übersicht" in Wittgensteins late writings on the philosophy of psychology, which is arguably one of the links between the seemingly disconnected fields of mathematics and psychology.

48 *Contra* Mühlhölzer, 2010, the use of "übersehen" in Ms-121, 89r.3 / §62 should be interpreted as being intimately linked with the earlier use of "übersehbar" in the context of mathematical proofs (Ms-121, 55v.3, see Section 1.4). The variety of English translations ("surveyable", "perspicuous", "synoptic" and others) often obscures that "Übersicht" and "übersichtliche Darstellung" are in fact some of the central terms in Wittgenstein's philosophy (Majetschak, 2016). In fact, the lack of "surveyability" caused by 'abbreviating' definitions is one of the core issues in Wittgenstein's philosophy of mathematics and explains why only a philosophical investigation, resulting in a surveyable representation, can clarify these situations: For a mathematician, the difference in surveyability between unary and decimal notation is negligible, a (surveyable) unary number with 3 strokes is just as much as number as one with a million. The only way to clarify these mathematical concepts, then, is to describe and compare the practical *techniques* in their specific language games, a task which falls to Wittgenstein and his *surveyable representation*. As a consequence, surveyability in mathematics and philosophy are not two separate concepts, but heavily interrelated, with the unsurveyability of mathematical logic and the 'founda-

Furthermore, the remark and its use of "survey" / "übersehen" highlight the fundamental importance of Wittgenstein's comparison of two different ways of counting in Ms-121, 91r.2 / §59 (the unary "| | | |" and the decimal "4"), which can only be understood in light of his development of the concept of "Übersicht" / "übersehen": Mathematically, all natural numbers can be defined as the repeated application of a successor "operation" (to use Wittgenstein's own term from the *Tractatus*) to the "base" 0, in other words as the repeated addition of a single stroke. The difference between the notations "S(S(S(S(0))))", "| | | |" and "4" is merely a difference of notation, easily bridged by suitable definitions (which is the central point of *TLP* 6.02). Such a view considers numbers as a single *system*, which produces its element through the repeated application of an operation. Cantor's diagonal method extends this viewpoint to *systems of numbers*, by giving us an operation which can be used to repeatedly produce elements that are not yet elements in the (assumed ordering of the) existing system. This repeated application of the diagonalisation operation allows us to 'escape' from any assumed ordering of the real numbers into an even greater infinity than that of the natural numbers, but it also allows us to build up a hierarchy of these ever greater infinities in the form of cardinal numbers (through the repeated construction of the diagonal set, see footnote in Section 1.2). This two-step procedure of first 'escaping' a system but then 'approximately capturing'[49] these escaped elements in an infinite hierarchy lies at the root of many other applications of the diagonal argument, with Cantor's own diagonal argument merely being the original example. It can more generally be used as a systematic method to produce elements that cannot be captured in the totality of a system and forms the nucleus of Gödel's incompleteness theorem, Turing's diagonal argument (see Section 3.1) and (less directly through analogy with the diagonal set) even Russell's paradox. While the diagonal method acknowledges that there is no possibility of ever finding a single system, in return it offers the possibility of a *systematic* series of systems (the systematic treatment of infinity as the cardinal numbers, for example). It systematises a series of systems and presents us with the "general form of the operation" (of moving from one system to the next), similar to what Wittgenstein attempted to do in the *Tractatus* (*TLP* 6.01), as he points out in Ms-121:

> Man könnte das auch so sagen: Es gibt nicht (wie ich in der Log. Phil. Abh. {gemeint habe // meinte}) eine 'allgemeine Form der Operation', die {eine Zahl in eine andere verwandelt // aus einer Kardinalzahl eine andre macht} – das wäre ein *System* der Operationen; [Ms-121, 90v.1]

---

tions of mathematics' as the impetus for Wittgenstein's philosophical aspiration for surveyability.

49 Compare Turing's "approximation of truth" in Section 3.6.

> One could also put it this way: There is no such thing as a 'general form of the operation' (as I thought in the *Tractatus Logico-Philosophicus*), which transforms one number into another – that would be a *system* of operations;

Mathematically (and philosophically from the perspective of the Wittgenstein of the *Tractatus*), the difference between "| | | |" and "4" is a negligible definition and numbers can be entirely explained by a finite successor operation, which is then applied ad infinitum ("u.s.f. ad inf."). In other words, in the same way that "| | | |" and "4" are equal in the finite case, the relation between picture and calculus holds even in the infinite case, thanks to a seemingly innocuous application of inductive inference.[50]

From the philosophical standpoint of the Wittgenstein of Ms-121, however, the picture changes radically if we consider the similarity of different notations regarding their *use*: "| | | |" and "4" are very similar, but "| | | | | | | | | | |" and "11" much less so, while even longer sequences of strokes and their decimal 'counterparts' may not be alike at all. This is because "| | | |" is surveyable, it can be taken in at a glance, exactly like "4". But while 11 and 1000 are still surveyable in decimal notation, their unary 'counterparts' are not, and even decimal notation itself stops being surveyable in the face of large numbers, which we can bring into a more surveyable form if we employ a notation such as "10^{10}".[51] The notations "| | | |" and "4" have a very similar use, but the difference in "Übersicht" concerning "| | | | | | | | | | |" and "11" has implications for their use: We cannot add numbers made up of thousands of strokes without making mistakes, while we can easily perform these additions if we use decimal notation. As a result, a philosophical investigation of their grammar must adequately describe these differences, however subtle they may be, while for a mathematician these notational differences do not matter and cannot impact the mathematical validity of a proof in any way, at most its practical feasibility. But for a philosopher, these notational differences matter, as they are linked to different techniques:

> {Du denkst // Man denkt}: alles was notwendig ist sind geeignete Definitionen. Und man vergißt, daß eine Definition in der Mathematik nicht bloß ein 'Aktuar zu' der Schreibweise ist, sondern die Einführung einer (*mehr* oder *weniger*) verwandten Technik des Rechnens. Wo aber steht geschrieben, *wie* ich Russells Technik durch andre Techniken fortsetzen soll? [Ms-121, 92v.3]

> {You think // one thinks}: all that is necessary are suitable definitions. And one forgets that a definition in mathematics is not merely an 'actuary to' the notation, but the introduction of a (*more* or *less*) related technique of calculation. Where however is it written *how* I should continue Russell's technique through other techniques?

50 See Ramharter, 2014, pp. 186–190 for a discussion of the relation between numerals and induction in the context of Wittgenstein's middle period, where his remarks on induction are more frequent but often show a revisionist bent.

51 Another example of such a shift from a less to a more surveyable notation is Knuth's "up-arrow notation" for numbers that are so large that even exponentiation stops being surveyable, see Knuth, 1976.

Philosophically, the different language games played with numbers cannot be reduced to a single general form, which could then be uniformly calculated in a formal system such as *Principia Mathematica*. Instead, different 'parts' of the real numbers each form their own system and stand in complex relationships with each other. Mathematically, we can of course play Cantor's diagonal game and view everything uniformly and extensionally, but philosophically, the game loses its point once we see that the envisaged uniformity fails to capture the variety of different language games.[52]

But even if we were to follow Wittgenstein in his emphasis of the "Übersicht", his critique might still seem overly pedantic. After all, does Cantor not show us something entirely compatible with this view, namely that the real numbers (or any other system considered under diagonalisation) lead to a variety of infinities, not just 'one' uniform infinity? Is Wittgenstein's argument for the plurality of systems not also an argument for the infinity of the cardinal numbers instead of a single 'general' infinity? If we only consider the diagonal argument as a piece of mathematics, Wittgenstein's critique can indeed appear rather shallow. But the crucial difference in interpretation is related to the conclusion that we draw from Cantor's diagonal argument, not the argument itself: Wittgenstein does not want to deny that we can 'escape' from a system through the 'loop hole' of diag-

52 Ramharter, 2018, pp. 138–39 notes the connection between surveyability and Cantor's diagonal argument and points out that for mathematicians, something being "merely a question of representation" ("nur eine Frage der Darstellung") is often taken to be "merely a question of didactics" ("nur eine Frage der Didaktik"), in other words unimportant to the core business of mathematics. Ramharter emphasises the importance of this question for Wittgenstein, but comes to the conclusion that Wittgenstein's critique is void if cardinal arithmetic and the diagonal argument are properly considered inside their mathematical surroundings with their connections to established practices. Compared to Wittgenstein's remarks on Russell's *Principia Mathematica*, his discussion of Cantor's diagonal method shows an "opposite turn" ("gegenteilige Wendung") regarding surveyability: arithmetic becomes surveyable only when Russell's logic is enriched with suitable *abbreviations* for numbers, whereas Cantor's diagonal argument becomes surveyable only when *expanded* with its surrounding mathematical practice. This interpretation suggests that there was nothing wrong with Cantor's diagonal argument to begin with and that Wittgenstein failed to do what he claims to do, namely to look at the *use* of this piece of mathematics. But such an interpretation ignores that Wittgenstein is interested in the *standing* of the diagonal method and its (ab-)use in our prose, not its intra-mathematical rigorous use. To show that Russell's logic depends (if not mathematically, at least philosophically and practically) on seemingly secondary definitions does not just make the system more surveyable for its own sake, it is meant to *break the spell* by showing how *"boastful"* (Ms-117, 109.3 / §21) it really is. This can only be achieved by a combination of (mathematical) surveyability as abbreviating definitions and (philosophical) surveyability as a description of our techniques of arithmetic, in fact the latter must include a description of these definitions. The same is true for Cantor's diagonal method (with its "u.s.f. ad inf" as an abbreviating definition) and is not obviated by intra-mathematical practice as long as the *standing* of Cantor's diagonal method (including at the heart of foundational proofs) inspires astonishment.

onalisation, but he wants to question the *importance* of this method of constructing loop holes (and consequently the extra-mathematical role and standing of the diagonal method and its diagonalised result). What is philosophically questionable is the need to construct an infinite hierarchy in the first place, a hierarchy that was motivated mainly by the necessity to banish the troubling contradiction that the diagonalisation produced. Instead of accepting the contradictory aspect of the diagonalised element, the hierarchy seems to assign a place even to this diagonalised result, which allows us to cling to our idea of a systematisation, even if it is only of a 'higher order'.

1.7 ENUMERATING RULES

Wittgenstein's discussion of Cantor's diagonal argument continues seamlessly in Ms-162a (and later in Ms-162b), with the first dated remark in Ms-162a from 6 January 1939, a day after his last entry in Ms-121. Unfortunately, none of these remarks have been published in *RFM II*, despite them being Wittgenstein's last extensive treatment of the diagonal argument (there are remarks in later notebooks, most notably Ms-135, 59v.2–60v.4 from 1947, discussed in Chapter 3, but the remarks in those later notebooks remain relatively short and isolated). One reason could have been that Ms-162a/b, being part of the "draft" pocket notebooks in the 160's range of *Nachlass* documents, were considered too unfinished by the editors to be included (though Ms-164 was later published as *RFM VI*), but such a verdict is hard to justify given the remarkable quality of most of Ms-162a and at least the first part of Ms-162b. It is more likely that the editors considered these remarks to be too 'scandalous' or too focused on a specific mathematical issue, similar to many of Wittgenstein's remarks on Gödel, and decided to stick to the (comparatively) more conservative remarks in Ms-117 and Ms-121.[53]

Wittgenstein begins by imagining "*everyday* purposes", a "*practical* purpose" (Ms-162a, 20.2, Ms-162a, 21.1) for Cantor's diagonal method. The remarks that follow on the next 10 pages in Ms-162a are interesting and revisit many issues brought up in Ms-117 and Ms-121, but will not be discussed here. Instead, the following text will highlight only a few other remarks in passing and then focus on remarks on Cantor that stand in particularly close connection to Wittgenstein's

53 Floyd, 2020, pp. 249–250, on Ms-162b: "The notebook itself is an immediate continuation of what Felix Mühlhölzer rightly called the "wonderful" pocket notebook 162a. Notebook 162b is less wonderful, because it is much less clear. And yet its ambition is wonderful in its own way. For here we see Wittgenstein attempting to draw lessons from his discussion of Cantor's diagonal method into the wider purview of his mature philosophy." And concerning the remarks on Cantor in both Ms-162a and Ms-162b: "It seems the editors, perhaps attempting to protect Wittgenstein, decided against publication of such remarks".

thoughts on formal systems, more specifically Russell's logic.[54] Leading up to the discussion of Russell, Wittgenstein writes on page 30:

> Der allgemeine Satz mag sagen, was alle speziellen sagen, aber die allgemeine *Technik* {tut // lehrt} nicht was alle besonderen Techniken tun. [Ms-162a, 30.2]
>
> ---
>
> The general proposition may say what all the special ones say, but the general *technique* does not {do // teach} not what all the special techniques do.

This remark does not introduce an entirely new aspect, but very succinctly distinguishes Wittgenstein's aim from that of many mathematicians and helps to explain the need for a surveyable representation. Propositions of formal logic (such as in Russell's *Principia Mathematica*) can be seen to abstract from concrete situations in order to arrive at a general proposition, and similarly Cantor's diagonal argument can be seen as a stepping stone away from numbers as concrete intensions (the "special cases", Ms-127, 23.3 / *RFM V* §37, quoted above) towards a more general extensionalist view that treats all real numbers uniformly. Of course Wittgenstein does not explicitly say that he associates general propositions with mathematics and specific techniques with philosophy (and of course many mathematicians are in fact very interested in specific techniques), nor should he be interpreted as advocating for a clear demarcation between mathematics and philosophy based on this distinction. Instead, the remark articulates a troubling tendency that Wittgenstein perceives both in the logical work of Russell and the mathematical work of Cantor, or rather *our tendency* to perceive more general pieces of mathematics as more interesting or astonishing than their more specific counterparts. But while these general propositions might be mathematically fascinating, they do little to explain our *concepts* (of numbers for example), because these concepts are reflected in a myriad of specific techniques, each with their own language games. Irrational numbers are "special cases", they form a *family* of related kinds of numbers and are themselves related to cardinal numbers and rational numbers through different techniques (Ts-227a, 58.2–58.3 / *PI* §§67–68), they share a *family resemblance* in other words. Reducing this variety to its most general essence would fail to explain these concepts, the philosophical alternative is then to describe the concepts with the help of a surveyable representation of their techniques.

The remarks that follow in Ms-162a continue this line of thought and are remarkably clear reflections that revisit many of the issues from Ms-121, most notably the three remarks Ms-162a, 33.2–36.2, with Ms-162a, 39.3 re-emphasising the freedom of reaching a different conceptual decision. While certainly interesting, these remarks will be skipped, in favour of remarks several pages later:

---

54 For a discussion of many of the remaining remarks in Ms-162a/b, see Floyd, 2020 and Mühlhölzer, 2020.

> Es scheint Cantor {lehrt // lehre} mich keine neue Technik; er braucht mir nur das Bild zu zeigen, & ich kann sie schon; er nenne mir nur eine die ich schon kannte. Aber das {Schema // Bild} ist neu & bringt etwas neues in Vorschlag wenn wir den Vorschlag allerdings auch gleich verstehen. [Ms-162a, 49.2]

> It seems Cantor is not teaching me a new technique; he only needs to show me the picture, & I already know it; he is only telling me one that I already know. But the {scheme // picture} is new & brings something new to the proposal, even if we understand the proposal at the outset.

Wittgenstein wants to emphasise that Cantor teaches us a *new technique*, which we find so intuitive that we adopt it immediately and cannot even imagine an alternative. As a result, we fail to see that it is a technique and instead interpret it as a fact. Exactly because the argument is so persuasive and seemingly employs only concepts that we supposedly already knew (though perhaps only vaguely), we are mislead into thinking that Cantor had made a discovery about the real numbers, when he has instead proposed a conceptual choice. We cannot order the real numbers because we have *decided* that there cannot be anything that we call an ordering of the real numbers. As Wittgenstein writes shortly after:

> Man kann 'sie' nicht in eine Reihe ordnen, – wer sind die sie, die man nicht ordnen kann? Ich habe eine Begriffsbestimmung gemacht, in der ich die *Ordnung* ausgeschlossen habe, indem ich bestimme, jede Ordnung sei immer nur als *Teilordnung* anzusehen; nun darin '*kann*' man diesen Begriff nicht ordnen. Der Schein der Unmöglichkeit (des Nicht-Könnens) entsteht hier durch die *Art* wie wir den Begriff einführen. Indem {die Bestimmung // eine Bestimmung}, die das Ordnen ausschließt, nachträglich wie eine Entdeckung über den schon fertigen Begriff eingeführt wird. [Ms-162a, 51.2]

> One cannot order 'them' in a series, – who are the them that one cannot order? I have made a definition in which I have excluded the *order* by stating that every order is always to be regarded only as *part of an order*; now in this sense one *cannot* order this concept. The appearance of impossibility (of not being able) arises here through the *way* we introduce the concept. In that {the determination // a determination}, which excludes ordering, is introduced afterwards like a discovery about the already final concept.

But if the impossibility of ordering the real numbers is, as Wittgenstein says, only an "appearance of impossibility", what is the alternative that would be open to us if we made a different conceptual decision? This talk about the "freedom" to "go as you please" (Ms-121, 75r.3) is all well and good, but, we might want to object, is there an *actual alternative*? Wittgenstein does not propose a *general* alternative here, which is only natural considering his explicit emphasis on the use of the diagonal method as a specific technique, but he offers us a therapeutic alternative in the specific context of Russell's logic. A preliminary hint is found on page 68, where Wittgenstein explicitly connects countability with propositions in Russell's system:

> 'Eigenschaft einer Zahl'

O'a = b

> Man zeigt, daß die Operationen mit Kardinalzahlen nicht abzählbar sind, indem man zeigt daß jedem System S solcher Operationen eine neue Operation D'S entspricht. Dann sind aber auch die Sätze der Arithmetik nicht abzählbar. Dagegen sind aber die in Russells System beweisbaren Sätze abzählbar.
>
> Wenn wir die Zeichen Russells als Ziffern auffassen, so wird jeder seiner Sätze ein Zahlzeichen & jeder seiner Beweise eine bestimmte Konstruktionsart einer Zahl (aus den Zahlen der primitive propositions). Wir könnten jeden solchen Satz schreiben: "die Zahl n ist aus r, s, t, u, beweisbar" wo Beweisbarkeit eben eine Eigenschaft von Zahlen ist. [Ms-162a, 68.2–69.2]

'Property of a number'

O'a = b

> One shows that the operations with cardinal numbers are not countable by showing that a new operation D'S corresponds to each system S of such operations. But then the propositions of arithmetic are not countable either. On the other hand, the propositions that are provable in Russell's system are countable.
>
> If we take Russell's signs to be numbers, then each of his propositions becomes a number sign & each of his proofs a particular form of constructing a number (out of the numbers of the primitive propositions). We could write each such proposition: "the number n is provable from r, s, t, u" where provability is precisely a property of numbers.

By "D'S" Wittgenstein means the diagonalisation operation of a particular system S, which, when applied to S, forms a new element that is not a part of the system S (with "O'a" being Wittgenstein's expression for the application of an operation "O" to an argument "a", reaching back to the *Tractatus*). As Wittgenstein already expressed in Ms-121, 90v.1 (see Section 1.6), this diagonalisation operation 'escapes' the countable system of operations, the collection of all operations is thus uncountable. There are of course systems that are closed under diagonalisation, so that their diagonalisation operation would not 'escape' from the system but already be a part of it, which Wittgenstein certainly does not want to deny. The remark is rather vague and could certainly be interpreted in a number of ways, which is not surprising given that Ms-162a is primarily a draft notebook. What matters is that in the following remarks Wittgenstein considers a system that is open to diagonalisation and thus uncountable (such as the real numbers in Cantor's diagonal argument) and compares this system with Russell's system, where propositions are countable. Countability in Russell's system follows immediately from the fact that we can describe a systematic way to (recursively) construct all possible propositions and could thus number each constructed proposition one by one. Russell's system is of course not unique in this regard, the following discussion will treat it interchangeably with other formal systems for which propositions are recursively enumerable.

What Wittgenstein points out may sound rather trivial, but has interesting philosophical repercussions: Formal systems are supposed to be general enough to allow for the expression of arbitrary propositions of arithmetic, yet their propositions are *countable*, while the propositions of arithmetic are *uncountable*, assuming that we can construct a proposition involving any of the uncountable real numbers. How can this be?

Wittgenstein does not answer the question immediately. Instead, the remarks Ms-162a, 70.4 until at least Ms-162a, 89.2, but arguably up to and including Ms-162a, 100.2, discuss provability in Russell's logic, more specifically the difference between being true and being provable, both seen from within Russell's system and 'from the outside'. These issues lie at the heart of Wittgenstein's remarks on Gödel (with his application of the diagonal method) and should be read in the context of Wittgenstein's earlier writings on that topic, even if Wittgenstein does not explicitly refer to Gödel here. While certainly interesting, such an interpretation would go beyond the scope of the present text, these remarks will thus be skipped in favour of remarks at the very end of Ms-162a, where Wittgenstein returns to the issue of countability:

> 'Du kannst nicht alle Kisten der Welt, in eine Kiste legen.' Warum? Weil ihrer zu viele sind? – Ich werde Dir beweisen, daß es eine unendliche Zahl von Kisten gibt; denn keine Kiste, wie groß Du sie auch machst kann alle Kisten {enthalten // beherbergen}.
>
> Man kann nicht alle Systeme auf die Kardinalzahlen aufteilen, weil, sie aufteilen, ein System bilden heißt. [Ms-162a, 101.2–102.1]
>
> ---
>
> 'You can't put all the boxes in the world in one box.' Why? Because there are too many of them? – I will prove to you that there is an infinite number of boxes; for no box, no matter how large you make it, can contain all the boxes.
>
> One cannot apportion all the systems to the cardinal numbers, because to apportion them is to form a system.

The reason why we cannot put all of the world's boxes into a single box is not that there were such an enormous number of boxes that it would be physically impossible to construct a sufficiently large box to hold them all. This would be a *physical impossibility* and Cantor's diagonal argument has the tendency to paint an analogous picture for the real numbers, so that we are lead to believe that the real numbers are too 'numerous', 'too many' to be ordered. But as the box example makes clear, the impossibility at play here is a *logical impossibility*, because even if there were only a single box in the world, the moment we constructed another box to hold the first box this new box would not be contained in a box and could not be called a box that contains all boxes. The logical impossibility of constructing a box that contains all boxes is thus a direct consequence of our *concept* of boxes, because our concept of boxes excludes boxes that contain themselves.

What Wittgenstein's example makes remarkably clear is twofold: First, the 'unboxability' of boxes does not depend on any large numbers, the situation already occurs with just a single box in the world.

Similarly, the uncountability of the real numbers does not depend on an 'enormously large' number of real numbers, only on our ability to construct a new diagonalised number once an ordering system is fixed. As Wittgenstein has made clear before, it is a consequence of an endless permission, not of an astonishingly large number. Second, the simplicity of the box example shows us the 'freedom' to make a different conceptual decision: It may be hard to see how a box in our physical world could contain itself, but we have no trouble imagining such a *mise en abyme* in literature, for example in a fairy tale, a picture that Wittgenstein himself uses in a very similar context in *RFM V*:

> Denke Dir unendliche Zahlen in: einem Märchen gebraucht. Die Zwerge haben soviele Goldstücke aufeinander {gelegt // getürmt}, als es Kardinalzahlen gibt – etc. Was in einem Märchen vorkommen kann, muß doch Sinn haben. – [Ms-126, 54.3 / *BGM V* §6]

> Imagine infinite numbers used in a fairy tale. The dwarves have piled up as many gold pieces as there are cardinal numbers – etc. What can occur in this fairy tale must surely make sense. – [*RFM V* §6]

All well and good, but what use is the example of boxes containing themselves outside of this kind of fairy tale logic? If we are free to make a different conceptual decision even in the case of uncountable systems, as Wittgenstein has repeatedly suggested, then how would such a different conceptual decision look like? This is what Wittgenstein investigates next:

> Numeriere die Systeme {einfach // eben} mit Brüchen zwischen 1 & 2 & behalte die Brüche zwischen 2 & 3, 3 & 4, u.s.w., in Vorrat. Dann kannst Du die Systeme der Systeme nach Herzenslust numerieren, wenn auch nicht in eine Reihe ordnen.
>
> "Du kannst nicht alle Systeme in ein System bringen; daher kannst Du nicht allen Systemen Namen geben; denn Du kannst alle Namen in ein System bringen." – Du kannst allen Systemen Namen geben, solange Du nur die Namen nicht (dadurch) verschwendest, {daß // indem} Du {mit dem System *aller* Namen anfängst // das System *aller* Namen verwendest}. [Ms-162a, 101.2–BCr.1]

> Number the systems simply by using fractions between 1 & 2 & keep the fractions between 2 & 3, 3 & 4, etc., in stock. Then you can number the systems of systems to your heart's content, though not in order.
>
> "You cannot fit all systems into one system; therefore, you cannot give names to all systems; for you can fit all names into one system." - You can give names to all systems, as long as you just don't waste the names by {starting with the system of *all* names // using the system of *all* names}.

The first remark might appear trivial: Of course we can assign numbers to all results of the repeated diagonalisation operation if we use only the fractions between 1 and 2 (which are countably infinite, after all) to enumerate the 'regular' non-diagonalised elements and keep all the other numbers "in stock". Then, whenever the diagonalisation operation produces a new element that cannot be part of the infinite list of elements already in the system, we can assign it a number from the (countably infinite) stockpile of numbers greater than 2, in contrast to Cantor's diagonal argument. As Wittgenstein notes in the second remark, we thus *can* enumerate all systems (including their diagonalisations), as long as we do not "squander" the "system of *all* names" by immediately using up all names for our 'regular' ordered systems without keeping a "stock", a reserve of names for all diagonalisations to come.

That Wittgenstein uses rational instead of natural numbers in his example is not essential in this context. Any 1:1 correspondence between systems and fractions could of course also be enumerated using just natural numbers, given that the rational numbers are countably infinite. The choice of fractions in this example illuminates Wittgenstein's point in a way that would have been obscured by the natural numbers, however, because in contrast to the natural numbers the rational numbers cannot be countably ordered according to their size, as Wittgenstein has pointed out multiple times before. The fact that there are always infinitely many fractions between two fractions is exactly what we need in this picture, because in this way fractions allow us to generate numbers as we go, while leaving the stockpile of numbers for diagonalisations untouched. By using fractions instead of natural numbers, Wittgenstein emphasises a certain aspect of countably infinite numbers, namely that we can picture them as systems of series (in plural, "Reihen"):

> Wenn die Brüche die Namen sind so kann man sie {den // allen} Systemen zuordnen, indem man [break between Ms-162a and Ms-162b] Brüche ihrer Größe nach in Reihen ordnet. Aber nicht {dadurch daß // indem} man alle Brüche in eine Reihe ordnet. [Ms-162a, BCr.2 & Ms-162b, 1r.1]

> If the fractions are the names then they can be assigned to {the // all} systems by [break between Ms-162a and Ms-162b] arranging fractions in series according to their size. But not by arranging all fractions into a series.

But given that we can name all systems in this way using countably infinite names, the next remarks are rather puzzling:

> "Ich kann Einem nicht alle Techniken durch eine Technik beibringen."
>
> Man kann wohl die Namen aller Systeme in eine Reihe ordnen, {aber sie nicht alle den Systemen der Reihe nach zuteilen // aber nicht die Namen allen Systemen der Reihe nach zuordnen}.
>
> D.h. man hat, wenn man eine endlose Reihe von Namen hat, nicht *zu wenig* Namen; wenn sie (nur) nicht *so* aufteilt, daß dem System von Namen ein System von Systemen entspricht. Verwendet man diese besondere Art der Zuteilung, dann hat man zu wenig Namen; und das kann man so ausdrücken: "man kann nicht alle Namen allen Systemen zuteilen" – weil man unter dem Verteilen aller Namen an Systeme das Verteilen aller Namen an ein System von Systemen versteht. [Ms-162b, 1r.2–1v.2]

> "I can't teach one person all the techniques through one technique."
>
> One may well put the names of all the systems in order, {but not assign them all to the systems in order // but not assign the names to all the systems in order}.

> That is, if one has an endless series of names, one does not have *too few* names; if (only) one does not divide them *in such a way* that a system of systems corresponds to the system of names. If one uses this particular kind of allocation, then one has too few names; and this can be expressed as “one cannot allocate all names to all systems” - because by allocating all names to systems one understands allocating all names to a system of systems.

What does Wittgenstein mean when he says that we “can order the names of all systems in a series”, but “cannot assign the names to all systems sequentially"? There are two ways to read this remark:

The more straightforward reading is that the assignment between names and systems is a ‘one way’ assignment, going from systems to names. Given a system, we can assign it a name (a fraction between 1 and 2 for non-diagonalised numbers, for example), but we cannot know which system corresponds to a number reserved in the stockpile until we have produced the system via diagonalisation. This way, it is not possible to diagonalise over the whole system of names, because we cannot produce all the systems as a series. In other words, we have no technique to produce all techniques.[55]

Another, more interesting reading follows from the idea of the box of all boxes and the third remark quoted above, according to which countably infinite names are sufficient if “the system of names does not correspond to a system of systems”. The previous reading assumes that the “system of names” does not correspond to *any* system, it is out of reach *inside* the systems and thus immune to diagonalisation. But the alternative is that the system of names is simply a system, period, not a “system of systems”. In other words, countably infinite names are sufficient not only if there is nothing that corresponds to the system of names, but also if the system of names is itself just a regular element in the series, not a higher-order system. But there is a price to pay for the inclusion of the system of names as a regular first-order system, because the diagonalisation operation will then operate on itself, meaning that the (diagonalising) system corresponding to the system of names is paradoxical.

The two readings correspond to the choice between incompleteness and inconsistency, but they are not as far apart as it might seem, because they both illuminate a different aspect of what it might mean to say that we cannot teach all techniques through a single technique.

---

55 An example would be Turing’s computing machines, which are enumerable, but for which it is not possible to decide in general using finite means whether or not a computing machine corresponds to a number. Computable numbers are thus enumerable, just not using finite means. As a result, we can easily assign numbers to computing machines and list these numbers sequentially, but we cannot know in general which of these numbers correspond to a ‘number system’ or just nonsense. But given a computing machine that computes a number, we can immediately and unambiguously assign it a number, its “description number”. The correspondence between computable number and ‘computable-machine-number-as-name’ is thus a ‘one way’ correspondence. For more details, see Chapter 3.

The incompleteness reading emphasises that a single general technique can give us access to a complete system of names, but that these names are rather shallow if the correspondence with the techniques that they name is only 'one way'. The inconsistency reading emphasises that we can have not only a system of names but also their corresponding techniques, at the cost of not knowing which of these techniques are worth learning or using, because the system contains both 'useful' techniques as well as paradoxical diagonalisations. To say that we can teach all techniques through a single techniques is not wrong, but obviously quite meaningless, as it is akin to showing someone the letters of the alphabet and then saying: "See, now you can form all possible sentences in the English language!"

The first reading corresponds to saying that there is nothing that we call a box that contains all boxes, the second reading to saying that the box of all boxes is a paradoxical box, because it must contain itself. That Wittgenstein is interested not only in the first but also the second reading is made clear not only by his use of the box-of-all-boxes example, but also his explicit remark about a self-referential function $f_x(x)$:

> {'Wir wollen {nicht // unter keinen Umständen} sagen: eine Reihe von Funktionen $f_1(x)$, $f_2(x)$, $f_3(x)$, $\ldots$ enthielte alle Funktionen, da sie $f_x(x)$ nicht enthält.' // 'Wir wollen unter keinen Umständen von einer Reihe von Funktionen $f_1(x)$, $f_2(x)$, $\ldots$ sagen sie enthielte alle Funktionen, {weil $f_x(x)$, (z.B.) nicht in ihr enthalten ist.' // weil $f_x(x)$ (z.B.) keines ihrer Glieder ist.'}}
>
> 'Wir wollen unter keinen Umständen von einer Reihe aller Reihen sprechen.'
>
> 'Wir wollen unter keinen Umständen von einer Kiste sagen, sie enthielte alle Kisten.' [Ms-162b, 2v.2–3r.2]
>
> ---
>
> {'We do {not // not under any circumstances} want to say: a set of functions $f_1(x)$, $f_2(x)$, $f_3(x)$, ... contained all functions, because it does not contain $f_x(x)$.' // 'Under no circumstances do we want to say of a set of functions $f_1(x)$, $f_2(x)$, ... it contains all functions, {because $f_x(x)$, (e.g.) is not contained in it.' // because $f_x(x)$, (e.g.) is not one of its members.'}}
>
> 'We do not want to speak of a series of all series under any circumstances.'
>
> 'Under no circumstances do we want to say of a box that it contains all boxes.'

As Wittgenstein adds in Ms-162b, 5r.1, knowing a "technique to form the signs $f_\nu(x)$ does not teach us anything about the particular systems behind the signs of this series, unlike the series π^x, which is explained by the system of exponentiation (Ms-162b, 4v.2). In the case of the system of all names $f_\nu(x)$, we can order these signs systematically, but such an ordering is not a systematic *explanation*:

> Man kann alle Funktionsnamen in {eine Reihe // ein System} ordnen; aber man kann sie nicht alle systematisch erklären.
>
> 'Ein System von Funktionsnamen systematisch erklären' heißt: eine Erklärung an den Kopf zu stellen.

> 'Den Gebrauch {eines Systems von Zeichen // der Zeichen eines Systems} systematisch erklären' heißt: eine Erklärung an den Kopf des Systems stellen, die die richtige Verwendung der Zeichen des Systems {bewirkt // verbürgt}.
>
> Kann diese Erklärung selbst ein Zeichen des Systems sein?
>
> Die Erklärung der Funktionszeichen geschieht systemweise. [Ms-162b, 5v.2–6r.3]

> One can order all function names into {a series // a system}; but one cannot explain them all systematically.
>
> 'To explain a system of function names systematically' means: to place an explanation at the beginning.
>
> 'To explain systematically the use {of a system of signs // of the signs of a system}' means: to place at the beginning of the system an explanation which {ensures // vouches for} the correct use of the signs of the system.
>
> Can this explanation itself be a sign of the system?
>
> The explanation of the functional signs is done system by system.

Wittgenstein shows us the different implications of the conceptual decision at the heart of Cantor's diagonal argument: In so far as a systematic way to construct the elements of a series can be understood as an explanation of this series, we can choose whether we want to preface our series of systems with this system of systems (and thus exclude it from the series proper) or whether we want to include it as a sign inside the system. In the latter case, it is clear that the inclusion of this system of systems makes the series (of the real numbers, for example) more general than without it (the series of algebraic numbers, for example), but the general series cannot explain the variety of special cases in the same way that specialised explanations could ("systemweise"). The conceptual decision is one between inclusion and exclusion, between including the diagonalisation in the real numbers (*calling* it a number and thus excluding it from the assumed ordering of the real numbers) on the one hand and including it (as a paradoxical diagonalisation) in the ordering (and thus excluding it from the non-paradoxical real numbers) on the other hand. Cantor himself chooses the first option and excludes the diagonalisation from his concept of numbers:

> 'Habe ich Dich {gelehrt // geübt}, ein System von Funktionen zu beherrschen, so habe ich Dich damit auch abgerichtet, eine außerhalb des Systems stehende Funktion zu beherrschen.' [Ms-162b, 6v.1]

> If I have {taught // trained} you to master a system of functions, I have also trained you to master a function that stands outside the system.

What is philosophically problematic about Cantor's diagonal argument is that the diagonalised number is never given 'a fair chance'. Of course it is not possible to order all real numbers if we have already made up our mind that we will not accept the diagonalised number in this series but that we *also* want to call this diagonalised number a real number. If we give up one of these two assumptions, Cantor's proof might not need to bother us very much, let alone astonish us with its uncountable infinity.

We could say that the diagonalisation does not produce a real number of the same sort as other real numbers (and part of the problem of Cantor's proof is that it leads us to believe that we have a clear picture of this 'sort' of number), it can even be argued that we only accept the diagonalised number as a real number because the concept of real numbers is sufficiently vague to begin with. Yes, we know what counts as a real number by viewing all numbers as decimal expansions, but such an extensionalist view does not give us a systematic way to produce the variety of systems that are the irrational numbers.

The more interesting perspective would be to freely accept the diagonalised number in the ordering of the real numbers, an option that is only available because the diagonalised number is not fixed by its use in the same sort of language games as other numbers. If Cantor's proof convinces us that the diagonalised number *is* a real number, then, we might choose to say, why not include it in the ordering if we want to order *all* real numbers? This would of course mean that we need to go 'all the way' and consider what happens if a diagonalisation meets itself (and it is clear that we must leave the neatly organised land of the real numbers that we have grown accustomed to and change our concept of what we call a real number in the process).

Wittgenstein's aim is not to advocate for a particular perspective. Instead, his remarks investigate the different conceptual paths that are open to us, so that we gain an understanding of the philosophical landscape:

> Meine Aufgabe ist es Euch die Geographie eines Labyrinths zu lehren, so zwar, daß Ihr Euch vollkommen darin auskennt. [Ms-162b, 6v.2]
>
> ---
>
> My task is to teach you the geography of a labyrinth, in such a way that you are perfectly familiar with it.

One result of this survey is that we gain an understanding of what it means for a system to be uncountably large. Yes, the real numbers are uncountable and in a certain sense we can even say that the infinity of the real numbers is greater than that of the natural numbers. But we can turn the tables: Everything expressible in a natural language, English for example, can be expressed with a (finite, but arbitrarily long) string of symbols (such as the letters of the alphabet and digits) and can thus be counted, simply lexicographically. We can then imagine these countably infinite strings written down in an infinitely long book (evoking Borges' infinite *Library of Babel*, Borges, 1964), with some of these strings describing real numbers, such as "2", "The first prime number greater than 3" and even Berry's paradox: "The first number not nameable in under ten words".[56] Any possible description of real numbers, including any imaginable description of diagonalised numbers, would then be a part of this book of numbers, which will of course also include an immensely large number of nonsense descriptions. This countably infinite collection of number descriptions is thus in one sense larger than the real numbers (because all the real numbers that we could ever possibly express are expressed somewhere in this collection and additionally many diagonalised and nonsensical descriptions), but in another sense smaller, because the number descriptions are obviously countable, whereas the real numbers are not. Or in other words: By adding the irrational numbers to the rational numbers, the resulting real numbers seem to grow so 'large' that they cannot be counted by the natural numbers. But by then adding the paradoxical diagonalised numbers to the real numbers (together with many other gibberish descriptions, of course), we get an even larger set of 'numbers', which is now countable again, however.[57]

There is nothing surprising about this, at least not if we understand Cantor's diagonal argument in a philosophically 'homespun' way. The image of a countably infinite book with all possible combinations of letters and digits can help us understand that Cantor has not discovered a new and unimaginably large collection of 'things', but rather invented a new concept, which we can then contrast with other concepts to dissolve our astonishment in the face of this seemingly larger-than-infinite collection of numbers. Instead of saying that the set of all real numbers is too *large* to be counted by the natural numbers, we could just as well say that it is too *small*, because it could possibly be counted if it only included the diagonalised number. The price, of course, would be that this diagonalised number became paradoxical, because it will be contradictory at the decimal place where it 'meets itself'. A better, more 'homespun' way to phrase the situation would be to leave aside all talk of different infinities and instead say that the real numbers give us an "endless permission" to produce new numbers and that we will not call anything an "ordering of all real numbers".

---

56 This version of Berry's paradox is due to Bennett (Bennett, 1979, p. 3), see Appendix A for a more detailed discussion of Berry's paradox.

57 Such an observation by no means originates in Wittgenstein's remarks, but was already noted in 1908 by Russell as the fifth contradiction in his list that motivates his theory of types and originates in even earlier publications from 1905 and 1906 (Russell, 1908, p. 153):

> Among transfinite ordinals some can be defined, while others cannot; for the total number of possible definitions is $\aleph_0$, while the number of transfinite ordinals exceeds $\aleph_0$. Hence there must be indefinable ordinals, and among these there must be a least. But this is defined as "the least indefinable ordinal", which is a contradiction.

It is fitting that Wittgenstein ends his reflections on countability and Cantor in Ms-162b with remarks on surveyability, which then gradually develop into more general thoughts on *Gestalt* psychology:

> "{Sei // Aber sei} nicht lächerlich! Freilich bedienen wir uns zum Erkennen der Anzahl gewisser Mittel, die Anzahl übersichtlich zu machen; z.B. des Dezimalsystems." – Aber was ist hier Zweck, & was Mittel?
>
> Wir würden also nicht *erkennen*, daß 10.000 Variable in dieser Klammer stehen. Ist das nicht, als sagte man: "wir würden nicht erkennen, wieviel Jahre der Elefant lebt, wenn die Erde nicht um die Sonne ginge"? [Ms-162b, 7v.1–7v.2]
>
> ---
>
> "But don't be ridiculous! Admittedly, in order to recognise the number, we use certain means to make the number surveyable; e.g. by using the decimal system." - But what is the end here, & what the means?
>
> So, we would not *recognise* that there are 10,000 variables in this bracket. Isn't that like saying, "we wouldn't recognise how many years the elephant lives if the earth didn't go around the sun"?

Surveyable descriptions are not just "means" to an end, as mere notations that help us discover eternal mathematical truths. Such a view is the misleading side effect of mathematical generality and the search for foundational systems. A philosophical investigation can show us that only the variety of particular mathematical systems, each with their own associated techniques, gives meaning to the more general formalisations, not the other way around.

2 GÖDEL, THEOREMS AND PROVABILITY

> Man {könnte // kann} mit Recht fragen, welche Wichtigkeit Gödel's Beweis für unsre Arbeit habe. Denn ein Stück Mathematik {kann nicht Probleme von der Art der unsern // kann Probleme von der Art, die *uns* beunruhigen, nicht lösen. // kann kein Problem von der Art, die *uns* beunruhigt lösen. // kann nicht Probleme von der Art, die uns beunruhigt, lösen.} – Die Antwort ist: daß die *Situation* uns interessiert, in die ein solcher Beweis uns bringt. 'Was sollen {wir // sie} nun sagen?' – das ist unser Thema. [Ms-124, 94.2 / *BGM VII* §22]

> It might justly be asked what importance Gödel's proof has for our work. For a piece of mathematics cannot solve problems of the sort that trouble *us*. – The answer is that the *situation*, into which such a proof brings us, is of interest to us. 'What are we to say now?' – That is our theme. [*RFM VII* §22]

Wittgenstein's remarks on Kurt Gödel's *First Incompleteness Theorem*, written in 1937 and posthumously published as appendix III of part I of the *Remarks on the Foundations of Mathematics*, form without doubt the most infamous of all of the remarks in the *RFM* and were upon their first publication heavily criticised for misunderstanding the mathematical content of Gödel's seminal result.[1] Instead of diving directly into Wittgenstein's thoughts on the matter, it is therefore useful to face this criticism head-on, by properly setting the mathematical stage for Gödel's result. This will require a certain degree of mathematical detail, as the technicalities of Gödel's result are far from trivial. It is nevertheless important not to lose sight of *why* Wittgenstein devoted his time to write about such a highly technical result, which is why the first step towards a closer examination of Wittgenstein's remarks should start with a broad (and necessarily simplified) exposition of Wittgenstein's interest in the matter and the philological status of the remarks in comparison to other parts of the *Nachlass*.

In contrast to the impression one might get from the rather disparaging initial reviews of the remarks on Gödel, Wittgenstein's interest in this particular theorem is not limited to a single passage of remarks in the *Nachlass*, but rather forms a thread that runs through several years and documents. The remarks of *RFM I; App. III* originate in Ms-118105–116 and were written in 1937, that they were then included in not just one but several typescripts (Ts-221a/b246–255, Ts-223246–255) shows that Wittgenstein considered them, at least during

1 See Kreisel, 1958, Dummett, 1959 and Bernays, 1959. However, beginning in the 80s, these early and dismissive reviews were increasingly challenged, for example in Shanker, 1987, p. viii: "But what were presented as *corrections* were, in fact, covert *philosophical objections* which, because of the prior assumption, were developed without any effort to clarify, let alone challenge, the philosophical background on which Wittgenstein had based his approach to the foundations dispute."

that time period, to be of sufficient quality to survive the draft stage and merit inclusion in typescripts, sometimes with minor revisions. There are also (unpublished) mentions of Gödel in Ms-12172–85, written in December of 1938 and January 1939, as well as a longer passage of remarks on Gödel in Ms-124, 87–95, published as part of *RFM VII* and written between 2.7.1941 and 4.7.1941. Wittgenstein's interest in Gödel's incompleteness theorem is not limited to the *Nachlass*, considering that he gave his "Lectures on Gödel" during the Easter Term of 1938 as part of the "Whewell's Court Lectures".

But why was Wittgenstein interested enough in such a highly technical result that he revisited it over the span of five years? Gödel's incompleteness theorem is of course first and foremost a strictly mathematical result, with direct applications to formal logic and its metatheory, but with little practical consequences for how we calculate and use mathematics in practice. As Wittgenstein hints at in the remark quoted above, Gödel's proof cannot solve any philosophical problem (at least in the eyes of Wittgenstein, Gödel's view on this issue might be another matter) and it would go against Wittgenstein's philosophical convictions to interfere in mathematical instead of philosophical matters. The topic in question is not the mathematical validity of Gödel's result itself, but rather the "*situation*, into which such a proof brings us". Wittgenstein is interested in what happens after we accept the proof as valid, in how we act or are unable to act based upon this proof, how we are puzzled by its conclusion. This is why Wittgenstein's remarks are focused on the frontier between purely mathematical results and their philosophical interpretations. The exact nature of this "situation" will be investigated in more detail below, what is most relevant for now is that even Wittgenstein's more mathematical remarks should not be too quickly interpreted as comments on the mathematical details. They often only set the stage for an investigation that revolves around the *interpretation* of these formal results in informal "prose", which leads otherwise careful mathematicians to sometimes venture into philosophical dogmatism. This is what Wittgenstein wants to attack, not the mathematical result itself.

Before looking at Gödel's proof in more detail, two connections to other topics in Wittgenstein's *Nachlass* should briefly be pointed out to further help explain Wittgenstein's interest in the proof. The first is that Gödel's theorem sent a shock wave through the field loosely known as the "foundations of mathematics", with a particularly serious impact on Hilbert's attempt to give a definitive formalisation of arithmetic.[2] Wittgenstein was of course heavily influenced by the

2 As has been pointed out in Shanker, 1988, p. 177, it can appear peculiar in the context of Hilbert's program that Wittgenstein criticised Gödel's theorem so vehemently: "After all, [Gödel's proof] is an esoteric work which if anything should have contributed to the enhancement of Wittgenstein's own stated ambitions. Or at least so it has been argued, on the grounds that given the purpose of Godel's theorem – to establish the impossibility of constructing a finitary consistency proof for arithmetic

foundational endeavours of Frege and Russell and highly suspicious of metamathematical constructs such as Russell's theory of types.[3] It is no accident that the phrase "Foundations of Mathematics" appears in the (posthumously chosen) titles of both the "Remarks on the Foundations of Mathematics" as well as his 1939 "Lectures on the Foundations of Mathematics": Foundational issues remained Wittgenstein's main interest in his philosophy of mathematics right up until his death, which explains why he was interested in Gödel's seminal proof.[4]

Secondly, Gödel's proof is at its heart a *diagonal argument*, with similarities to Cantor's original use of the diagonal method in his proof of the uncountability of real numbers. Wittgenstein wrote on multiple occasions on Cantor's proof and it is no accident that Gödel is mentioned in Ms-121, a document that is otherwise focused on Cantor, as both arguments can lead to similarly misleading philosophical conclusions. This is not to suggest that Wittgenstein sees these different diagonal arguments only as superficial symptoms of the same

– it is far from clear why Wittgenstein should have sacrificed such an important potential ally. For the one thing that is manifest in Wittgenstein's writings on the philosophy of mathematics is his desire to challenge the two central themes that together constitute Hilbert's Programme: the formalization of mathematical thought, in order to produce a finitary proof of the reliability of mathematical reasoning." However, as both Shanker's interpretation and the present chapter will show, the reasons for rejecting Hilbert's attempt to fully formalise mathematics differ considerably in the case of Gödel and Wittgenstein. In fact, Wittgenstein reads Gödel's interpretation as giving the *appearance* of an answer to what is actually a *philosophical* problem, which can only be clarified through a philosophical investigation, not on the basis of a mathematical proof. (Shanker, 1988, p. 183: "The heart of Wittgenstein's critique of the platonist interpretation of Godel's theorem lies in the principle that it is only possible to resolve a philosophical problem *philosophically*;") Furthermore, in an important sense Gödel's result also has a *positive* impact on Hilbert's program by legitimising its meta-mathematical approach, as Shanker, 1988, p. 225 emphasises: "Gödel was aware of the *negative* impact of his theorem on Hilbert's Programme, but even more important to him was the *positive* role which he hoped his proof would perform: if genuinely successful it would utilize 'Hilbert's meta-mathematical progeny' to revitalize platonism in the midst of the positivist mood dominating analytic philosophy in the 1930s." It is this tendency that Wittgenstein is criticising in his remarks above all else. See also Floyd, 2001, p. 287: "Gödel's theorems are not a threat to Wittgenstein, and he has no special animus toward or against them, except in their philosophical misuses."

3 As Shanker, 1988, p. 192 correctly points out, Wittgenstein's distrust of the need for higher-order meta-theories precedes Gödel's proof (Shanker cites Ts-209, 74.10 / *PR 153*: "On the other hand there can't in any fundamental sense be such a thing as meta-mathematics. Everything must be of one type (or, what comes to the same thing, not of a type)."). This philosophical context is certainly one of the primary reasons why Wittgenstein chose to investigate this particular mathematical result.

4 Additionally, Gödel's proof shows how seemingly rather specialised issues can have far reaching consequences, with the effect that a philosophical investigation of Gödel's proof does more than just investigate a single and isolated piece of mathematics. As Ramharter, 2008, p. 8 notes: "Gerade in der Auseinandersetzung mit Gödel zeigt sich, wie weit die Überlegungen ausstrahlen, die sich Wittgenstein zu sehr speziell anmutenden Fragen macht, und welche allgemeinen Konsequenzen diese Überlegungen haben."

'essential' illness, as such an interpretation would misattribute a generalising tendency to Wittgenstein's investigations, which stay in fact always close to their particular subject matter. But as this discussion of Wittgenstein's remarks on Gödel attempts to show, there is a strong family resemblance between the philosophical abuses of the different diagonal arguments and it is Wittgenstein's goal to draw our attention to these abuses, by demonstrating how they are favoured by the prosaic interpretation of perfectly valid mathematical proofs.

So did Wittgenstein actually disagree with Gödel? Of course any claim will have to be substantiated in the following discussion, but a preliminary answer might nevertheless help in navigating the more technical parts of the discussion. In this text, the following stance will be defended: While Wittgenstein's remarks need not be read as disagreeing with *the mathematician Gödel*, they are certainly in disagreement with *the philosopher Gödel* and his mathematical platonism.[5] Gödel's philosophical interpretation of his own proof is the sort of "one-sided diet" (Ms-116, 255.3; Ms-120, 135v.2; Ts-227a/b, 291.2; Ts-228, 53.3; Ts-230a/b/c, 11.5 / *PI* §593) that Wittgenstein attempts to dissolve through a surveyable presentation, with the more mathematical remarks as means to an end, not as a critique of the mathematical result itself.[6]

The structure of this chapter roughly follows the segmentation of Wittgenstein's remarks on Gödel into three main passages in different documents (Ts-221a/b, Ms-121 and Ms-124/Ms-163). However, the considerable mathematical ingenuity of Gödel's proof makes it necessary to briefly describe what exactly Gödel's (first) incompleteness theorem proves and how Gödel's own informal introduction relates to his precise mathematical result (Section 2.1). Given that the primary focus of this chapter is Wittgenstein's philosophical investigation on Gödel, not the finer mathematical points of Gödel's results, the presentation of Gödel's theorem will necessarily lack in mathematical detail but should nevertheless provide a sufficient understanding of the context of Wittgenstein's remarks. The first two of the main sections on Wittgenstein discuss his remarks in Ts-221a/b / *RFM I; App. III*, which are the most heavily criticised writings on Gödel in the *Nachlass*, by first examining the role of truth and prov-

5 This is not to suggest that Wittgenstein's intent would be to replace Gödel's platonism with an equally dogmatic outlook, such as for example finitism. While it cannot be denied that Wittgenstein's writings in the early 1930s sometimes exhibit finitist tendencies, these more dogmatic remarks do not play a major role in his discussion of Gödel and will not be discussed in the following text. As Shanker, 1988, p. 215 notes, Wittgenstein's "objection to Hilbert's Programme stemmed from his remarks on the logical nature of mathematical propositions, and not from any 'finitistic' misgivings."

6 As Floyd, 2001, pp. 288–289 puts it: "But does this mathematical work resolve or uniquely interpret the general philosophical questions that have traditionally been asked about (our notions of) *truth*, *proof*, and *mathematics*? [...] Gödel's answer was, Yes. Wittgenstein's was, No."

ability in Wittgenstein's investigation (Section 2.2) and then his peculiar attitude towards inconsistency (Section 2.3). The next two sections deal with the entirely unpublished remarks in Ms-121 (Section 2.4 and Section 2.5), which are not always as 'mature' as some of the published remarks, but nevertheless introduce new aspects in Wittgenstein's thoughts on Gödel and provide evidence that Wittgenstein did understand many of the mathematical details of Gödel's proof. The last sections are focused on the remarks in Ms-124 / *RFM VII*, which constitute Wittgenstein's 'last word' on Gödel and are some of the most undogmatic and mature remarks on this topic. The central aspect in this passage is the concept of *surveyability* (Section 2.6), which will be shown to be essential for an understanding of why Wittgenstein chose to attack Gödel's interpretation of the theorem. The remarks in Ms-124 are continued in the pocket notebook Ms-163, which shows interesting connections to the notion of "ultraphysical" impossibility. The remarks are briefly examined in the last section (Section 2.7).

2.1 GÖDEL'S DIAGONAL ARGUMENT

In 1931, Gödel published his seminal paper "Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I", with its two famous incompleteness theorems (the theorems VI and XI).[7] As we will see, Wittgenstein's remarks are focused on the first of these two theorems and so "Gödel's incompleteness theorem" will in the following discussion usually refer to theorem VI of Gödel's paper, or, if the context renders the usage unambiguous, to the conjunction of the two theorems.

Gödel begins his paper with an informal introduction to his formal mathematical result and it can be useful to briefly walk through this introduction to 'get a feeling' for what Gödel's incompleteness theorem is about. It should be clear, however, that 'what counts in the end' is Gödel's formal result and that his informal introduction must sometimes trade in exactness and correctness to achieve a more easily understandable exposition. Gödel himself is very clear about this: "Before going into details, we shall first sketch the main idea of the proof, of course without any claim to complete precision" (Gödel, 1986, p. 147). Similarly, after finishing the sketch of the proof, Gödel notes: "We now proceed to carry out with full precision the proof sketched above" (Gödel, 1986, p. 151). Not too much weight should thus be placed on the merely informal comments by Gödel, they are meant as a guide to the larger proof, not as an independent reformu-

7 Its importance in the field of logic can hardly be overstated (Berto, 2009, p. xii), see e.g. Gödel, 1986, p. 126: "Gödel's 1931 was undoubtedly the most exciting and the most cited article in mathematical logic and foundations to appear in the first eighty years of this century."

lation of it. Wittgenstein has been criticised as misunderstanding this crucial point: According to such a reading, all that Wittgenstein's remarks do is attack perceived flaws in the informal introduction, while completely missing the formal point of Gödel's paper.[8] Leaving aside for now whether this charge is warranted, Gödel's informal remarks are certainly important for a proper understanding of Wittgenstein's line of investigation, as they represent at least in parts Gödel's own prosaic interpretation of his formal result. Let us thus look at the informal proof sketch, before supplementing it with selected passages from Gödel's more precise proof.

As Gödel explains in the introduction, the incompleteness theorem demonstrates a limitation of those formal systems that fulfil certain technical requirements (most importantly ω-consistency, a stronger form of logical consistency) and are strong enough to formalise arithmetic. The classic example is Whitehead and Russell's *Principia Mathematica* ("PM"), a logical system that is general enough that we might expect it to *decide* any mathematical question (that is, derive from its axioms either the proposition in question or its negation).[9] This assumption seems natural, but as Gödel will show in his paper, is mistaken:

> Es liegt daher die Vermutung nahe, daß diese Axiome und Schlußregeln dazu ausreichen, *alle* mathematischen Fragen, die sich in den betreffenden Systemen überhaupt formal ausdrücken lassen, auch zu entscheiden. Im folgenden wird gezeigt, daß dies nicht der Fall ist, sondern daß es in den beiden angeführten Systemen sogar relativ einfache Probleme aus der Theorie der gewöhnlichen ganzen Zahlen gibt, die sich aus den Axiomen nicht entscheiden lassen. [Gödel, 1986, p. 144]

> One might therefore conjecture that these axioms and rules of inference are sufficient to decide any mathematical question that can at all be formally expressed in these systems. It will be shown below that this is not the case, that on the contrary there are in the two systems mentioned relatively simple problems in the theory of integers that cannot be decided on the basis of the axioms. [Gödel, 1986, p. 145]

These systems are thus (syntactically) *incomplete*, as there exist "simple problems" of arithmetic that cannot be decided in these systems.[10]

8 Such a reading usually interprets Wittgenstein's remarks as follows: Wittgenstein's main mistake is to misunderstand Gödel's informal mention of a "proposition that says about itself that it is not provable" (Gödel, 1986, pp. 149–151) as an exact description of the mathematical machinery and Gödel's full result as a logical paradox, when there is actually no direct reference of a proposition to "itself" and no paradox in the formal proof. Wittgenstein certainly takes up Gödel's phrasing of the "proposition that says about itself that it is not provable", but this is on its own insufficient to disqualify Wittgenstein's remarks (see Ramharter, 2008, pp. 14–15).

9 While Gödel's proof is focused only on *Principia Mathematica*, the results obtained by Gödel are generalisable to a large range of formal systems that are strong enough to formalise certain basic propositions of arithmetic, as Gödel himself was well aware.

10 Whether these problems can be decided *outside* these systems and what decidability-outside-the-system might mean are some of the philosophical questions that Wittgenstein was interested in. These issues will be ignored during the more mathematical discussion of Gödel's result, but will re-emerge as central later on.

The proof of the undecidability of certain propositions in these systems, as sketched out by Gödel in the introduction, proceeds as follows: Formulas in a system such as PM are "finite sequences of primitive signs (variables, logical constants, and parentheses or punctuation dots)", while "proofs, from a formal point of view, are nothing but finite sequences of formulas" (Gödel, 1986, p. 147). By a procedure that has since become known as "Gödelization", these sequences of (sequences of) primitive signs are encoded as numbers so that arithmetical formulas in the system can be interpreted as metamathematical propositions: "The metamathematical notions (propositions) thus become notions (propositions) about natural numbers or sequences of them; therefore they can (at least in part) be expressed by the symbols of the system PM itself" (Gödel, 1986, p. 147). For example, instead of talking meta-theoretically about how a chain of formulas is a proof of a particular theorem, this encoding allows Gödel to specify a function that does the same job, by taking as inputs two numbers, namely the number encoding of a chain of formulas (corresponding to the proof that is to be checked) and the number encoding of a single formula (corresponding to the theorem), and returning 1 if and only if the chain of formulas is a proof of the theorem and 0 otherwise.

Gödel then goes on to construct an undecidable proposition, that is, a proposition for which neither it nor its negation can be derived in the system. Why and how is this proposition undecidable? It is constructed in such a way that it corresponds (via the encoding of metamathematical propositions as propositions about natural numbers) to a metamathematical proposition about the unprovability of the proposition with the encoding as number q. In other words, it is an arithmetical proposition that, when interpreted according to its Gödelization as a metamathematical proposition, states the unprovability of a particular proposition in the system, with this particular proposition having the encoding q. This in itself would not be remarkable and does not yet lead to undecidability: The metamathematical proposition stating the unprovability could either be false (if the proposition corresponding to q is after all provable in the system) or it could be true but the negation of the proposition corresponding to q could be provable in the system. In both cases it would be possible to either prove the proposition in question or its negation and therefore no undecidability results. But the way Gödel constructs the metamathematical proposition, it turns out that the proposition corresponding to q is this metamathematical proposition itself, so that the metamathematical proposition states the unprovability of itself. This is then how a proposition can talk "about itself" in Gödel's paper: Not directly, but only as an arithmetical proposition about numbers, which (if interpreted through the lens of the gödelised correspondence of numbers as propositions and proofs) corresponds to a metamathematical proposition q about the proposition q.

This is then how undecidability results: Let us assume that the proposition corresponding to q is provable in the system (and here "is provable" is still understood as corresponding to an arithmetic function). The metamathematical proposition corresponding to q states that the proposition corresponding to q is unprovable. A proof of it would thus be a proof that it is not provable and thus lead to a contradiction, from which we conclude that the assumption that the proposition corresponding to q is provable must be wrong. Let us assume that the *negation* of the proposition corresponding to q is provable in the system. This negation states that q is provable in the system (again understood as corresponding to an arithmetic function). But this would mean that both the negation of the proposition corresponding to q and the proposition corresponding to q itself are true, which is also a contradiction. As both assumptions lead to a contradiction, neither the proposition corresponding to q nor its negation can be proved in the system (without producing a contradiction), the proposition corresponding to q is thus undecidable.

The phrasing of 'the proposition corresponding to a number x' (or even more precisely 'the proposition corresponding to the formula with the encoding as number x') might seem overly verbose, but it allows Gödel to avoid a direct self-reference and distinguishes his proof from a paradox. Provability is in the context of Gödel's paper always understood in terms of recursive functions, meaning functions whose operation can be carried out mechanically step by step, based on unambiguous and finite rules.[11] By using only a syntactic and mechanisable notion of provability, Gödel can avoid any circularity, as he explains in a footnote:

> Ein solcher Satz hat entgegen dem Anschein nichts Zirkelhaftes an sich, denn er behauptet zunachst die Unbeweisbarkeit einer ganz bestimmten Formel (namlich der q-ten in der lexikographischen Anordnung bei einer bestimmten Einsetzung), und erst nachtraglich (gewissermaßen zufällig) stellt sich heraus, daß diese Formel gerade die ist, in der er selbst ausgedrückt wurde. [Gödel, 1986, p. 150]

> Contrary to appearances, such a proposition involves no faulty circularity, for initially it [only] asserts that a certain well-defined formula (namely, the one obtained from the qth formula in the lexicographic order by a certain substitution) is unprovable. Only subsequently (and so to speak by chance) does it turn out that this formula is precisely the one by which the proposition itself was expressed. [Gödel, 1986, p. 151]

11 Even in the introduction, Gödel explicitly mentions that the metamathematical propositions are to be understood in this way, as corresponding via Gödelization to propositions of arithmetic:

> In other words, the procedure described above yields an isomorphic image of the system PM in the domain of arithmetic, and all metamathematical arguments can just as well be carried out in this isomorphic image. This is what we do below when we sketch the proof; that is, by "formula", "proposition", "variable", and so on, *we must always understand the corresponding objects of the isomorphic image*. [Gödel, 1986, p. 147]

It is of course quite hard to grasp the nature of the proof from the condensed and informal description that Gödel gives in the beginning of his paper. But as the footnote above makes clear, the "proposition that says about itself that it is not provable" is not a directly self-referential proposition in the sense of the Liar (even though Gödel is the first to point out the similarities with the Liar and other antinomies, see Gödel, 1986, p. 149). There is no circularity, but the proposition can nevertheless be interpreted as a metamathematical proposition about itself thanks to a diagonalisation very similar to Cantor's original argument: All formulas in the system are uniquely encoded as a number (in an enumerable arrangement of numbers) and just like in the case of the uncountability of the real numbers it is possible to construct through diagonalisation a special object that conflicts with its number in the enumeration. In Gödel's case this conflicts is resolved by the undecidability of the formula, because the assumption of decidability (of either the formula or its negation) leads to a contradiction.

So far, most of the remarks in Gödel's introduction that have been discussed are only a slightly imprecise sketch of his larger proof and there is little to criticise from the perspective of Wittgenstein, as long as one keeps in mind that metamathematical notions such as provability are used in a very particular sense by Gödel and might not always directly correspond to our understanding of these terms in more informal language. There is however one passage in Gödel's introductory remarks which is worth highlighting, as it introduces an aspect that goes beyond the formal proof and ventures into philosophical waters ($[R(q); q]$ is the "proposition that says about itself that it is not provable"):

> Aus der Bemerkung, daß $[R(q); q]$ seine eigene Unbeweisbarkeit behauptet, folgt sofort, daB $[R(q); q]$ richtig ist, denn $[R(q); q]$ ist ja unbeweisbar (weil unentscheidbar). Der *im System PM* unentscheidbare Satz wurde also durch metamathematische Uberlegungen doch entschieden. [Gödel, 1986, p. 150, Gödel's emphasis]

> From the remark that $[R(q); q]$ says about itself that it is not provable, it follows at once that $[R(q); q]$ is true, for $[R(q); q]$ is indeed unprovable (being undecidable). Thus, the proposition that is undecidable in the system PM still was decided by metamathematical considerations. [Gödel, 1986, p. 151]

The phrase "it follows at once that $[R(q); q]$ is true" introduces a notion of truth that was not present until now: It is not a notion that has been defined recursively as an arithmetic function *inside the system*, but rather marks the transition into meta-theoretical considerations *without any correspondence in the system*. From the (intra-systemic) result that the proposition is undecidable (and thus unprovable intra-systemically) Gödel concludes (extra-systemically) that it must be *true* (extra-systemically), because it (intra-systemically) asserts its own unprovability. The (intra-systemically) undecidable proposition has thus been (extra-systemically) decided.

Gödel's introductory remarks, at least in this very specific passage, thus go beyond what his formal proof can do, since the reasoning employed by Gödel steps "outside" the formal system to conclude the truth of the proposition based on the inability of the formal system to decide it. At least in this passage, Gödel's introduction is 'prosaic' and does more than simply sketch out in a slightly imprecise way something that will follow more precisely later on.[12] There is no formalised or refined notion of truth that would lend meaning to Gödel's assertion that "$[R(q); q]$ is true", this notion of truth has meaning only in terms of considerations that must step outside the formal system and rely solely on "prose".[13]

One important aspect of Gödel's result that has been glanced over in the discussion so far (but which becomes crucial for the detailed proof) is the requirement of *ω-consistency*. In contrast to 'simple' inconsistency, which results if for some formula p both p and $\neg p$ can be derived in the system, ω-inconsistency results if for a formula $a(x)$ with a single free variable x all the formulas $a(0)$, $a(1)$, ..., $a(n)$ and the formula $\neg\forall x\ a(x)$ can be derived in the system. A system can thus be ω-inconsistent without being inconsistent and ω-consistency entails consistency.

12 Shanker, 1988, p. 220: "It is all too easy to overlook the radical philosophical claim here presented. Contained in these brief words is not simply an obituary of Hilbert's Programme (and with it a covert plea for platonism); at stake is our understanding of the logical character of mathematical propositions: of the nature of the relation between a mathematical proposition and its proof."

13 It should be noted that Wittgenstein's conception of "prose" is independent from the question of whether a mathematical result is expressed in ordinary informal language or formal logic, in other words it is possible to express Gödel's incompleteness theorem in ordinary language without producing philosophically-problematic prose and it is also possible to give prose a formal clothing (although it is certainly true that philosophical abuses are more likely to arise in contexts where ordinary language is used to talk about formal results). Berto, 2009 gives an example of a description of Gödel's (first) incompleteness theorem in ordinary language that Wittgenstein would have most likely considered to be unproblematic. The description uses somewhat different terms, with "G1" corresponding to Gödel's theorem VI, "**TNT**" corresponding to *Principia Mathematica* and γ corresponding to Gödel's "proposition that says about itself that it is not provable" with the encoding as number q:

> Specifically, G1 attests only to the syntactic incompleteness of **TNT**, which means (putting things down as formalistically as possible): assuming the (omega-) consistency of **TNT**, there exists a string of symbols, which we have dubbed γ, not producible as a theorem of **TNT**, and such that the same string with a "$\neg$" put in front of it is also not producible as a theorem. Or, given the axioms of **TNT**, which are finite strings of symbols, and the rules of inference, which are instructions for the manipulation of symbols, if **TNT** is (omega-) consistent, then neither the string called γ nor the string called $\neg\gamma$ will ever show up in the last line of a proof of **TNT**, that is, in the last line of a sequence of strings obtained starting with the axioms, and by means of manipulations allowed by the rules.

Here, there is no talk of a proposition saying something "about itself", no mention of truth or semantic completeness. Instead, this description reflects precisely what happens on a *syntactic* level for "manipulations allowed by the rules", with Gödel's theorem showing that there are theorems that are "not producible" in the system, because they will not "show up in the last line of a proof". The description presents a precise picture of Gödel's theorem, without venturing into unfounded philosophical interpretations.

The difference between ω-consistency and simple consistency will not play a big role for the following discussion and it is in fact possible to prove the undecidability of certain propositions in systems with very similar requirements as in Gödel's proof, but for simple consistency instead of the stronger requirement of ω-consistency.[14] The proof for simple consistency constructs a different and somewhat more intricate (and for simple consistency undecidable) proposition than the (for ω-consistency undecidable) "proposition that says about itself that it is not provable". A discussion of this more intricate undecidable proposition would go beyond the scope of this text, but it should be noted that while Gödel's result is often treated as being easily adaptable to simple instead of ω-consistency, this seemingly innocuous change goes beyond Gödel's own proof and requires a different undecidable proposition.[15]

More importantly, Gödel's first incompleteness theorem is a *conditional* result: The undecidable "proposition that says about itself that it is not provable" is only undecidable *if the system in question is ω-consistent*. On the one hand, it seems obvious that Gödel has to make *some* assumptions about the consistency of the formal system, since in classical logic everything follows from a contradiction and thus of course even the undecidable proposition (and its negation) could be derived in such a (simply) inconsistent formal system. In other words, undecidability depends on some form of consistency. On the other hand, Gödel's assumption of ω-consistency is rather strong and leaves open the question of whether this consistency (or even simple consistency) can be assumed to hold in informal mathematics, including the "metamathematical considerations" that lead Gödel to proclaim that "the proposition that is undecidable *in the system PM* still was decided" (Gödel, 1986, p. 151). Such questions would go beyond the (mathematical) scope and goal of Gödel's paper, but will

14 This was proved in 1936 by Rosser, a few years after Gödel's result, see Berto, 2009, pp. 100–101.

15 This distinction lies at the heart of Floyd and Putnam's "note" on Wittgenstein's "notorious paragraph" on Gödel (Floyd and Putnam, 2000), which attributes to Wittgenstein an advanced understanding of model theory. According to their interpretation, Wittgenstein's insight depends on the fact that Gödel's result requires ω-consistency and not simple consistency and the remarks on Gödel only appear as a misunderstanding if one ignores the difference between Gödel's and Rosser's undecidable propositions. Floyd and Putnam's interpretation certainly offers a very insightful and charitable reading of Wittgenstein's remarks, but has the disadvantage of reducing Wittgenstein's philosophical critique to a very specific observation that only applies to Gödel, but not to Rosser or diagonal arguments as a whole. The "Floyd-Putnam Thesis" has led to a lively debate, with objections coming from Rodych, 2003 and Bays, 2004 (building on Steiner, 2001), followed by Floyd and Putnam, 2006, which was later revised and published in Floyd and Putnam, 2008 to include a rebuttal of Bays, 2006. While the Floyd-Putnam Thesis is certainly interesting, this chapter will instead propose a different interpretation that is closer in spirit to a paraconsistent reading along the lines of Priest, 2004 and Berto, 2009, pp. 189–213. Accordingly, the debate that grew from the original Floyd-Putnam paper will not be examined here (although it is not over, see e.g. the recent Lajevardi, 2021).

become relevant in the context of Wittgenstein's (philosophical) remarks. While it is understandable that as a mathematician Gödel is interested primarily in (ω-)consistent formal systems, it is important to keep in mind that this assumption of (ω-)consistency is not philosophically 'privileged' and should be critically examined whenever philosophical interpretations are drawn from Gödel's result: Consistency might be a mathematical ideal, but whether this ideal is justified or even an apt description of our mathematical practice is a question that cannot be answered on the basis of Gödel's mathematical proof alone.

Lastly, it can be useful to very briefly discuss Gödel's *second* incompleteness theorem (theorem XI in the paper). It is a corollary of the first incompleteness theorem and overall less relevant in the context of Wittgenstein's remarks than the first, but nevertheless important to understand the relevance of Gödel's result for mathematics and especially the foundations of mathematics. Gödel himself only sketched out some of the parts leading up to theorem XI and no detailed discussion of the theorem will be attempted here either. But slightly simplified, the theorem states that no formal system of the aforementioned nature (being ω-consistent, axiomatised so that its set of axioms is decidable, and strong enough to formalise arithmetic and primitive recursive functions) can prove its own ω-consistency. If such a proof could be carried out inside the formal system, the antecedent of the first incompleteness theorem (which states that if the formal system is ω-consistent, there is a proposition that is not decidable in the system) would be derivable and thus via *modus ponens* the undecidable proposition itself, which would make it decidable after all and thus lead to a contradiction.[16]

Before finally focusing on Wittgenstein, let us recapitulate the most salient points for a philosophical discussion of Gödel's result, keeping in mind that any brief and informal exposition will necessarily have to ignore some of the finer mathematical details. What has actually been proved by Gödel?

The first incompleteness theorem states that, for systems that satisfy specific requirements (being ω-consistent, axiomatised so that its set of axioms is decidable, and strong enough to formalise arithmetic and primitive recursive functions), it is possible to construct a proposition that is not derivable from the rules of the system and neither is its negation. The formal proof of the theorem carefully avoids

16 Gödel's second incompleteness theorem is not as simple as this informal treatment suggests, see Berto, 2009, pp. 102–107. The difficulty of proving it arises partly from the fact that Gödel's first incompleteness theorem needs to be translated from its metamathematical expression into the arithmetical language of the formal system itself before it is possible to apply *modus ponens* and arrive at the undecidable proposition. The amount of arithmetic machinery that is necessary for this does not correspond exactly to what is used in the first incompleteness theorem Berto, 2009, p. 102, further complicating the matter. These technical details are not relevant for the following discussion of Wittgenstein's remarks, however.

any notion of truth and is concerned only with the syntactic property of provability (in terms of derivability from the axioms). Gödel's assertion that "the proposition that is undecidable *in the system PM* still was decided by metamathematical considerations" (Gödel, 1986, p. 151) thus presupposes a notion of truth that is only present outside the system. It is therefore no accident that the assertion only occurs in the informal introduction of the paper.

The second incompleteness theorem states that a system of the above nature cannot prove its own ω-consistency (in other words no such proof can be derived as a theorem in the system), as such a proof would immediately imply the undecidable proposition, in contradiction to it being undecidable.

Statements such as "in a formal system there are true but undecidable propositions" or "a formal system cannot prove its own consistency" are not necessarily incorrect, but definitely imprecise characterisations of Gödel's results. Gödel's proof is only applicable to a particular notion of "formal system", the question of whether it is applicable to our ordinary notion of mathematical reasoning and of whether certain philosophical interpretations are justified go beyond its mathematical scope. This is where Wittgenstein finally comes in.

2.2 TRUTH AND PROVABILITY

After a few introductory remarks, the main discussion of Gödel's "proposition that says about itself that it is not provable" in appendix III of the *RFM* starts with Ts-221a, 248.1 / §5[17], where Wittgenstein asks what we call a true proposition in Russell's system:

> Gibt es wahre Sätze in Russell's System, die nicht in seinem System zu beweisen sind? – Was nennt man denn einen wahren Satz in Russell's System? [Ts-221a, 248.1 / *BGM I; Appendix III* §5]

---

17 The introductory remarks are by no means clearly separated from the more explicit discussion of Gödel's result starting in §5 and have often been discussed along with it (e.g. in Floyd, 1995). Some even go so far as to ascribe to them a crucial importance for a proper understanding of Wittgenstein's approach, e.g. Kienzler and Grève, 2016, p. 86: "At a first glance, sections 1–4 and 20 can easily appear to be somewhat disconnected from the bulk of Appendix III. They are, however, of special importance for any adequate understanding of the text because they characterise the perspective from which Wittgenstein approaches the issues." In fact, it is hard to draw a sharp line between Appendix III and the rest of the *RFM*, because "the text of Appendix II can be interpreted as already dealing with Gödel in some sense" (Kienzler and Grève, 2016, p. 86), given that it deals with the issue of surprise in mathematics and that Gödel characterises his proof as demonstrating "surprising results" (Gödel, 1986, p. 151). As the following text intends to show, Wittgenstein's investigation of Gödel is only understandable if it is viewed not as an isolated examination of a particular mathematical theorem (as the early dismissive reviews had the tendency to interpret it), but rather as a reflection of more general philosophical values drawn from the *RFM* and the *PI* as a whole. For the interpretation in this chapter however, §§1–4 are of minor importance and will therefore not be discussed.

> Are there true propositions in Russell's system, which cannot be proved in his system? – What is called a true proposition in Russell's system, then? [*RFM I; App. III* §5]

As mentioned, the truth of propositions is irrelevant for the purely syntactic approach of Gödel's first incompleteness theorem, it is thus not the mathematical result that Wittgenstein is interested in, but rather Gödel's informal remark that "it follows at once that $[R(q); q]$ is true" because it "was decided by metamathematical considerations" (Gödel, 1986, p. 151).

In Ts-221a, 248.2 / §6, Wittgenstein gives a deflationary answer to his own question ("For what does a proposition's *'being true'* mean? *'p' is true* = *p*. (That is the answer.)") and draws our attention to the fact that "in Russell's game" a proposition is true if it occurs as an axiom or "at the end of one of his proofs". This should not be read as an endorsement of a particular theory of truth, but only as the rather trivial observation that, seen purely from 'within' a formal system such as *Principia Mathematica*, truth is defined in terms of provability.[18]

It is clear that Gödel's undecidable proposition is not true in this sense, as it will never appear at the end of a proof in *PM*. But it could certainly be true in a different system (where it would appear at the end of a proof), as Wittgenstein notes in Ts-221a, 248.3 / §7. Of course this observation is quite trivial and does not conflict with Gödel's result at all, so why does Wittgenstein mention it? There seem to be two reasons: First of all, Wittgenstein emphasises the role of truth in the context of different language games, instead of thinking of it as a global notion with the same meaning in every formal system.[19] This is how something can be true or false "in a different sense" depending on the language game in question. In the context of Wittgenstein's philosophy, this is not surprising, but it nevertheless marks a departure from Gödel's 1931 paper, where different notions of truth play no role for the mathematical result. Secondly and more importantly, the trivial observation that false propositions in one system may be true propositions in another (Wittgenstein mentions non-Euclidean

18 In other words, Wittenstein merely adopts the perspective of a Russellian logician for this argument, but does not himself endorse it. As Floyd, 1995, p. 377 notes, this rejection of the "logistic reduction" of mathematical truth goes back to Wittgenstein's *Notebooks 1914–16* as well as the *Tractatus*. See also Floyd, 1995, p. 401: "Read in context, this remark need not be read as an analysis of the truth of sentences generally, i.e., the so-called redundancy theory of truth, but rather perhaps only of those of logic, or, even more, specifically, of the sentences of *Principia*". See also Kienzler and Grève, 2016, p. 95, stressing a similar point.

19 As Floyd, 1995, p. 403 notes, "all the notions in which philosophers have been most interested - "truth", "provability", even "derivability" - find a home only within a specific, ongoing technique of use, and never within a formalism itself." Of course, the hard part is showing *why and how* an overly general conception of these concepts produces misleading interpretations, which is what Wittgenstein sets out to do in the context of his remarks on Gödel.

geometries as examples[20]) draws our attention to the fact that Gödel's undecidable proposition is 'non-trivially' true-but-unprovable, which in Gödel's own words leads to "surprising results concerning consistency proofs for formal systems" (Gödel, 1986, p. 151). Gödel's undecidable proposition is remarkable precisely because the truth of it is not established in an unrelated system, but rather seems to follow directly from the formal system in question itself, just on a meta-systemic level, which nevertheless fully depends on the formal system. As Gödel notes, the undecidable proposition is true, but as Wittgenstein asks, *in what sense*?

After the preparatory work in the aforementioned remarks, the real investigation of Gödel's incompleteness theorem begins in the "notorious paragraph"[21] Ts-221a, 249.2 / §8. It is thus worth quoting in full, while splitting into two parts:

> Ich stelle mir vor, es fragte mich Einer um Rat; er sagt: "Ich habe einen Satz (ich will ihn mit "P" bezeichnen) in Russell's Symbolen konstruiert, und den kann man {durch gewisse Definitionen und Transformationen so deuten, daß er sagt: 'P ist nicht in Russell's System beweisbar'. // auch in der Form aussprechen: 'P ist (in Russell's System) nicht beweisbar'.} Muß ich nun von diesem Satz nicht sagen: einerseits er sei wahr, anderseits er sei unbeweisbar? Denn angenommen, er wäre falsch, so ist es also wahr, daß er beweisbar ist! Und das kann doch nicht sein. Und ist er bewiesen, so ist bewiesen, daß er nicht beweisbar ist! So kann er also nur wahr, aber unbeweisbar sein." [Ts-221a, 249.2 / §8]

> I imagine someone asking my advice; he says "I have constructed a proposition (I will use '*P*' to designate it) in Russell's symbolism, and by means of certain definitions and transformations it can be so interpreted that it says: '*P* is not provable in Russell's system'. Must I not say that this proposition on the one hand is true, and on the other hand is unprovable? For suppose it were false; then it is true that it is provable. And that surely cannot be! And if it is proved, then it is proved that it is not provable. Thus it can only be true, but unprovable." [*RFM I; App. III* §8]

This may be a very brief summary of Gödel's informal assertion that the undecidable proposition is true, but it nevertheless adequately captures why someone might (based on Gödel's result) conclude that the "proposition that says about itself that it is not provable" is true but not provable in *Principia Mathematica*. Perhaps the only departure from most interpretations of Gödel's theorem is how Wittgenstein explicitly notes that Gödel's arithmetisation of metamathematical concepts requires an interpretation, so that "by means of certain definitions and transformations it can be so interpreted", leaving open the question of whether we *must* interpret a formula in this way.[22]

---

20 This should not be read as endorsing some kind of relativism where the axioms of geometry were arbitrary, however. As *WCL*, p. 56 makes clear, Wittgenstein is aware of the fact that Euclidean "Euclidean geometry has a great advantage over non-Euclidean geometry when it is a matter of measuring boards", that is to say, when it comes to the variety of everyday uses.

21 See Floyd and Putnam, 2000.

22 Not all interpretations agree that Wittgenstein's target here is the *informal and prosaic interpretation* of Gödel's theorem. For example, Steiner, 2001, p. 272 denies "that Witt-

Leaving aside this question for now, the the second part of the remark is more problematic and has often been interpreted as an attempted refutation of Gödel's theorem:

> So wie wir fragen: "in welchem System 'beweisbar'?", so müssen wir auch fragen: "in welchem System 'wahr'?". 'In Russell's System wahr' heißt, wie gesagt: in Russell's System bewiesen; und 'in Russell's System falsch' heißt: das Gegenteil sei in Russell's System bewiesen. – Was heißt nun Dein: "angenommen, er sei falsch"? *In Russell's Sinne* heißt es: "angenommen das Gegenteil sei in Russell's System bewiesen"; *ist das Deine Annahme*, so wirst Du jetzt die Deutung, er sei unbeweisbar, wohl aufgeben. Und unter dieser Deutung verstehe ich die Übersetzung in diesem deutschen Satz. – Nimmst Du an, der Satz sei in Russell's System beweisbar, so ist er damit *in Russell's Sinne* wahr und die Deutung "P ist nicht beweisbar" ist wieder aufzugeben. Nimmst Du an, der Satz sei in Russell's Sinne wahr, so folgt das *Gleiche*. Ferner: soll der Satz in einem andern als Russell's Sinne falsch sein: so widerspricht dem nicht, daß er in Russell's System bewiesen ist. (Was im Schach "verlieren" heißt, {kann doch in einem andern Spiel das Gewinnen ausmachen.) // darin kann doch in einem andern Spiel das Gewinnen bestehen.)} [Ts-221a, 249.2 / §8]

> Just as we ask: "'provable' in what system?", so we must also ask: " 'true' in what system?" 'True in Russell's system' means, as was said: proved in Russell's system; and 'false in Russell's system' means: the opposite has been proved in Russell's system. – Now what does your "suppose it is false" mean? *In the Russell sense* it means 'suppose the opposite is proved in Russell's system'; *if that is your assumption*, you will now presumably give up the interpretation that is unprovable. And by 'this interpretation' I understand the translation into this English sentence. – If you assume that the proposition is provable in Russell's system, that means it is true *in the Russell sense*, and the interpretation "*P* is not provable" again has to be

---

genstein saw the introductory remarks as Gödel's own attempt to give a philosophical interpretation of his theorem, and his criticism was of the interpretation." Instead, Steiner reads §8 as an attempted refutation of "a sketch of a perfectly valid, semantic version of Gödel's theorem" (Steiner, 2001, p. 267), and interprets Wittgenstein as mistakenly violating his principle of not interfering in mathematics. Even if it is true that Wittgenstein's description in §8 could be read as a semantic version of Gödel's theorem, such a reading not only needs to assume that Wittgenstein carelessly violated his principle of non-interference, but also deal with the fact that especially the later remarks in Ms-121 and Ms-124 heavily favour the view that Wittgenstein is attacking the *interpretation* of the theorem, not any version of the *mathematical theorem itself*. Furthermore, Steiner's argument makes heavy use of Tarski's theory of truth, when it is quite doubtful that Wittgenstein would have accorded one particular theory such a position, given that his remarks on Gödel all lead to the consequence that a mathematical theory cannot justify its own use, nor its applicability as a picture of our informal concepts. This is convincingly shown in Floyd, 2001, p. 304:

> Wittgenstein's dialectical treatment of Gödel's theorem allows us to raise the philosophical question of whether Tarski's model-theoretic account of truth definitions for formalized languages exhibits the core of our notion of mathematical truth. Wittgenstein would deny that this account resolves the philosophical questions at stake in arguments about the nature and scope of mathematics – even if he accepted, as I suppose he would, Tarski's mathematical theorems, and a Tarskian semantical proof of Gödel's theorem.

Further arguments against Steiner's interpretation are given in Rodych, 2006, where Wittgenstein is however interpreted as a "finitist, formalist, and radical constructivist" (Rodych, 2006, p. 56), endorsing a particular theory of truth (merely one different from Steiner's interpretation). But as this chapter attempts to show, a finitistic reading would swap one dogmatic interpretation for another.

> given up. If you assume that the proposition is true in the Russell sense, *the same* thing follows. Further: if the proposition is supposed to be false in some other than the Russell sense, then it does not contradict this for it to be proved in Russell's system. (What I called "losing" in chess may constitute winning in another game.) [*RFM I; App. III* §8]

What does it mean to "give up the interpretation that it is unprovable"? If we interpret Gödel's arithmetisation as a faithful translation of the concept of provability, the assumption that the undecidable proposition is false in the sense of Russell (and thus provable in *PM*) leads to the contradictory conclusion that it is unprovable in *PM* (and thus true in the sense of Russell), *and vice versa*, since from the perspective of Wittgenstein truth in the sense of Russell is equivalent to provability in *Principia Mathematica*. If truth and provability coincide, it is clear that the only way 'out' is to give up the interpretation that Gödel's translation is "the translation into this English sentence".[23]

This certainly marks a departure from Gödel's own interpretation of his result in the informal introduction, but is it actually a refutation of Gödel's incompleteness theorem? For Gödel, truth and provability are distinct notions, which is a reasonable interpretation if one assumes that Gödel's arithmetisation adequately captures the meaning of the English translation. For Wittgenstein, the situation is reversed, because the equivalence of truth and provability is assumed and the translation can consequently not be said to adequately capture the English meaning. While these two views certainly offer very different perspectives on Gödel's incompleteness theorem, we should be careful not to ascribe a particular theory of truth to Wittgenstein or, even worse, read the above passage as a refutation of Gödel's mathematical result. If there is a disagreement between Gödel and Wittgenstein here at all, it is a disagreement concerning the interpretation of the

23 Rodych, who otherwise proposes a very charitable interpretation of Wittgenstein's remarks on Gödel, considers §8 (and also §10) to contain Wittgenstein's main "mistake" (Rodych, 1999, pp. 181–82): "Gödel's proof of [the first incompleteness theorem] *does not require* that 'P' "be so interpreted that it *says*: 'P is not provable in Russell's system'". We do not *need* to assume a natural language meaning for 'P' (e.g., an *English* meaning) to obtain the threatened contradiction, for it is just a number-theoretic 'fact' that an actual proof of 'P' would enable us to *calculate* the relevant Gödel numbers and thereby arrive at '¬ P' by existential generalization." Rodych is certainly correct in pointing out that Wittgenstein's peculiar and rather informal description of Gödel's theorem operates on the level of informal natural language interpretations and *reductios* unlike Gödel's *mathematical* result, but there is no reason to assume that this would be Wittgenstein's intention. Rodych himself points out that §8 is applicable to Gödel's informal introduction: "Wittgenstein again seems to think that a natural language interpretation of 'P', such as Gödel's original 'the undecidable proposition [R(q); q] states ... that [R(q); q] is not provable', is *essential* to the proof" (Rodych, 1999, p. 183). As later remarks will show, Wittgenstein was aware of the number-theoretic details and the indirect-reference-via-substitution of Gödel's proof, it is therefore likely that Wittgenstein deliberately chose to present an *interpretation* of Gödel's result in terms of a prosaic description, because this kind of interpretation is what he wants to attack, not the mathematical result itself. Consequently, a more adequate interpretation of §8 would be Priest, 2004, pp. 211–13 or Floyd, 1995, pp. 404–406.

proof, not an attempt on the part of Wittgenstein to contest the validity of Gödel's mathematical result. Wittgenstein merely adopts the perspective of truth as seen from within *Principia Mathematica*, but he should not be read as endorsing it, because the investigation was launched in §5 not by dogmatically asserting that truth is provability, but rather by asking "What is called a true proposition in Russell's system, then?". The answer to this is indeed a deflationary account of truth as provability, but only *as seen from within Russell's system*, in other words in a very restricted language game, which by no means precludes the possibility of other sensible notions of truth such as Gödel's. Wittgenstein very deliberately adopts a different perspective than Gödel, but this does not immediately make Wittgenstein a dogmatic proponent of such a perspective.

We might still object, however, that Wittgenstein's perspective is entirely nonsensical. If truth and provability coincide as Wittgenstein assumes (even if only for the sake of the argument), to "suppose it is false" means to "suppose the opposite is proved in Russell's system" or to "assume that the proposition is provable in Russell's system". But in Gödel's proof, provability is understood precisely in terms of its arithmetical translation, in such a way that we can follow the same rules as we usually would while deriving a theorem, only now while transforming numbers instead of mathematical symbols. This is why 'supposing it is false' seems nonsensical: We would have to suppose that, by following only well-defined rules in a mechanical manner, we reach a point which we, by following only these same well-defined rules, cannot reach, if the rules are guaranteed never to lead to an ω-inconsistent state. Wittgenstein's remark appears to misunderstand Gödel by assuming that there is a gap between the arithmetisation of provability and 'regular' provability, so that we could suppose something to be proved 'the regular way' (by following the rules of *Principia Mathematica*) but unprovable when interpreted via its arithmetical translation. The beauty of Gödel's result, however, is that it does not leave such a gap, since it presents an 'implementation' of exactly the rules of a system such as *PM* developed rigorously as recursive functions. How could we then suppose the undecidable proposition P to be false *in Russell's system*?

A tentative answer is given in Ts-221a, 250.3 / §10 after Wittgenstein lays some groundwork in Ts-221a, 250.2 / §9, where he notes that the arithmetic translation of Gödel's undecidable proposition can be interpreted in two ways:

> Was heißt es denn: "P" und "P ist unbeweisbar" seien der gleiche Satz? Es heißt, daß diese *zwei* deutschen Sätze in der und der Notation *einen* Ausdruck haben. [Ts-221a, 250.2 / §9]

> For what does it mean to say that *P* and "*P* is unprovable" are the same proposition? It means that these *two* English sentences have a *single* expression in such-and-such a notation. [*RFM I; App. III* §9]

Strictly speaking, the undecidable proposition should be translated as saying something along the lines of "the proposition with the number r (obtained via substitution) is unprovable", where r is precisely the number that corresponds to this proposition (the proposition "P") via its arithmetical translation, but crucially not by *naming* the number r directly, but rather by substituting a variable in a formula q in such a way that the variable is replaced with the number corresponding to the formula q. Gödel thus constructs a *fixed point* and it is in this very restricted sense that we might be willing to say that "P" and "P is unprovable" are both valid translations for the same mathematical proposition, because once we ask what "P" stands for in "P is unprovable" the answer will (after translating the result of the substitution into English) again be "P is unprovable".[24]

Directly naming the proposition P (with number r) to construct a new proposition r' that says "the proposition with the number r (named directly without substitution)" is of course also possible, the corresponding number would then be larger than the number for r (because the number for r occurs in r'), but the translation could still be said to be "P is unprovable" and in this sense "P" and "P is unprovable" correspond to different expressions in the notation. The method of quoting P therefore matters (and Gödel's construction only works if an indirect quoting-by-substituting is used) and Wittgenstein's quite informal discussion of it is certainly problematic, but not without merit.

Returning to the question of how the seemingly nonsensical provability of P could be assumed, Wittgenstein notes that a human *error* could lead to the proposition P occurring at the end of a proof:

> "Aber P kann doch nicht beweisbar sein, denn, angenommen es wäre bewiesen, so wäre der Satz bewiesen, er sei nicht beweisbar." Aber wenn dies nun bewiesen wäre, oder wenn ich glaubte – vielleicht durch Irrtum – ich hätte es bewiesen, warum sollte ich den Beweis nicht gelten lassen und sagen, ich müsse meine Deutung *"unbeweisbar"* zurückziehen? [Ts-221a, 250.3 / §10]

---

24 Kienzler and Grève, 2016, p. 99 interpret this remark as stressing an important difference between a formal representation of P and its informal expression in English:

> "It becomes apparent, however, that what is expressed here for consideration is the unorthodox view according to which *first* there are two distinct sentences of English, namely on the one hand 'P' and on the other 'P is unprovable', and *then* a notation (or formal system) is constructed in which, for some reason, both these sentences share just *one* formalised counterpart expression (rather than two, e.g. P and Q, as – all things being equal – would seem to be the normal thing to do)."

However, this way of emphasising the difference between the formal notation and its informal counterpart is misleading, because strictly speaking in Gödel's proof 'P' and 'P is unprovable' operate on different levels (and thus have different formal expression), because in 'P is unprovable' the proposition 'P' is never named directly, but is referred to only via its Gödel-number and with the help of substitution. This corresponds to the English expression 'The sentence with the number X is unprovable' where this very sentence occurs as number X in a list of sentences.

> "But surely *P* cannot be provable, for, supposing it were proved, then the proposition that it is not provable would be proved." But if this were now proved, or if I believed – perhaps through an error – that I had proved it, why should I not let the proof stand and say I must withdraw my interpretation "*unprovable*"? [*RFM I; App. III* §10]

It is easy to imagine that we could end up with a 'proof' of P if we forget a negation symbol or add one along the way, we are then faced with a derivation of P, a proposition which, translated into English, says that P is unprovable. What are we to do then? One option would be to take it as a sign of an error and double-check the proof. But what if can find no fault in it and all other proofs lead to the expected results, in other words if we can use the formal system in a useful way? Why should we then not "let the proof stand" and instead "withdraw" our interpretation of it being "unprovable"? After all, we have proved the proposition, so it cannot be unprovable *in this sense*.[25]

2.3 HARMLESS INCONSISTENCY

But there is another interesting possibility left open by the use of the word "perhaps" in remark §10: What if the 'proof' of P really is a proper proof and there is no error? A proof of P would then entail an inconsistency, as P would be provable, but would also (by being true) unprovable. That Wittgenstein indeed leaves behind any assumption of ω-consistency and instead considers a contradictory notion of truth is made clear in the next remark:

> Nehmen wir an, ich beweise die Unbeweisbarkeit (in Russell's System) von P; so habe ich mit diesem Beweis P bewiesen. Wenn nun dieser Beweis einer in Russell's System wäre, – dann hätte ich also zu gleicher Zeit seine Zugehörigkeit und Unzugehörigkeit zum Russell'schen System bewiesen. – Das kommt davon, wenn man solche Sätze bildet. – Aber hier {ist // wäre} ja ein Widerspruch! – Nun so ist hier ein Widerspruch. Schadet er hier etwas? [Ts-221a, 251.2 / §11]

> Let us suppose I prove the unprovability (in Russell's system) of *P*; then by this proof I have proved *P*. Now if this proof were one in Russell's system – I

25 It is interesting that Priest, 2004, p. 225 (note 6) finds "the expression of the last sentence of 10 [to be] a bit odd" and suggests to read it as "But if this were now proved, or if I believed – perhaps through an error – that I had proved it, why should I not let the proof stand and [why] say I must withdraw my interpretation 'unprovable'?". Priest thus interprets the last part of the sentence in the opposite way than most German readers would, as *not* withdrawing the interpretation "unprovable", but rather as 'letting it stand' together with the proof. Such a reading seems highly unlikely in the context of the German sentence and is completely unnecessary for Priest's interpretation: The point of Wittgenstein's remark is precisely that if we let P stand, any interpretation of it is contradictory and so the interpretation "unprovable" can be both withdrawn (because we have just proved the supposedly unprovable proposition) and asserted (because the proved proposition says that it is unprovable). What we have to withdraw in both cases is the expectation of consistency that we had assumed as part of our notion of provability.

> should in that case have proved at once that it belonged and did not belong to Russell's system. – That is what comes of making up such sentences. – But there is a contradiction here! – Well, then there is a contradiction here. Does it do any harm here? [*RFM I; App. III* §11]

This remark, together with Ts-221a, 251.3 / §12 and Ts-221a, 252.1 / §13, exhibits a striking openness towards inconsistency.[26] In itself, this does not conflict with Gödel's result, which is after all conditional: If the formal system in question is ω-consistent, the proposition P is undecidable, but if the system is inconsistent, all bets are off and both P and its negation could be provable.[27]

Inconsistency provides a new perspective in the matter of what it might mean to "give up" or "withdraw" the interpretation that P is unprovable: Giving up this interpretation could either mean that one rejects Gödel's translation of formal concepts such as provability into arithmetic operations (which would amount to a misguided attempt to refute the mathematical validity of Gödel's proof) or it could mean that an ω-consistent concept of provability is simply not powerful enough to fully capture what we mean by provability in the case of propositions such as P. The latter perspective does in no way contradict Gödel's result, it simply departs from Gödel's own interpretation of P as "undecidable *in the system PM* [but metamathematically] still

26 This also seems to provide one of the strongest arguments against the model-theoretic interpretation of Floyd and Putnam, 2000: Even if Wittgenstein had a knowledge of model theory on the level suggested by Floyd and Putnam, the remarks that follow the "notorious paragraph" in Ts-221a, 249.2 / §8 show a strong paraconsistent tendency and do not exhibit much of the model-theoretic subtlety that is supposedly at work in the "notorious paragraph". Of course this does not preclude the possibility that Wittgenstein went on a brief detour in Ts-221a, 249.2/ §8, but it seems more likely that even this earlier paragraph should be read from a paraconsistent perspective, even if such an interpretation might be at odds with the philosophical *Weltanschauung* of most working mathematicians. (Another argument against ascribing the Floyd-Putnam Thesis to Wittgenstein is given in Lajevardi, 2021: If Wittgenstein had indeed intended §8 in such a model-theoretic sense, it is unclear why he did not extend his insight to the *second* incompleteness theorem, where the observation would have been of greater importance.)

27 It is unfortunate that the earliest charitable interpretations of Wittgenstein's remarks on Gödel, Shanker, 1988 and Floyd, 1995, both largely ignore the remarks §§10–13, which represent of course some of the most troubling remarks from the perspective of classical logicians or anyone unable to entertain the idea that inconsistency might be acceptable within a formal system. It is here that paraconsistent interpretations such as Priest, 2004 and Berto, 2009 can shed new light on the remarks, although these interpretations end up going too far, by ascribing a particular position of logic to Wittgenstein (e.g. Priest, 2004, p. 214: "In fact, if one identifies truth with provability, as does Wittgenstein, Gödel's paradox and the liar collapse into each other", and Berto, 2009, p. 213: "As I have claimed before, by subscribing to such a way of making sense of Wittgenstein's remarks on Gödel's Theorem we will not be allowed to claim – as many commentators did – that such a philosophy does not require any logical-mathematical revisionism, being directed only against the foundational demands of philosophers."). As this chapter attempts to show, Wittgenstein's embrace of inconsistency should be taken seriously, but it must not be understood as a dogmatic position, only as an antidote against a philosophically one-sided diet.

[...] decided" in favour of an inconsistent and deflationary notion of truth as provability.[28]

Giving up the interpretation that P is unprovable means just that: Clearly P is not unprovable if we can prove it. So where does that leave the truth of P, a proposition which says about itself that it is not provable? The informal translation of P into English might seem dubious at this point, but it is important to remember that in the case of P there is no pre-existing 'standard of measurement' that would apply here: In the case of 'regular' mathematical propositions inconsistency is a sign of an error, because we do not accept '$2+2=4$' and '$2+2=5$' to be true at the same time, a formal system that would prove both an arithmetic proposition and its negation would thus be useless for practical purposes. But as Wittgenstein reminds us in Ts-221a, 251.3 / §12, in the case of P "the proposition *itself* is unusable, and these inferences equally".[29] Wittgenstein's embrace of inconsistency is not an invitation for trivialism, because a formal system that could prove anything would be useless in practice. It is instead the basis for the very insightful observation that in the case of propositions

---

28 Kienzler and Grève, 2016, p. 101 consider a paraconsistent interpretation of §11 to "misconstrue Wittgenstein's investigations of Gödel in Appendix III" and instead hold the view that "Wittgenstein's point is that, if there is no proof of P in PM, then it is first and foremost questionable whether P should be regarded as a (well-formed) part of PM or not." It is certainly correct that paraconsistent interpretations of Wittgenstein have usually coopted him in the service of a (from Wittgenstein's perspective) dogmatic logical position, but Wittgenstein's discussion of contradictions in Appendix III and other parts of the *Nachlass* should not be too quickly interpreted as excluding such a contradictory proposition from the formal system, no matter how useless the proposition might be. One might follow the interpretation in Kienzler and Grève, 2016 to assert that, while P is indeed unprovable in the sense of Gödel's arithmetised concept of provability, the formalisation of P is not understandable in terms of our English-language concept of provability, but this again raises the question of whether the formalisation of P would be understandable if one leaves behind the formal ideal of consistency (and thus entertains a paraconsistent reading, but only in this particular instance). The important point is that Gödel's proof does not and cannot answer this question, because the usefulness of a proposition such as P is determined by factors outside the mathematical proof itself. It is thus a misinterpretation to say that "Wittgenstein shows in detail that there is no way that the Gödelian construct of a string of signs could be assigned a useful function within (ordinary) mathematics" (Kienzler and Grève, 2016, p. 76). Wittgenstein certainly gives us good reasons to say that P is useless, but this does not mean that it could not be assigned a useful function (possibly through the help of a paraconsistent formal system).

29 The remark bears a striking resemblance to Wittgenstein's remarks on Cantor's diagonal argument, most importantly Ms-121, 41v.2 / §35 (see Section 1.3), where Wittgenstein emphasises that the diagonalised number constructed in Cantor's proof is not used for the same practical purposes as other real numbers. This resemblance is no coincidence, as Gödel's proof is at its heart a diagonal argument and the construction of Gödel's undecidable proposition shows many analogies to Cantor's diagonalised number. In both cases, the practical purposes of such a construction matter little to mathematicians, whereas this aspect is of paramount importance for Wittgenstein, because the question of whether the resulting contradiction could be accepted instead of excluded depends on how we use the language game in practice.

such as P, which are fixed points constructed through diagonalisation, there is no practical use and therefore no pre-existing standard for the importance of consistency. Of course we might be tempted to apply the same standard of consistency and the same concept of provability that governs the 'regular' cases, for formal reasons or otherwise, but are we sure that we are not extending the concept beyond its reach? Coincidentally, this is exactly the phrasing that Wittgenstein uses in an earlier variant of Ts-221a, 250.3 / §10 in Ms-118:

> "Aber P kann doch nicht beweisbar sein, denn angenommen es wäre bewiesen so wäre der Satz bewiesen, er sei nicht beweisbar." Aber wenn dies nun bewiesen wäre, oder wenn ich glaubte – vielleicht {durch einen Irrtum – // irrtümlich –} ich hätte es bewiesen, warum sollte ich den Beweis nicht {{als solchen anerkennen & sagen, // gelten lassen & sagen,} ich habe meine Deutung: "unbeweisbar" zu {weit // sehr} ausgedehnt? // sagen, ich müsse meine Deutung ... zurückziehen? } [Ms-118, 111r.4]

> "But P cannot be provable, for if it were proved, the proposition that it would not be provable would be proved." But if this were now proved, or if I believed – {perhaps by an error // erroneously} – I had proved it, why should I not {{acknowledge the proof as such & say, // allow it to stand & say,} I have extended my interpretation: "unprovable" too {far // much}? // say I have to withdraw my interpretation ...? }

Of course one should not interpret Wittgenstein as necessarily endorsing an inconsistent concept of provability here. What Wittgenstein's investigation appears to aim at is an openness for alternative philosophical interpretations of Gödel's result, in other words for a different and less one-sided diet. As Ts-221a, 251.3 / §12 and Ts-221a, 252.1 / §13 make clear, consistency is not an ideal in itself for Wittgenstein and we therefore cannot simply assume a formal system to be consistent or ω-consistent if the system is supposed to reflect our understanding of provability, because the 'motivating examples' of provable or unprovable useful propositions simply do not apply in the case of the undecidable proposition P, as we have not given it any use.

The next remark goes further into this direction, by considering how a proof of the unprovability (and undecidability) of a particular proposition can be used as a *prediction*. The remark is a central point in the investigation of Gödel's theorem in *RFM I; App. III* and thus worth quoting in full, as it introduces a reflection on the "physical element" of a proof:

> Ein Beweis der Unbeweisbarkeit ist quasi ein geometrischer Beweis; ein Beweis, die Geometrie der Beweise betreffend. Ganz analog einem Beweise etwa, daß die und die Konstruktion nicht mit Zirkel und Lineal ausführbar ist. Nun enthält so ein Beweis ein Element der Vorhersage, ein physikalisches Element. Denn als Folge dieses Beweises sagen wir ja einem Menschen: "Bemüh' Dich nicht, eine Konstruktion (der Dreiteilung des Winkels, etwa) zu finden, – man kann beweisen, daß es nicht geht." Das heißt: es ist wesentlich, daß sich der Beweis der Unbeweisbarkeit in dieser Weise soll anwenden lassen. Er muß – könnte man sagen – für uns ein *triftiger Grund*

> sein, die Suche nach einem Beweis (also einer Konstruktion der und der Art) aufzugeben.
>
> Ein Widerspruch ist als eine solche Vorhersage unbrauchbar. [Ts-221a, 252.2 / §14]

> A proof of unprovability is as it were a geometrical proof; a proof concerning the geometry of proofs. Quite analogous e.g. to a proof that such-and-such a construction is impossible with ruler and compass. Now such a proof contains an element of prediction, a physical element. For in consequence of such a proof we say to a man: "Don't exert yourself to find a construction (of the trisection of an angle, say) – it can be proved that it can't be done". That is to say: it is essential that the proof of unprovability should be capable of being applied in this way. It must – we might say – be a *forcible reason* for giving up the search for a proof (i.e. for a construction of such-and-such a kind).
>
> A contradiction is unusable as such a prediction. [*RFM I; App. III* §14]

Apart from being a succinct explanation of what Wittgenstein has in mind when he speaks of a "geometrical proof", the remark is interesting due to how it draws a line between a Euclidean proof that can be used to practically predict what can and cannot be done on the one hand and an impossibility proof proceeding via *reduction ad absurdum* on the other. But can Gödel's proof not also be used as a prediction that the undecidable proposition P cannot be proved in a ω-consistent formal system? Is that not a very practical result that holds for all systems that meet certain requirements?

Of course Gödel's proof shows that *under the assumption of* ω*-consistency* we cannot prove proposition P or its negation and there is no reason to read Wittgenstein as objecting to this. But crucially and in contrast to a Euclidean proof with compass and straightedge, ω-consistency leads us to *reject* any candidate for a proof of P (or its negation), because accepting such a candidate would also entail accepting a proof of its negation. In other words, it is not that we lack the methods to construct a proof of P simply by following the rules of the system, as would be the case in a Euclidean proof, but rather that what we construct by following the rules is explicitly excluded by another rule (the assumption of ω-consistency).[30]

---

30 This is also emphasised by Rodych, who reminds us that Gödel's result cannot be used as a prediction because the mathematical proof is conditional: "What must be borne in mind (again and again) is that Gödel proves the conditional proposition 'If *this* system is consistent, then 'P' is not derivable' - he does *not* prove that the system is consistent" (Rodych, 1999, p. 190). This is an important disagreement with Floyd, 1995 and Floyd, 2001, where no interpretation of the last sentence of §14 is provided (Floyd, 2001, pp. 292–293 considers the last sentence to be "enigmatic" and an "afterghought", sidestepping its interpretation by focusing on the "more telling remark" in Ms-118, 113r.2 instead). Although Rodych's discussion of Floyd's interpretation is otherwise not particularly charitable (Rodych, 1999, pp. 191–193), he correctly points out that the last sentence of §14 leads to "a rock-solid disagreement" between the two interpretations. The interpretation presented in this chapter follows Rodych in this instance, by distinguishing between Euclidean proofs (with predictive power) and Gödel's *conditional* proof (which gains its predictive power only in combination with a philosophical interpretation that presupposes ω-consistency).

We could object here that the assumption of ω-consistency is just as much a rule in the formal system as the laws of Euclidean geometry. But what Wittgenstein seems to be driving at is that any requirement of consistency is not a "forcible reason" to give up our search for a proof of P inside the formal system, because the aversion to contradictions is, at least in the specific case of P, only a formal requirement, without sufficient grounding in our practice. A contradiction simply shows that we can prove 'too much', not just the proposition itself, but also its negation, but why not "let the proof stand", as Wittgenstein suggested before?

The above remark (with its emphasis on the "physical element" of a proof, the "element of prediction" and the uselessness of a contradiction as a prediction) introduces in a very condensed form some of the most central aspects of Wittgenstein's reflections on Gödel, many of which are expanded upon in Wittgenstein's Whewell's Court Lectures from 1938. The discussion of Gödel's proof in the lectures is by no means very detailed (and spans less than 10 pages of the published lecture notes), but is worth including here:

> We call a thing a proof of improvability if we can say: 'Don't try.' If people seemed to find it, who is to say they only seemed ⟨to find it⟩ or ⟨to⟩ say they found it? We would have to make a decision.
>
> Suppose we say (1) 'Assume P is provable.' We have done this by fixing a proof of provability. We assume, by (1), a provability proof. We assume we have a provability proof, and thereby we assume $P \wedge \neg P$. This doesn't contradict common sense; this is not a kind of proposition used in common life.
>
> The assumption of Gödel that $\Pi(P)$[31] is not provable without a provability proof is useless. [*WCL*, p. 53][32]

As Wittgenstein notes, the (metamathematical) proof of the unprovability of P is useless as a prediction, because we might decide to let the proof of P stand and accept the contradiction. But this "decision" is only open to us because a proposition such as P is not used in the same way as 'regular' propositions: A 'regular' proposition such as "it is raining" would be useless if both it and its negation could be proved at the same time, because we draw practical conclusions from it. But we do not draw any (practical) conclusions from (diagonalised) propositions such as P, so a contradiction could potentially be harmless in this situation. The decision is up to us, but only in this very special case. A proposition such as P is in other words an *exception* and we could treat it as such:

> You could say: 'You contradict the law of contradiction – $\neg(p \wedge \neg p)$.' – 'All right, then do what you like, e.g. make an exception, etc.' Or else, Mr Russell can say: 'I won't call $\Pi(P)$ a provability proof.' [*WCL*, p. 53]

31 This is the notation used by Wittgenstein in his lectures to signify the arithmetical translation of 'P is not provable'.

32 The Russellian notation of logical symbols here and in all other quotations has been replaced with the symbols commonly used today, "Π" has additionally been replaced by "Π(P)", as suggested by a footnote from the editors of the *WCL*.

This is a central point of disagreement between the philosophical interpretations of the incompleteness theorem by Gödel and Wittgenstein. Gödel wants us to see that the undecidable proposition P ist a proposition of arithmetic just like any other regular proposition that we might usually want to prove or disprove, because all of the propositions in the formal systems discussed by Gödel share a uniform way of constructing them, including proposition P. But Wittgenstein wants to emphasise that *despite the uniform construction*, which he certainly does not want to deny, the propositions are used in entirely different ways and we might therefore distinguish between them on the basis of their application. Proposition P is a "queer" proposition: "We could say it was not an improvability proof in some sense, or say it led to a contradiction: a queer proposition has led to a queer proposition – well, what about it?" (*WCL*, p. 54)

At the end of the lectures on Gödel, Wittgenstein gives one of the clearest explanations of what he perceives as misleading about Gödel's argument (or rather, about Gödel's prosaic exposition of his mathematical result), which is worth quoting in full, split into two parts:

> Gödel draws a line between logic and mathematics. We all have an idea what mathematics or geometry is. This doesn't contain the idea that all mathematics can be derived from a few primitive propositions.
>
> Gödel says, whatever primitive propositions we start with, we can always construct a mathematical proposition which cannot be derived from these primitive propositions. [*WCL*, p. 55]

It is important to remember that Wittgenstein himself does not see mathematics as simply reducible to logic, but for entirely different reasons than Gödel:

For Wittgenstein, a proof or calculation has to be *surveyable*, which is often not the case when mathematical proofs are translated into the uniform language of logic. We might mistakenly believe that the logical translation could exist independently from its 'informal' mathematical counterpart, whereas from Wittgenstein's perspective an unsurveyable translation into logic is useless in a way that it ceases to be a proof or a calculation, because it loses its ability to act as a *rule* or standard of measurement in our language games.

For Gödel, there 'are' propositions of arithmetic which 'escape' (ω-consistent) logic by being undecidable within the formal system, but can be proved to be true metamathematically. This, however, is misleading, according to Wittgenstein:

> If you say: 'There are mathematical propositions which are neither provable nor disprovable', this is extremely misleading. It suggests, 'Only *God* knows.' We can always say there will be propositions which in one system are provable, in others are not. There are no propositions true outside a system – in regions we cannot comprehend, etc.
>
> We confuse mathematical propositions with some kind of propositions of physics. 'If we only could see, we would then see the truth about it', 'There are always higher and higher truths to which we cannot come', which

> means nothing whatsoever. Whatever propositions are true will in another system be false. [*WCL*, p. 55]

The misleading aspect of Gödel's interpretation is that it suggests the existence of true propositions outside of any formal system. The truth of these propositions appears to hold outside of any *logic*, with the propositions ceasing to be propositions of logic. Instead, they seem to be propositions of physics (or rather 'ultraphysics') in the platonic realm of metamathematics. This is of course entirely nonsensical from the perspective of Wittgenstein, who rejects the idea of dogmatic platonism and of truth independent from any particular formal system or language game.[33]

Again, this should not be read as a refutation of Gödel's proof, merely as Wittgenstein's reminder that Gödel's own interpretation of P as metamathematically true goes beyond his own proof (where incompleteness is a conditional result that depends on ω-consistency).[34] There may be many good reasons to indeed presuppose ω-consistency in our mathematical reasoning, in fact even as a purely formal requirement it might be sufficient to convince people to give up any search for a proof of P within a formal system, but Wittgenstein correctly points out that a contradiction is a weaker reason to give up such a search than in many other situations and that some people might not be convinced by the philosophical conclusions about truth and provability drawn from Gödel's proof.[35]

Returning to *RFM I; App. III*, Wittgenstein slightly broadens his focus after §14 and considers *criteria* for unprovability, echoing his remark from the early 1930s ("Sieh auf die Beweise und entscheide dann, was sie beweisen." in Ms-154, 49v.2; Ms-113, 108r.4; Ts-211, 682.3; Ts-212, 1579.2; Ts-213, 632r.3) that only the proof of a proposition shows its sense:

> Ob etwas mit Recht der Satz genannt wird "X ist unbeweisbar", hängt davon ab, wie wir diesen Satz beweisen. Nur der Beweis zeigt, was als das Kriterium der Unbeweisbarkeit gilt. Der Beweis ist ein Teil des Systems von Operationen, des Spiels, worin der Satz gebraucht wird, und zeigt uns seinen 'Sinn'.

---

33 For an exploration of how this relates to the idea that "only *God* knows", see item A, item A and item A.

34 This is also Shanker's interpretation (albeit without emphasising Wittgenstein's openness towards inconsistency), see Shanker, 1988, p. 234: "Here is a case in point of the subtlety with which Wittgenstein felt philosophy must draw its conclusions. Often all that is called for is a modest change in point of view; alter the focus slightly and Godel's proof will not have lost any of its significance: only the platonism will have been abandoned."

35 These people would for example include proponents of paraconsistent logic, who do not accept $\neg(p \land \neg p)$ as unconditionally true, but do not advocate trivialism either. Wittgenstein should not be read as dogmatically endorsing the view that 'there are' true contradictions, merely as reminding us that there might be languages games where the law of non-contradiction is not part of the rules of the game.

> Es ist also die Frage ob der "Beweis der Unbeweisbarkeit' von p hier ein triftiger Grund ist zur Annahme daß ein Beweis von p nicht gefunden werden wird. [Ts-223, 252.3 / §15]

> Whether something is rightly called the proposition "*X* is unprovable" depends on how we prove this proposition. The proof alone shews what counts as the criterion of unprovability. The proof is part of the system of operations, of the game, in which the proposition is used, and shews us its 'sense'.
>
> Thus the question is whether the 'proof of the unprovability of *P*' is here a forcible reason for the assumption that a proof of *P* will not be found. [*RFM I; App. III* §15]

Obviously the "proof of the unprovability of P" is not a "forcible reason" if we give up the assumption of ω-consistency, because then the "proof of the unprovability of P" would be what P says, but the derivability of P would also be a proof of the provability of P. The proposition is thus provable and unprovable in a different sense, because its unprovability is proved by what the proposition 'says', by its truth, whereas the provability of P is proved by P being the "terminal pattern in the proof of unprovability" that we have derived, as Wittgenstein further explains in the next remark:

> Der Satz "p ist unbeweisbar" hat einen andern Sinn, nach dem – als ehe er bewiesen ist.
>
> Ist er bewiesen, so ist er die Schlußfigur des Unbeweisbarkeitsbeweises. – Ist er unbewiesen, so ist ja noch nicht *klar*, was als Kriterium seiner Wahrheit zu gelten hat, und sein Sinn ist – kann man sagen – noch verschleiert. [Ts-223, 253.1 / §16]

> The proposition "*P* is unprovable" has a different sense afterwards – from before it was proved.
>
> If it is proved, then it is the terminal pattern in the proof of unprovability. – If it is unproved, then *what* is to count as a criterion of its truth is not yet *clear*, and – we can say – its sense is still veiled. [*RFM I; App. III* §16]

When we prove P by letting the contradictory proof stand, we see that it is not provable or unprovable in the sense that we had in mind before we proved it, but rather in a different sense: Before the proof, we assumed that a proposition would be either provable or unprovable, but not both, in analogy to how we use the concept of provability for 'regular' (that is to say *non-diagonalised*) propositions. When we see the proof, we realise that accepting a proof of P would also entail accepting an interpretation of it as unprovable (the paraconsistent option) or alternatively reject both its provability and unprovability, which excludes the proposition from the game of provability and considers it to be undecidable (the Gödelian option).

On the other hand, Gödel's "metamathematical considerations" are of course also a proof of the unprovability of P, now carried out in a 'higher-level' system than the formal system in question itself, which leads Gödel to conclude in his introduction that P is unprovable, but true. This seems to be the sort of proof of unprovability that Wittgenstein has in mind in the first paragraph of remark §17, contrasting it to the direct proof that is only available in an ω-inconsistent formal system:

> Wie, soll ich nun annehmen, ist P bewiesen? Durch einen Unbeweisbarkeitsbeweis? oder auf eine andere Weise? Nimm an, durch einen Unbeweisbarkeitsbeweis. Nun, um zu sehen, *was* bewiesen ist, schau auf den Beweis! Vielleicht ist hier bewiesen, daß die und die Form des Beweises nicht zu P führt. – Oder, es sei P auf eine direkte Art bewiesen – wie ich einmal sagen will –, dann folgt also der Satz "P ist unbeweisbar", und es muß sich nun zeigen, wie diese Deutung der Symbole von P mit der Tatsache des Beweises kollidiert und warum sie hier aufzugeben sei. [Ts-221a, 253.2 / §17]

> Now how am I to take *P* as having been proved? By a proof of unprovability? Or in some other way? Suppose it is by a proof of unprovability. Now, in order to see *what* has been proved, look at the proof. Perhaps it has here been proved that such-and-such forms of proofs do not lead to *P*. – Or, suppose *P* has been proved in a direct way – as I should like to put it – and so in that case there follows the proposition "*P* is unprovable", and it must now come out how this interpretation of the symbols of *P* collides with the fact of the proof, and why it has to be given up here. [*RFM I; App. III* §17]

Wittgenstein spells out these different senses of provability in the next paragraph of the remark. Most of it is only a succinct summary and culmination of the previous remarks, but Wittgenstein also clearly mentions that we end up in a situation where we have to *decide* between two options: We can either accept the proof of P (derived within the formal system) and revise our concept of the unprovability claimed by P to include inconsistency, or we can reject the proof of P and "still call [the proposition P] the statement of unprovability", thereby continuing to use the concept of provability that we had in mind before the 'proof':

> Angenommen aber, ¬P sei bewiesen. – *Wie* bewiesen? Etwa dadurch, daß P direkt bewiesen ist – denn daraus folgt, daß es beweisbar ist, also ¬P. Was soll ich nun aussagen: "P", oder "¬P"? Warum nicht beides? Wenn mich jemand fragt: "Was ist der Fall: P, oder {nicht-P // ¬P}?" so antworte ich: "⊢ P"" steht am Ende eines Russellschen Beweises, Du schreibst also im Russellschen System: "⊢ P"; anderseits ist es aber eben beweisbar & dies drückt man durch ⊢ ¬P aus. Dieser Satz aber steht nicht am Ende eines Russellschen Beweises gehört also nicht zum Russellschen System. – Als die Deutung "P ist unbeweisbar" für P gegeben wurde, da kannte man ja {den // einen} Beweis für P nicht & man {kann // muß} also nicht sagen "P" sage: *dieser* Beweis existierte nicht. – Ist der Beweis {konstruiert // hergestellt}, so ist damit eine *neue Lage* geschaffen: Und wir haben uns nun zu entscheiden, ob wir *dies* einen Beweis (*noch* einen Beweis) oder ob wir *dies* noch die Aussage der Unbeweisbarkeit nennen wollen.
>
> Angenommen ¬P sei direkt bewiesen; es ist also bewiesen, daß sich P direkt beweisen läßt! Das ist also wieder eine Frage der Deutung – es sei denn, daß wir nun auch einen direkten Beweis von P haben. Wäre es nun so, nun, so wäre es so. –
>
> (Sehr komisch ist die abergläubische Angst & Verehrung der Mathematiker vor dem Widerspruch.) [Ms-118, 114v.2–115v.2 / §17][36]

---

36 The typescript versions of this remark in Ts-221a, 253.2 and Ts-223, 253.2 omit all logical symbols, which is why the version from Ms-118 is quoted here instead.

> Suppose however that not-*P* is proved. – Proved *how*? Say by *P*'s being proved directly – for from that follows that it is provable, and hence not-*P*. What am I to say now, "*P*" or "not-*P*" then I reply: *P* stands at the end of a Russellian proof, so you write *P* in the Russellian system; on the other hand, however, it is then provable and this is expressed by not-*P*, but this proposition does not stand at the end of a Russellian proof, and so does not belong to the Russellian system. – When the interpretation "*P* is unprovable" was given to *P*, this proof of *P* was not known, and so one cannot say that *P* says: *this* proof did not exist. – Once the proof has been constructed, this has created *a new situation*: and now we have to decide whether we will still call *this* a proof (a *further* proof), or whether we will still call *this* the statement of unprovability.
>
> Suppose not-*P* is directly proved; it is therefore proved that *P* can be directly proved! So this is once more a question of interpretation – unless we now also have a direct proof of *P*. If it were like that, well, that is how it would be.
>
> (The superstitious dread and veneration by mathematicians in face of contradiction.) [*RFM I; App. III* §17[37]]

If Wittgenstein's remarks are intended as a criticism of Gödel's incompleteness theorem, then only in the sense that Gödel's own informal interpretation has the tendency to obscure that *there is a decision to be made* and that we cannot point to the mathematical proof in defense of the option that we ultimately choose, because both options are equally compatible with Gödel's mathematical result.

This is also why Gödel's prosaic interpretation can lead us to believe that there is a "physical element" in the proof and that it can be used to predict that there *are* undecidable propositions. This immediately suggests a platonic picture of propositions as independent entities whose truth can sometimes only be grasped outside their own formal system, through metamathematical considerations. But as Wittgenstein points out, it is ultimately our decision to *call* these propositions 'undecidable' and to reject any proof of them from the language game of provability (within the formal system).

As Wittgenstein points out in Ts-221a, 254.2 / §18, we often play language games where rules as fundamental as the law of non-contradiction do not hold, for example when we answer a question with "yes and no". It is doubtful that Gödel would disagree with this rather obvious fact, but it nevertheless shows once again that Wittgenstein is investigating Gödel's incompleteness theorem from a very unorthodox perspective that is interested primarily in how we use concepts such as provability in a variety of language games.

The following remark is one the most provocative in the *RFM*, as it might appear to equate Gödel's fundamental result with an entirely useless châlet in the Baroque style on Mount Everest, with the immediate association with a 'Luftschloss' / 'castle in the air' certainly intended:

37 See Kienzler and Grève, 2016, pp. 107–109 for a discussion of several problematic decisions in Anscombe's translation. For the sake of consistency, Anscombe's translation has not been altered here.

> Du sagst: "....." also ist P wahr und unbeweisbar." Das heißt wohl: "Also ⊢ P." Von mir aus – aber zu welchem Zweck schreibst Du diese 'Behauptung' hin? (Das ist, als hätte jemand aus gewissen Prinzipien über Naturformen und Baustil abgeleitet, auf den Mount Everest, wo niemand wohnen kann, gehöre ein Schlößchen im Barockstile.) Und wie könntest Du mir die Wahrheit der Behauptung plausibel machen, da Du sie ja zu nichts weiter brauchen kannst als zu jenen Kunststückchen? [Ts-221a, 255.1 / §19]

> You say: "..., so *P* is true and unprovable". That presumably means: "Therefore *P*". That is all right with me – but for what purpose do you write down this 'assertion'? (It is as if someone had extracted from certain principles about natural forms and architectural style the idea that on Mount Everest, where no one can live, there belonged a châlet in the Baroque style. And how could you make the truth of the assertion plausible to me, since you can make no use of it except to do these of legerdemain? [*RFM I; App. III* §19]

On closer inspection, however, it becomes clear that the useless architectural châlet is not Gödel's result at all, but only the undecidable proposition P. Wittgenstein explicitly speaks of the "purpose" of "this 'assertion'" and of the inability to "make the truth of the assertion plausible", which is not in any way an indictment of Gödel's incompleteness theorem, but rather a very insightful observation about the undecidable proposition constructed by Gödel. In contrast to other propositions, which may be used to refer to an external state of affairs (such as when we infer q from $p \to q$ and p by *modus ponens* and let p correspond to "It is raining" and q to "I will get wet"), the undecidable proposition P refers (through an indirect substitution) only to itself and to nothing outside the formal system. It cannot be used in the same way as other propositions and is in this sense merely the product of "certain principles". If it has any use at all, it is only as an instrument in Gödel's proof, because it almost parasitically depends on the precise specification of the formal system and cannot be separated from it, as every slight change in the specification of the rules would lead to an entirely different undecidable proposition P.[38]

This is why the undecidable proposition does not convey any "information", it cannot be used as a picture of the world (to borrow a Tractarian terminology). This distinguishes such an undecidable proposition from propositions of logic, because the undecidable proposition is nonsensical (at least from an ω-consistent perspective), whereas tautologies and contradictions are senseless, "for the one allows every possible state of affairs, the other none" (*TLP* 4.462). An undecidable proposition such as P is thus a "proposition-like structure of another kind",[39] as Wittgenstein notes in the next remark:

38 This demonstrates another resemblance to Cantor's diagonal argument, since Cantor's diagonalised number is similarly useless outside the diagonal proof. We cannot use it to calculate anything practical, because its only purpose is to act as an instrument in the proof.

39 Anscombe translates it as "sentence-like structure of another kind", which is a dubious choice in a context where the aim is to emphasise the similarity to propositions of logic.

> Man muß sich hier daran erinnern, daß die Sätze der Logik so konstruiert sind, daß sie als *Information keine* Anwendung in der Praxis haben. Man könnte also sehr wohl sagen, sie seien garnicht *Sätze*; und daß man sie überhaupt hinschreibt, bedarf einer Rechtfertigung. Fügt man diesen 'Sätzen' nun ein weiteres satzartiges Gebilde andrer Art hinzu, so sind wir hier schon erst recht im Dunkeln darüber, was dieses System von Zeichenkombinationen nun für eine Anwendung, für einen Sinn haben soll, denn der bloße *Satzklang* dieser Zeichenverbindungen gibt ihnen ja eine Bedeutung noch nicht. [Ts-221a, 255.2 / §20]

> Here one needs to remember that the propositions of logic are so constructed as to have *no* application as *information* in practice. So it could very well be said that they were not *propositions* at all; and one's writing them down at all stands in need of justification. Now if we append to these 'propositions' a further sentence-like structure of another kind, then we are all the more in the dark about what kind of application this system of sign-combinations is supposed to have; for the mere *ring of a sentence* is not enough to give these connexions of signs any meaning. [*RFM I; App. III* §20]

It is unlikely that Wittgenstein's wants to deny that the undecidable proposition P has a use inside Gödel's proof, instead the remark clarifies that there is no extra-mathematical application that would correspond to its intra-mathematical and purely formal use. Wittgenstein thus emphasises the differences between propositions and proposition-like structures, whereas most mathematicians, including Gödel, prefer to emphasise their uniform constructibility as formulas of the system, similar to how Cantor's diagonal argument leads to a uniform extensionalist treatment of all real numbers in terms of their decimal expansions.

Before focusing on Wittgenstein's later remarks on Gödel, the end of the remarks in *RFM I; App. III* offer an opportunity to briefly review the most relevant aspects of Wittgenstein's critique of Gödel's (first) incompleteness theorem: Most importantly, Wittgenstein's remarks are targeted less at the mathematical result itself and more at Gödel's own prosaic interpretation of the undecidable proposition P as 'true but unprovable'. In contrast to Gödel, for whom (ω-)consistency is an implicit assumption in his own philosophical interpretation, Wittgenstein investigates an (ω-)inconsistent interpretation for which provability (in the formal system) and truth (in the formal system) coincide. This should not be read as an endorsement of inconsistent logic, but only as a reminder that Gödel's mathematical result offers us a conceptual *decision*. The mathematical result cannot in itself provide justifications for one option over another, because both interpretations are compatible with Gödel's incompleteness theorem. Any decision can only be justified by the *application* of the formal system, which is why Wittgenstein is interested in how we *use* undecidable propositions such as P. Being a result of diagonalisation, these propositions depend on the exact specification of the formal systems, they are thus of a different kind than the propositions we normally encounter in our use of formal systems, a difference that is emphasised by Wittgenstein.

2.4 INCONSISTENCY AND USE

Wittgenstein returns to Gödel's incompleteness theorem at the end of December 1938, in a passage from Ms-121, 72r.3–85r.2 that is clearly delineated from (but also strongly influenced by) the reflections on Cantor's diagonal argument in the rest of the document. The remarks on Gödel were almost entirely excluded from *RFM*,[40] which is understandable given their sometimes unfinished appearance and their role as a 'detour' from the rest of *RFM II*. Wittgenstein did not include any of these remarks in later manuscripts or typescripts, they should therefore be considered merely as first drafts, with no plans to eventually publish them (and all the caveats for interpretation that this entails). They are nevertheless an interesting reflection of Wittgenstein's thought process regarding Gödel's theorem and should thus be included in an interpretation of Wittgenstein's views on the incompleteness theorem, although it should be pointed out that only a selection of remarks from the passage will be discussed here.

Wittgenstein begins his reflections on Gödel in Ms-121 by considering the status of the law of non-contradiction in logic. As a fundamental principle of logic, the law seems to hold unconditionally:

> Aber gilt also der Satz {vom Widerspruch // des Widerspruches} nicht? [Ms-121, 72r.3]
>
> ---
>
> But does the law of non-contradiction not hold then?

However, there are many occasions where a contradiction is not immediately rejected, as Wittgenstein points out shortly after. More importantly, when a contradiction is rejected, it is not rejected as *false*, but as a *nonsensical pseudo-proposition*, because it has no *use* in our language games:

> Aber überall weist man ja den Widerspruch nicht zurück. Es gibt (ja) Gelegenheiten, wo wir den Satz gelten lassen & wo wir für den Satz Verwendung haben: es verhalte sich so, & doch nicht so.
>
> Auch wird der Widerspruch nicht zurückgewiesen als eine falsche Mitteilung, sondern als Unsinn, als Scheinsatz, als etwas wofür in unsern Sprachspielen kein Gebrauch ist. [Ms-121, 72v.2–73r.2]
>
> ---
>
> But one does not reject the contradiction everywhere. There are (after all) occasions when we allow the sentence to stand & where we have use for the sentence: it behaves like this, & yet not like this.
>
> Nor is the contradiction rejected as a false statement, but as nonsense, as a sham, as something for which there is no use in our language games.

---

40 Only Ms-121, 74v.2 was published as *RFM IV* §60.

But if a contradiction is rejected because it has no use, could we not give it a use in a language game, perhaps as a "rare" or "exotic" curiosity of a formal system? The obvious objection (which is a valid objection in the case of classical logic) would be that anything can be inferred from a contradiction, in other words trivialism would result and the formal system would immediately become useless. This is not the case in a paraconsistent logic, however, where a contradiction is not immediately 'explosive' and not everything can be inferred from it. Even though Wittgenstein does not explicitly propose such a logic, he certainly recognises that *inferring nothing* (contra classical logic) could be an option:

> Warum sollte die {Russellsche // symbolische} Logik nicht zu einem Widerspruch führen dürfen? {Warum // Ja, warum} {sollte man dieses nicht als die seltenste Blume dieses Systems empfinden. // sollte man in diesem nicht die seltenste Blume dieses Systems sehen? // sollte man in diesem nicht eine exotische Blume dieses Systems sehen?}
>
> "Aber aus einem Widerspruch folgt ja *jeder* Satz! Was würde dann aus der Logik?"
> Nun so folgere nichts aus einem Widerspruch!
>
> Wenn Mathematiker sich abergläubisch vor dem Widerspruch wie vor dem leibhaftigen Teufel gebärden, warum sollten nicht andere eine Art schwarze Messe feiern (&) sich in Widersprüchen ergehen? [Ms-121, 74r.2–74r.4]

> Why should the {Russellian // symbolic} logic not be allowed to lead to a contradiction? {Why // Yes, why} {should one not see this as the rarest flower of this system. // shouldn't one see in this the rarest flower of this system? // shouldn't one see in this an exotic flower of this system?}
>
> "But from a contradiction follows *any* proposition! What would then become of logic?"
> Well so infer nothing from a contradiction!
>
> If mathematicians behave superstitiously before contradiction as before the devil incarnate, why should not others celebrate a kind of black mass (&) indulge in contradictions?

As these remarks make clear, Wittgenstein does not want to prescribe a particular logic. Instead, he wants to clarify that we can imagine language games that go beyond the restrictive picture painted only by classical logic (which should not be read as a value judgement about the usefulness or uselessness of classical logic).

According to Wittgenstein, a contradiction in logic is not dangerous in and of itself. Instead, a contradiction becomes dangerous only when it can occur in regions where we did not expect it, such as when we proceed by natural deduction from consistent assumptions and suddenly arrive at a contradiction such as "it is raining and it is not raining". Here, a contradiction is a sign of an error and throws into question our whole logical argument, with the result that we cannot trust any of the results that we draw from the argument.

But if the contradiction does not result from faulty reasoning and is instead only the expression of an inherently inconsistent and useless part of our language game (such as in the case of self-referential statements along the lines of "This sentence is false."), the contradiction need not lead to any harm, as long as we do not use to infer arbitrary propositions as in classical logic:

> Nicht das // dies ist ein Unglück // perniziös: einen Widerspruch zu erzeugen {in der Region, in der // dort, wo} weder der widerspruchsfreie noch der widerspruchsvolle Satz {eine // irgend welche} Arbeit zu leisten hat; wohl aber das: nicht zu wissen, {wo man in diese Region eingetreten ist // wie man dorthin gekommen ist // wo man in diese Region gekommen ist wo der Widerspruch nicht mehr schadet}.
>
> Frag nicht: "Ist p wahr, oder falsch?, sondern: "Soll ich schreiben '⊢ p', oder '⊢ ¬p'?" – Und darauf wird {oft // manchmal} die Antwort sein: "Das kommt darauf an, was Du mit dem Satz machen willst". [Ms-121, 74v.2–75r.2 / *RFM IV* §60 (first remark only)]
>
> ---
>
> The pernicious thing is not: to produce a contradiction in the region which neither the consistent nor the contradictory proposition has any kind of work to accomplish; no, what *is* pernicious is: not to know how one reached the place where contradiction no longer does any harm.
>
> Ask not, "Is p true, or false?" but, "Should I write '⊢ p', or '⊢ ¬p'?" - And to that {often // sometimes} the answer will be: "That depends on what you want to do with the proposition". [*RFM IV* §60 (first remark only)]

In these special cases, where a contradictory proposition does not conflict with the normal use of propositions, we are free to *decide* what we do with the contradiction. We could exclude it from the language game, or we could give it a use:

> Erinnere Dich *hier* Deiner Freiheit, möchte ich sagen, zu gehen, wie Du willst.
>
> Und heißt das nicht: Verstehe, was Dich sonst gebunden hat & daß Du also hier frei bist? [Ms-121, 75r.3–75r.4]
>
> ---
>
> Remember *here* your freedom, I want to say, to go as you will.
>
> And doesn't that mean: Understand what has otherwise bound you & that you are therefore free here?

It is no accident that the next remark explicitly mentions Gödel, because Gödel's undecidable proposition P, the "proposition that says about itself that it is not provable", grants us this freedom. It does not have a mathematical use comparable to other propositions, so why does Gödel feel the need to exclude the resulting contradiction from the formal system and considers it to be *metamathematically true*?

> "Ja, soll ich diesen Satz (Gödels z.B.) anerkennen, oder nicht? –"
>
> {Worin besteht // Was heißt} es denn: einen Satz {anzuerkennen // anerkennen}?
>
> "Es ist eine besondere geistige Handlung." Nun, dann interessiert es mich {hier // jetzt} nicht. Erkenne ihn nur immer an, wenn Du dazu Zeit & Lust hast! – Aber redet man nicht davon, daß man einen Satz mit der Tat – oder nur mit dem Mund, anerkennt? Nun das bringt uns schon näher, {daran, zu sehen // läßt uns sehen // es läßt uns sehen // zu sehen // zu erkennen }

> was es mit dem Anerkennen {eines Satzes // der Wahrheit des Satzes} für eine Bewandtnis hat.
>
> [Setze statt der Gefühle (Gebärden) der Anerkennung: was Du mit dem Satz tust.][Ms-121, 75v.2–76r.2]

> "Well, shall I recognise this proposition (Gödel's, for example) [as true], or not? -"
>
> {What does it consist of // What does it mean}: to recognise a proposition [as true]?
>
> "It is a special mental act." Well, then I don't care about it {here // now}. Just recognise it [as true] whenever you have time & feel like it! - But doesn't one talk about recognising a proposition [as true] by deed - or only by mouth? Well, {this already brings us closer to seeing // it makes us see // it makes us see // to see // to recognise } what recognising {a proposition [as true] // the truth of the proposition} is about.
>
> [Replace the feelings (gestures) of recognition with: what you do with the sentence].

In the case of 'regular' propositions, we can easily name the reasons that led us to recognise a proposition as true: We might point to the axioms of a formal system, explain that we recognise them as obviously true, and demonstrate that the rules of inference are truth-preserving, by checking that what we derive corresponds to our expectations. If we then recognise a propositions such as "it is raining" as true, we will choose to act differently based on it than if we had recognised "it is not raining" as true. In the case of the undecidable proposition P, we can also point to the axioms and the rules of inference, but we cannot check that the truth of P corresponds to our expectations and it is unclear how we might act differently on P. Instead of asking whether we should recognise proposition P as true (which we might do simply in analogy with 'regular' propositions), which does not give us any insight into the "freedom" involved in the decision of how to treat P, it is thus better to investigate the use of the proposition.

However, the question of how we use a proposition cannot be answered by mathematics alone, at least not by a mathematical argument such as Gödel's proof. It has to be answered through an investigation of a variety of language games, because the use of propositions is not *uniform*. The situation is comparable to the variety of numbers and the different systems that they form, each with different properties. If we restrict out attention to the natural numbers or the rationals, we can describe them as *a single system* where all numbers share essential properties (for example by defining the natural numbers as repeated applications of a successor operation or the rational numbers as signed fractions of natural numbers), but if we extend our scope to include all real numbers, we quickly see that they are characterised by a variety of heterogeneous uses with no essential property that would describe numbers as different as for example π and diagonalised numbers. It is of course possible to treat them uniformly as decimal expansions, but such a view is only possible 'after the fact', it cannot tell us why we would want to use a numbers such as π or how to arrive at this particular number.

In this way, Gödel's undecidable proposition P demonstrates that a conception of mathematics as a single system is unclear, similar to what Cantor's diagonal argument shows for the real numbers:

> Gödel zeigt uns eine Unklarheit im Begriff (der) 'Mathematik', die darin zum Ausdruck kam, daß man die Mathematik für ein *System* gehalten hat.
>
> Die '*Eigenschaft einer Zahl*' – wie schaut das aus? Ich vermute– – –.
>
> Wenn wir ein *System* mathematischer Sätze haben, so hat dies {selbst eine // seine eigene} Geometrie. [Ms-121, 76r.3–76r.5]
>
> ---
>
> Gödel shows us an ambiguity in the concept (of) 'mathematics', which was expressed in the fact that mathematics was thought to be a *system*.
>
> The 'property of a number' - what does that look like? I presume- - -.
>
> If we have a *system* of mathematical propositions, then this {itself has a geometry // has its own geometry}.

The above remarks can be read as preliminary summary of Wittgenstein's remarks on Gödel. Some of the details in Ms-121 differ from Wittgenstein's earlier remarks in Ts-221a/b / *RFM I; App. III*, but the overall direction has so far been similar. However, the next remarks in Ms-121 introduce a new element that was not present in Wittgenstein's earlier reflections, namely a closer examination of *how* Gödel's undecidable proposition P manages to (indirectly) speak of itself and what this means for its interpretation and use.

2.5 SELF-EVIDENT CONTRADICTIONS

The remarks in Ms-121 make clear that Wittgenstein was at least somewhat familiar with the mathematical details behind Gödel's proof, because in the following remarks he does not write about directly self-referential propositions, but rather about formulas ("Satzzeichen") that cannot be derived inside the system via substitution by following rules of inference ("kann nicht durch die Operationen ... erhalten werden"), where the formula is referred to only indirectly with the help of its (translation into a) number:[41]

> "Dieses Satzzeichen ist 25 cm lang." "Dieses Satzzeichen kann nicht durch die Operationen ... erhalten werden."
>
> "Das Satzzeichen № 512 kann nicht durch die Operationen ... erhalten werden."
> Die Frage ist: wie rechne ich aus, daß dieses Satzzeichen das 512te ist. [Ms-121, 76v.1–76v.2]
>
> ---

41 This view is shared by Rodych, 2002, p. 380, who interprets the remark as showing "Wittgenstein's genuine understanding of the mathematical nature of Gödel's proposition".

> "This propositional sign is 25 cm long." "This propositional sign cannot be obtained by the operations ..."
>
> "The propositional sign № 512 cannot be obtained by the operations ..."
> The question is: how do I calculate that this propositional sign is the 512th.

These details correspond quite closely to Gödel's mathematical proof. In other remarks, it might seem as if Wittgenstein brushed aside the details of translating formulas into numbers and interpreting numbers as propositions, but the above remark is a first hint that Wittgenstein is quite interested in exactly these details, even if he investigates them from a philosophical and not a mathematical perspective. In fact, Wittgenstein flips around the criticism and accuses mathematicians of carelessly glancing over these seemingly incidental questions of interpretation and application:

> Eine der {verderblichsten // peinlichsten} Unklarheiten ist die der Mathematiker über das, was sie – *jetzt* halb verächtlich – die '*Interpretation*' {der // ihrer} Zeichen nennen. Unter 'Interpretation' oder 'Auffassung' stellt man sich irgendwelche uns nicht interessierende psychologische Vorgänge vor, die die {Worte // Zeichen} begleiten, während die Interpretation eines Zeichens in seiner Anwendung liegt.
>
> Die Bedeutung eines Zeichens liegt von seltenen Fällen abgesehen nicht in seelischen Vorgängen, die sein Aussprechen, Schreiben, etc. begleiten sondern in der komplizierten, uns aber geläufigen, Praxis seiner Verwendung. [Ms-121, 77v.3–78r.2]
>
> ---
>
> One of the most {pernicious // embarrassing} obscurities is that of mathematicians about what they - *now* half-scornfully - call the '*interpretation*' {of // of their} signs. By 'interpretation' or 'conception' one imagines any psychological processes of no interest to us that accompany the {words // signs}, whereas the interpretation of a sign lies in its application.
>
> The meaning of a sign lies, except in rare cases, not in psychological processes that accompany its utterance, writing, etc., but in the complicated, but familiar, practice of its use.

From the perspective of a mathematician, there are two kinds of interpretations involved in Gödel's incompleteness theorem:

Gödel's method of translation from formulas to numbers allows us to interpret numbers as mathematical and even metamathematical propositions, which is of course central to Gödel's proof and of great importance to mathematicians. Gödel's particular proof is just one among many possible 'encodings', but it is beyond dispute that his translation captures the axioms and rules of inference of the formal systems in question, because Gödel's translation 'merely' shows us how we can manipulate numbers in place of logical symbols while adhering to the formal rules of the system in question.

There is also the question of how the purely mathematical concepts are to be interpreted and used in our practice, which is often an afterthought for mathematicians or at least considered to be of lesser importance, since such questions seem to be matters of psychological processes: For example, what is the psychological process associated with recognising something as true? Or the psychological process involved in asserting a proposition? These questions appear to be questions that should be answered by psychologists, but not mathematicians.

Wittgenstein's sometimes rather critical remarks on the unclear interpretation of Gödel's result must then appear to be either mathematically wrong (if he is read as attacking Gödel's translation of formulas into numbers) or missing the point (if he is read as attacking the purely psychological matters involved in any extra-mathematical interpretation). But Wittgenstein wants to emphasise that what misses the point is actually the simplistic mathematical picture of these two distinct kinds of interpretations, because even Gödel's own informal interpretation goes far beyond what his mathematical proof can provide (by presupposing a particular concept of truth that excludes inconsistency). Gödel ignores that the concept of provability used in a formal system such as *Principia Mathematica* is in practice not applied to propositions such as the undecidable proposition P and that is is therefore not clear whether *in this particular instance* it makes sense to regard his formalisation as an adequate translation of what we mean when we say that something is provable or unprovable. This is not a question of psychological processes, however, but of the use of certain concepts in a variety of mathematical language games.

In the following remarks, Wittgenstein experiments with informal propositions that show a resemblance to Gödel's proposition P, but where the concepts provability and truth are replaced by other concepts. These remarks are some of the least developed remarks in all of Ms-121 and often read like preliminary notes instead of fully formed philosophical thoughts, but they nevertheless contain interesting reflections concerning the negation of undecidable propositions such as P, which is why a selection will be discussed here.

Wittgenstein starts by discussing the proposition "This proposition is not a tautology", with 'is a tautology' acting as a semantic counterpart to Gödel's purely syntactic 'is provable'. More interesting is the aspect that Wittgenstein introduces in the last part of the remark, of a proposition being immediately recognisable as tautologic ("man sieht es ihm ja gleich an. // es ist ihm ja gleich anzusehen"):

> "Dieser Satz ist keine Tautologie." 'Dieser Satz kann keine Tautologie sein & er kann nicht falsch sein, denn . . . ' (Siehe Gödel)
>
> Argumentieren wir so: nehmen wir an dies wäre eine Tautologie, so gäbe es also eine Tautologie, die von sich selbst aussagte, sie sei *keine*. Und dann sagt sie doch nicht die Wahrheit.
>
> {"Aber das könnte doch ohnehin niemand glauben, daß der Satz eine Tautologie ist" // "Aber Dein Satz kann doch ohnehin keine Tautologie sein, {man sieht es ihm ja gleich an. // es ist ihm ja gleich anzusehen.}"} – Ich nehme an, jemand hatte // habe einen Rechenfehler gemacht, & ich kann ja einen beliebig dummen Rechenfehler annehmen. Es ist unbegreiflich –, aber er hat herausgebracht, daß der Satz eine Tautologie ist. [Ms-121, 78v.3]

> "This proposition is not a tautology." 'This proposition cannot be a tautology & it cannot be false, because ...' (See Gödel).
>
> Let's argue like this: suppose this were a tautology, then there would be a tautology that said of itself that it was *none*. And then it doesn't state the truth after all.
>
> {"But nobody could believe that the proposition is a tautology anyway." // "But your proposition can't be a tautology anyway, it's obvious from the look of it"} - I assume that somebody had made an arithmetical error, & I can assume an arbitrarily stupid arithmetical error. It is inexplicable - but he has brought out that the proposition is a tautology.

Wittgenstein continues this line of thought a few remarks later, when he introduces the notion of propositions being "immediately obvious" and "self-evident" ("unmittelbar einleuchtend" and "selbstverständlich"):

> "Dieser Satz ist nicht unmittelbar einleuchtend."
>
> Wie wenn ein Mensch wenn auch fälschlich vom Satz T sagen würde: "Nun, das ist eine offenbare Tautologie!" – Was machst Du? – "Nun, das ist doch selbstverständlich, daß das keine Tautologie ist!"
>
> "Dieser Satz ist nicht selbstverständlich." – Wie sollen wir uns zu diesem Satz stellen? Sollen wir sagen, er sei wahr? falsch? selbstverständlich? – 'Du mußt sagen: er sei wahr, aber nicht {selbstverständlich // , er sei selbstverständlich-wahr}. Denn . . . ' [Ms-121, 79v.4–80r.2]

> "This proposition is not immediately obvious."
>
> As if a person were to say, albeit falsely, of the proposition T, "Well, that's an obvious tautology!" - What do you do? - "Well, it is self-evident that that is not a tautology!"
>
> "That proposition is not self-evident." - How are we to take a stand on this proposition? Shall we say it is true? false? self-evident? - 'You must say: it is true, but not {self-evident // self-evidently-true}. Because ...'

Wittgenstein thus distinguishes between a proposition being "true" (or tautological) and a proposition being "self-evident" in a way that is apparently meant to correspond to Gödel's distinction (in his informal introduction) between a proposition being (metamathematically) true and a proposition being (within the system) unprovable.[42]

The beauty of Wittgenstein's reformulation lies in the fact that at least for propositions of logic (which as tautologies can be immediately recognised as true independently from any verification in the external world) truth and self-evidence normally coincide, exactly in the same way as truth and provability might have been assumed to coincide prior to Gödel's incompleteness theorem. Similar to how Gödel

42 Wittgenstein's remarks do not appear to be directly targeted at Gödel's undecidable proposition here, instead they seem to be intended as a detour that demonstrates a certain resemblance to Gödel's proposition but acts only as an intermediate link in a surveyable representation. Rodych, 2002, p. 382, in contrast, reads Wittgenstein's mention of self-evidence as referring to Gödel's undecidable proposition ("Surely we will not say that the Gödelian proposition, complex as it is, is 'self-evident!'"), but comes to a conclusion that appears nevertheless to be compatible with the interpretation of the remark offered here.

constructs an undecidable proposition for which these two notions no longer coincide (if we adopt Gödel's own informal interpretation), Wittgenstein constructs a semantic counterpart where truth and self-evidence become distinct.

Faced with an analogous contradiction as in the case of Gödel, we might be lead to declare the proposition (following Gödel) to be "true – not self-evident". But what does that mean? From Wittgenstein's perspective, it can seems as if mathematicians view undecidable propositions of this sort as true "for its own sake", because the truth of the proposition does not correspond to any use in practice. As Wittgenstein wrote before, we are therefore free to decide differently and accept the contradiction inside the system:

> Angenommen nun, Du gibst {es zu: // mir nach:} er sei wahr – nicht selbstverständlich – – – was hast Du da zugegeben?
>
> Du hast *den* Satz zugegeben. (Aber wie macht man das?) {Du sprichst ihn nun mit dem Ton der Überzeugung aus, lehrst Andere, es tun, nickst mit dem Kopf & sagst: "das stimmt" // Du sprichst ihn also mit dem Ton der Überzeugung aus; sagst: "das stimmt" & nickst mit dem Kopf; (&) lehrst Andere {dies // dasselbe} tun.} {Oder {sollten wir sagen: // will der Mathematiker sagen:} wir lieben die Wahrheit um ihrer selbst willen? // Oder entgegnet der Mathematiker: er liebe ... // Oder sagt der Mathematiker, es handelt sich nicht um Vorteil & Nachteil:}
>
> Aber welchen Nachteil hätte es hier gehabt, zu sagen: der Satz sei selbstverständlich, das käme aber hier auf das gleiche hinaus, als ihn, der scheinbar das Gegenteil sagt, behauptend auszusprechen. Wir hätten also hier einen äußerlichen Widerspruch; {aber es sei alles in Ordnung. // aber {unter den besonderen Umständen // durch die besonderen Umstände} sei alles in Ordnung. // aber, durch die Besonderheit der Aussage, sei ...} [Ms-121, 80v.2–81r.2]

> Now suppose you {admit it: // give in:} it is true - not self-evident - - what have you admitted?
>
> You admitted the proposition. (But how does one do that?) {You now say it with a tone of conviction, teach others to do it, nod your head & say: "that is true" // So you say it with a tone of conviction; say: "that is true", nod your head; (&) teach others to do {this // the same thing}}. {Or {should we say: // does the mathematician want to say:} we love truth for its own sake? // Or does the mathematician reply: he loves ... // Or does the mathematician say, it is not a matter of advantage & disadvantage:}
>
> But what disadvantage would it have had here to say: the proposition is self-evident, but that would amount to the same thing here as to say it assertively, which apparently says the opposite. So here we would have an outward contradiction; {but everything is all right. // but {under the particular circumstances // by the particular circumstances} everything is all right. // but, by the particularity of the statement, ...}

Once again, Wittgenstein emphasises that we are faced with *special circumstances* when deciding whether to accept the contradictory proposition within a formal system or instead exclude it as undecidable. It is here that his disagreement with the philosophical interpretation of Gödel is most clearly expressed, because Gödel does not view the undecidable proposition as being fundamentally different from other propositions in the system. On the contrary, Gödel treats all propositions uniformly: According to him, undecidable propositions such as P are "relatively simple problems in the theory of integers" (Gödel, 1986, p. 145), suggesting that propositions such as P are not fundamentally different from the 'regular' mathematical proposition involving integers that we might want to solve for practical purposes.

In Gödel's defence, he cannot distinguish these different kinds of propositions purely on the basis of his mathematical result (or at least he cannot justify leaving behind the mathematical ideal of consistency), because Wittgenstein's distinction between propositions such as P and 'regular' propositions hinges on the fact that we do not use a proposition such as P for anything useful. Only then are we free to decide whether or not to let the contradiction stand, because the 'freedom of movement' of P is not restricted by any existing language games.

Gödel's decision (of choosing consistency and thereby excluding the proposition P from the system) stems from the formal ideal of consistency, not from any regard for how we use concepts such as provability in practice. From the perspective of Wittgenstein, this is a "stupid reason"[43], which comes down to "keeping up appearances"[44], as Wittgenstein writes in one of the most succinct and clear remarks on Gödel's incompleteness theorem in the whole *Nachlass*, summarising all of the central aspects of Wittgenstein's investigation up until this point:

> "Aber zum Teufel, er ist selbstverständlich, oder nicht selbstverständlich!" – Die Wahrheit ist, daß Du zu so etwas normalerweise nicht "selbstverständlich!" sagst, noch es behauptest, noch sein Gegenteil. Du hast vor allem gar nicht den geringsten Gebrauch für so einen Satz. Und dränge ich Dich nun doch, Dich zu entscheiden, ob Du ihn anerkennen {wirst // willst}, etc., {so sollst Du sehen, daß es hier ganz gleichgültig ist wie Du Dich entscheidest, daß also hier die gewöhnliche {Entscheidung // Wahl} nicht vorliegt // so sollst Du sehen, daß es hier die gewöhnlichen Entscheidungsgründe nicht gibt // so sollst Du sehen, daß hier die gewöhnliche Situation der Entscheidung nicht vorliegt.} Ich möchte beinahe sagen: wofür immer Du Dich entscheidest, entscheide Dich nicht aus dem Gödelschen Grund, denn das ist ein dummer Grund. Ich wollte (lieber), Du hättest den Mut hier {etwas offenbar Unsinniges // einen offenbaren Unsinn} zu sagen, {statt daß Du vor *dieser* Konsequenz zurückscheust. // statt daß Du hier noch die Formen wahrst.} [Ms-121, 81r.3]

> "But hell, it is self-evident, or not self-evident!" - The truth is that you do not normally say "of course!" to such a thing, nor do you assert it, nor its opposite. Above all, you have not the slightest use for such a proposition. And if I now urge you to decide whether you {will // want to} recognise it, etc., {you shall see that here it is quite indifferent how you decide, that here the ordinary {decision // choice} is not present. // you shall see that

43 See Rodych, 2002, p. 384: "Wittgenstein's point here is that *a mere belief* in the consistency of PA is an exceedingly weak - indeed, stupid - reason for accepting the proposition in question as true and unprovable in PA."

44 Rodych, 2002, p. 385 notes the similarity of this charge to Wittgenstein's earlier remark about Cantor's diagonal argument in Ms-117, 109.3 / *RFM II* §21.

> here the ordinary situation of the decision is not present.} I almost want to say: whatever you decide, do not decide on the basis of Gödel's reason, because that is a stupid reason. I wish you (rather) had the courage to say something obviously nonsensical here, {instead of shying away from *this* consequence. // instead of still keeping up appearances here.}

After this summarising remark, Wittgenstein moves on to consider a new aspect of Gödel's proof that was not previously mentioned in any of the remarks in Ts-221a/b / *RFM I; App. III* or Ms-121: Which metamathematical translation corresponds to the *negation* of Gödel's undecidable proposition P? In the case of P itself, the metamathematical translation is "Proposition P is not provable" with P being the *this very proposition* (while keeping in mind, as usual, that the proposition does not directly refer to P but only indirectly via substitution). This is why Gödel himself called it the "proposition that says about itself that it is not provable" (Gödel, 1986, pp. 149–151) and why we might be tempted to translate P using a reflexive pronoun as "This proposition is not provable". At first glance, we might then naively expect the negation of P to translate to "This proposition is provable", but this is not the case, because the negation corresponds to the metamathematical proposition "Proposition P is provable", with P *still referring* to the proposition "Proposition P is not provable". In other words, the negation of P is not reflexive in the same way as P itself is, which is why we cannot simply negate the reflexive translation itself, but have to resort to the translation "The proposition 'This proposition is not provable' is not true". We have thus encountered a situation where a naive attempt to negate the informal interpretation of Gödel's proposition P might fail to match the mathematically precise and formal version of the negation of P, a situation where prose can quickly become misleading and we have to look at the proof to understand what we mean by 'negation'. Wittgenstein is well aware of this:

> Wie lautet denn das Gegenteil des Satzes "Dieser Satz ist nicht selbstverständlich"? So: "{Dieser // Der} Satz ist selbstverständlich"? Aber wenn hier "dieser" wieder reflexiv ist, dann ist es ja nicht das Negativ des oberen.
>
> Wenn ich dies ausspreche & Du willst es leugnen mußt Du bereit sein zu sagen: "Was {Du sagst // er sagt} ist falsch; {es // er} ist selbstverständlich." –
>
> Soll ich sagen, das Gegenteil lautet: "Der Satz: '{Dieser // Der} Satz ist nicht selbstverständlich.' *ist* selbstverständlich"; oder etwa: "Der Satz: '{Dieser // Der} Satz ist selbstverständlich' ist nicht wahr.".
>
> "Gödels Satz sagt in *indirekter* Weise aus, daß er nicht beweisbar ist." – Was sagt also der verneinte Gödelsche Satz aus? [Ms-121, 82r.1–82v.1]

> What is the opposite of the proposition "This proposition is not self-evident"? Is it: "{This // The} proposition is self-evident"? But if "this" is reflexive here again, then it is not the negative of the upper one.
>
> If I say this & you want to deny it you must be prepared to say, "What {you say // it says} is false; {this // it} is self-evident." -

> Shall I say the opposite is: "The proposition: '{This // The} proposition is not self-evident.' *is* self-evident"; or rather, "The proposition: '{This // The} proposition is self-evident' is not true."
>
> "Gödel's proposition states in an *indirect* way that it is not provable." - So what does the negated Gödelian proposition say?

We can write the different ways of (properly or naively) negating the undecidable proposition P as follows, while keeping in mind that the reference to 'itself' only happens indirectly via substitution:

1. $\underbrace{\neg\mathit{Provable}(p)}_{p}$

   "This prop. is not provable."
   (Gödel's undecidable proposition P)

2. $\neg\underbrace{\neg\mathit{Provable}(p)}_{p}$

   "The prop. 'This prop. is not provable.' is not true."
   ("Der Satz: 'Dieser Satz ist selbstverständlich' ist nicht wahr.")

3. $\mathit{Provable}(\underbrace{\neg\mathit{Provable}(p)}_{p})$

   "The prop. 'This prop. is not provable.' is provable."
   ("Der Satz: 'Dieser Satz ist nicht selbstverst.' *ist* selbstverst.")

4. $\underbrace{\neg\neg\mathit{Provable}(p')}_{p'} \equiv \underbrace{\mathit{Provable}(p')}_{p'}$

   "This prop. is provable."
   ("Dieser Satz ist selbstverständlich")

5. $\neg\underbrace{\neg\neg\mathit{Provable}(p')}_{p'} \equiv \neg\underbrace{\mathit{Provable}(p')}_{p'}$

   "The prop. 'This prop. is provable' is not true."
   (Negation of 4, analogous to 2)

6. $\neg\mathit{Provable}(\underbrace{\neg\neg\mathit{Provable}(p')}_{p'}) \equiv \neg\mathit{Provable}(\underbrace{\mathit{Provable}(p')}_{p'})$

   "The prop. 'This prop. is provable' is not provable."
   (Negation in terms of provability of 4, analogous to 3)

The correspondence to Gödel's or Wittgenstein's own wordings is given in parentheses.[45] Proposition 1 corresponds to Gödel's undecidable proposition P, proposition 2 is its (proper and equally undecidable) negation (corresponding to Wittgenstein's "Der Satz: '{Dieser

45 Wittgenstein himself seems to have experimented with a very similar style of diagrams in the loose-page fragment Ms-178c, most likely written in 1939, which

// Der} Satz ist selbstverständlich' ist nicht wahr."), proposition 3 corresponds to Wittgenstein's "Der Satz: '{Dieser // Der} Satz ist nicht selbstverständlich.' *ist* selbstverständlich". The propositions 4–6 correspond to the naive negation of P (proposition 4), its negation (proposition 5) and a negation in terms of provability (proposition 6, analogous to proposition 3).

Apart from drawing our attention to the fact that a prosaic interpretation of negation requires more care in the case of diagonalised propositions such as P, Wittgenstein's remarks lead to two interesting observations about Gödel's proof:

First of all, as Wittgenstein himself mentions, the proper negation of P is proposition 2, not proposition 3. But what is the difference in meaning between proposition 2 and 3? The beauty of Gödel's purely syntactic approach is that he sidesteps any discussion of the meaning of truth, with P itself saying only that it is not *provable*. But the negation of P, proposition 2, cannot be interpreted purely in terms of provability (this would be proposition 3), it has to be interpreted as saying that P is "not true". But as Wittgenstein alludes to in the remark quoted above, do we know what we mean by this without introducing a notion of truth that is carefully avoided in Gödel's own proof?

Secondly, why not focus briefly on the 'naive' negation of P, proposition 4, which corresponds to the interpretation "This proposition is provable"? Gödel himself does not consider such a proposition in his proof, but we might be tempted to ask whether it should be provable (or perhaps its negation). Gödel's undecidable proposition is undecidable because *either* it or its negation would lead to (ω-)inconsistency if provable, but for its naively-negated counterpart the opposite is the case: Either proposition 4 or its negation, proposition 5, could be provable and *neither* of them would lead to inconsistency (as long as only one of them is provable, of course). In other words, if we are asked to choose proposition 4 or 5 as a provable theorem in our system, either choice would be justified by and in accordance with the very proposition itself, whereas in the case of Gödel's undecidable proposition P or its negation (propositions 1 and 2 above) neither choice would be justified by and in accordance with the chosen proposition.

In contrast to a 'regular' proposition, for which we can give reasons why it (or its negation) is provable, the diagonalised propositions 1 and 4 are special: For Gödel's undecidable proposition P, *no choice* will be acceptable unless we accept inconsistency, whereas for its naively-negated counterpart *any choice* will be acceptable and the reason for it purely tautological.[46]

---

revolves around Cantor's diagonal argument and contains several drawings of diagonalisations. On page Ms-178c, 4, Wittgenstein writes "Ristunbew. = f(R)" and draws an arrow from this expression back to R, indicating its self-reference.

46 Priest calls these situations "overdetermined" and "underdetermined", see Priest, 2006b, p. 15. Wittgenstein himself discusses an underdetermined variant of Cantor's

The subtlety involved in the different ways of negating P or propositions similar to it shows once more that propositions such as P, which refer indirectly to their own number-translation through diagonalisation, "confront us with a new situation":

> Gödel konfrontiert uns mit einer neuen Situation: "was sollen wir nun *dazu* sagen?"
>
> Aber in der Entscheidung, was man sagen solle, darf man nun nicht vorschnell sein. (Besonders darf man nicht gleich das sagen wollen, was am Aufsehenerregendsten klingt.) Die Situation ist schwerer zu übersehen, als es scheint. [Ms-121, 84r.2–84r.3]
>
> ---
>
> Gödel confronts us with a new situation: "what shall we now say about *this*?"
>
> But in deciding what to say, one must not be hasty. (Especially one must not want to say immediately what sounds most sensational.) The situation is more difficult to survey than it seems.

Gödel implicitly assumes that even diagonalised propositions are to be treated analogously to 'regular' propositions and is of course mathematically justified in presenting a formalisation that does just that. When the purely mathematical result is interpreted metamathematically, it leads to the "sensational" result that there 'are' propositions that are true, but not provable. But are we sure that this treatment of diagonalised propositions is philosophically justified? Does it correspond to what we mean when we say that a proposition is true? Or do we lack an understanding of what truth means in the case of diagonalised propositions and are mislead into thinking that we have a clear picture due to the superficial analogy with regular propositions?

It should be pointed out that Wittgenstein ends his remarks on Gödel in Ms-121 by mentioning that the situation created by Gödel's result is "harder to survey than it appears". The concept of surveyability has so far played only a minor role in Wittgenstein's remarks on Gödel, but it will come to the forefront in the remarks in Ms-124 / *RFM VII*, which will be discussed next.

2.6 SURVEYABILITY AND DIAGONALISATION

The remarks on Gödel in Ms-124 stem from 1941 and form Wittgenstein's last extensive discussion of Gödel's proof. The first explicit mention of Gödel occurs on page 84, but the relevant remarks begin at least two pages before, with Wittgenstein reflecting on his philosophical method in a way that also applies to Gödel's result:

> Fordere nicht zuviel, & fürchte nicht, daß Deine gerechte Forderung in's Nichts zerrinnen wird.
>
> Meine Aufgabe ist es nicht, Russells Logik von *innen* anzugreifen, sondern von außen.

---

diagonal argument in the context of Turing's own use of the diagonal method, see Section 1.1 for a detailed discussion.

> D.h.: nicht, sie mathematisch anzugreifen – sonst triebe ich Mathematik – sondern ihre Stellung, ihr Amt. [Ms-124, 82.1–82.3 / *BGM VII* §19]

> Don't demand too much, and don't be afraid that your just demand will dwindle into nothing.
>
> It is my task, not to attack Russell's logic from within, but from without.
>
> That is to say: not to attack it mathematically – otherwise I should be doing mathematics – but its position, its office. [*RFM VII* §19]

Based on the latter two remarks, Wittgenstein's guideline of "not demanding too much" might be read as follows: Any *mathematical* critique of Gödel's proof by a philosopher such as Wittgenstein would be an 'unjust' demand and Wittgenstein would be "afraid" that his critique would later turn out to miss the mathematical point of Gödel's proof. But a philosophical investigation in the sense of Wittgenstein, which leaves the mathematical proof itself as it is and is occupied only with the *philosophical interpretation* of the proof, is not impacted by any mathematical discovery, therefore there is no need to be "afraid" that a new mathematical discovery could invalidate the philosophical investigation.[47]

The point is made more explicit if we remember that Gödel's result is usually understood as a *limitative* result: For certain (ω-consistent) systems, it is impossible to decide all formulas in the system (the first incompleteness theorem) and consequently impossible to prove the consistency of the system inside the system (the second incompleteness theorem). If Wittgenstein's critique were mathematical, it would focus on whether Gödel's proof is in fact a limitative result, similar to what would have happened if Gödel's proof had contained a mathematical error and another mathematician had later proved that it is in fact possible to decide all propositions in an ω-consistent system of the form Gödel has in mind. Whether or not the "demand" is just would then depend on something mathematical. But Wittgenstein does not intend to refute the limitative result itself, his aim is only to investigate what *kind* of limitation is being proved here: Gödel's proof is often interpreted (even by Gödel himself in the introduction)

---

47 Wittgenstein's remark on being "afraid" shows similarities to remarks in Ms-117 on the fear of contradictions. In the first of these (Ms-117, 246.2 / *RFM III* §85), Wittgenstein asks why we are not afraid of the possibility of an impossible Euclidean construction (such as a regular heptagon), whereas we are afraid of the possibility of constructing a contradiction. Shortly after (in Ms-117, 253.4 / *RFM III* §87), Wittgenstein notes that the fear of contradictions in formal systems and of certain Euclidean construction need not be dissimilar after all, because a fear of contradictions is an *indeterminate fear* and can disappear entirely if "it is enough for me to get a proof that a contradiction or a trisection of an angle cannot be constructed in *this* way." This is directly related to the importance of surveyability, because no consistency proof can calm our fear if the system is not surveyable and we have no understanding of how a contradiction would affect the system. A contradiction can be entirely harmless and a consistency proof might be unnecessary if we have a clear picture of how the contradiction is to be used, we can then decide whether we need to exclude it or let the contradiction stand.

as having an almost physical element (or rather "ultraphysical", in Wittgenstein's words), whereas from the perspective of Wittgenstein the proof (like all mathematical proofs) can only show a *logical* impossibility.

Consequently, Wittgenstein's critique is not focused on the mathematical system itself, he wants to attack the system not "from within, but from without". His remarks are focused on an aspect that is only of secondary importance to mathematicians such as Gödel, namely the variety of language games that give meaning to concepts such as provability, truth and proposition.

What Wittgenstein wants to attack is the undue adoration that so often comes to the fore in discussions of Gödel's theorem and which is similar to the adoration for the uniform treatment of all of mathematics in Russellian logic. In other words, not the mathematical or logical system itself is the issue, but its "position", our tendency to believe that the uniform and foundational treatment of these issues is *primary* and could exist independently of the *secondary* variety of mathematics in its actual use.

The end result of such a philosophical investigation is that we *see a mathematical result differently*, not that we abandon it as invalid. This change of perspective comes from seeing a particular *aspect* of mathematics more clearly, as Wittgenstein notes in an unpublished remark between two remarks of §19:

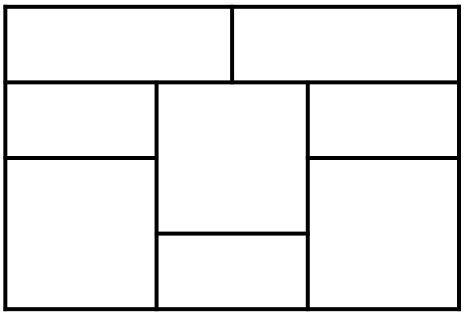
> Ich will einen bestimmten Aspekt der Mathematik herausarbeiten; & zwar den, der – meiner Meinung nach – {offenbar gemacht // klar geschildert} die Art & Weise beeinflußt, wie Mathematiker & Philosophen (heute) die Mathematik betrachten. [Ms-124, 83.2]

> I want to work out a certain aspect of mathematics; & namely the one that - in my opinion - {made obvious // clearly described}, influences the way mathematicians & philosophers (today) look at mathematics.

Closely connected with this change of perspective (by bringing an aspect into focus) is Wittgenstein's concept of surveyability, which has not played a major role in any of the preceding remarks on Gödel. Wittgenstein explicitly mentions Gödel[48] in Ms-124, 84.3 / *RFM VII* §19 and then immediately describes the surveyable example of joins in a wall:

> Meine Aufgabe ist es nicht über den Gödelschen Beweis (z.B.) zu reden; sondern an ihm vorbei zu reden.

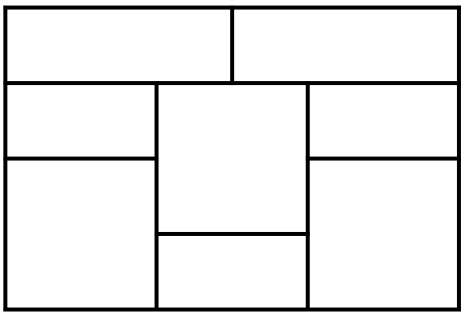
48 Wittgenstein's intent to "by-pass" Gödel's proof will not be explicitly discussed here, it should be clear from the preceding discussion that this expression does not indicate Wittgenstein's *mathematical* dismissal of Gödel's result but only Wittgenstein's belief that his task regarding Gödel's theorem is entirely *philosophical* and therefore operates on a different level than Gödel's mathematical work. This is also emphasised in Shanker, 1988, p. 235: "There is an understandable tendency to suppose that by announcing his desire to by-pass Gödel's proof Wittgenstein was indicating the low esteem in which he held Gödel's work. Yet had that been the case Wittgenstein would quite literally have by-passed Gödel's theorem: viz. by ignoring it."



Die Aufgabe, die Zahl der Wege zu finden, auf denen man den Fugen dieser Mauer ohne abzusetzen & ohne Wiederholung entlangfahren kann, erkennt jeder als *mathematische* Aufgabe. –

Wäre die {Zeichnung // Mauer} viel komplizierter & größer, nicht zu überblicken, so könnte man annehmen sie ändere sich, ohne das wir's merken, & dann wäre die Aufgabe, jene Zahl (die sich vielleicht gesetzmäßig ändert) zu finden, keine mathematische mehr. Aber auch wenn sie gleichbleibt, ist die Aufgabe dann nicht mathematisch. – Aber auch wenn {die Mauer // das Netz der Fugen} zu überblicken ist, {so heißt das nicht, die Aufgabe ist eine mathematische – als sagte man: *diese* Aufgabe ist nun eine der Embryologie. // so kann man nicht sagen: die Aufgabe wird dadurch zu einer mathematischen – wie man sagt: *diese* Aufgabe ist nun eine der Embryologie. // tritt die Aufgabe dadurch nun nicht in's Gebiet der Mathematik über – wie man sagt: *diese* Aufgabe ist nun eine der Embryologie.} Vielmehr: *hier* brauchen wir eine mathematische Lösung. (Wie: hier ist, was wir bedürfen, eine *Vorlage*.) [Ms-124, 84.3–84.5 / *BGM VII* §§19–20]

My task is, not to talk about (e.g.) Gödel's proof, but to by-pass it.



The problem: find the number of ways in which we can trace the joins in this wall continuously and without repetition, will be recognized by everyone as a *mathematical* problem. –

If the drawing were much bigger and more complicated, and could not be taken in at a glance, it could be supposed to change without our noticing; and then the problem of finding that number (which perhaps changes according to some law) would no longer be a mathematical one. But even if it does not change, the problem is, in this case, still not mathematical. – But even when the wall can be taken in at a glance, that cannot be said to make the question mathematical, as when we say: *this* question is now a question in embryology. Rather: *here* we need a mathematical solution. (Like: here what we need is a *model*). [*RFM VII* §19–20]

The problem described here is mathematical only because the picture of the wall is *surveyable*: We are able to treat it as a paradigmatic case, because the picture is simple enough that we will not overlook any joins, nor will any joins suddenly change without us noticing. Anscombe's translation of "zu überblicken" as "taken in at a glance" is unfortunate, because the notion at stake here is explicitly not primarily visual in the sense of being compact enough for us to glance at it and take it in, it is rather about surveyability in the sense of 'mechanical reproducibility',[49] our ability to treat it as an unchanging and ideal paradigm (even if this may require several steps and glances).

By considering the task of finding the different ways of following the joins in the wall as a mathematical problem, we treat it as an exercise in a realm of ideal lines and their connections, with no regard to any psychological processes (which might explain why one person is able to arrive at the correct number, while another is not) or physical properties of the joins and wall in question. It can then easily seem as if the mathematical problem were an *experiment* in this platonic realm of ideal objects, which is a conception of mathematics that Wittgenstein wants to dispel:

> 'Erkannten' wir das Problem als mathematisches, weil, die Mathematik vom Nachfahren von Zeichnungen handelt?
>
> Warum sind wir also geneigt, *dieses* Problem schlechtweg ein 'mathematisches' zu nennen? Weil wir es ihm gleich ansehen, daß hier die Beantwortung einer *mathematischen* Frage *so gut wie* alles ist, was wir brauchen. Obschon man das Problem, z.B., leicht als ein psychologisches sehen könnte.
> Ähnliches von der Aufgabe, aus einem Blatt Papier das & das zu falten.
>
> Es kann so ausschauen, als ob die Mathematik hier eine Wissenschaft ist, die mit *Einheiten* Experimente macht, Experimente, bei {denen // welchen} es nämlich nicht auf die Arten der Einheiten ankommt, also nicht darauf, ob sie Erbsen, Glaskugeln, Striche, usw. sind. – Nur was von *allen* diesen gilt, findet sie heraus. {Also nichts // Z.B. nichts} über ihren Schmelzpunkt, aber, daß 2 und 2 von ihnen 4 sind. Und das Problem der Mauer [№ 1] ist eben ein mathematisches, d.h.: kann durch *diese* Art von Experiment gelöst werden. – Und worin das math. Experiment besteht? Nun, im Hinlegen & Verschieben von Dingen, Ziehen von Strichen, Anschreiben von Ausdrücken, Sätzen, etc. Und man muß sich dadurch nicht stören lassen, daß die äußere Erscheinung dieser Experimente nicht die physikalischer & chemischer, etc. hat, es sind eben andersartige. Nur eine Schwierigkeit ist da: {der Vorgang // das, was vorgeht} ist leicht genug zu sehen, zu beschreiben – aber *wie* ist es als Experiment anzuschauen? Welches ist hier der Kopf, welches der Fuß des Experiments? Welches sind die Bedingungen des Experiments, welches sein Resultat? Ist das Resultat das Rechnungsergebnis, oder das Rechnungsbild, oder die Zustimmung (worin immer diese besteht) des Rechnenden?
>
> Werden aber, etwa, die Prinzipien der Dynamik zu Sätzen der reinen Mathematik dadurch, daß man ihre Interpretation offen läßt & sie nur zum Erzeugen eines Maßsystems verwendet? [Ms-124, 85.2–87.2 / *BGM VII* §20]

49 See Mühlhölzer, 2010.

> Did we 'recognize' the problem as a mathematical one because mathematics treats of making tracings from drawings?
>
> Why, then, are we inclined to call this problem straight away a 'mathematical' one? Because we see at once that here the answer to a *mathematical* question is *practically* all we need. Although the problem could easily be seen as, for example, a psychological one.
> Similarly with the task of folding a piece of paper in such-and-such a way.
>
> It may look as if mathematics were here a science that makes experiments with *units*; experiments, that is, in which it does not matter what kind of units they are, whether for instance they are peas, glass marbles, strokes and so on. – Mathematics discovers only what holds for *all* these things. And so it does not discover anything about e.g. their melting point, but that 2 and 2 of them are 4. And the first problem of the wall is a mathematical one, i.e. can be solved by means of *this* kind of experiment. – And what does the mathematical experiment consist in? Well, in setting things out and moving them about, in drawing lines, writing down expressions, propositions, etc. And we must not be disturbed by the fact that the outward appearance of these experiments is not that of physical or chemical experiments, etc.; they just are of a different kind. Only there is a difficulty here: the procedure is easy enough to see, to describe, – but *how* is it to be looked at as an experiment? What is the head and what the tails of the experiment here? What are the conditions of the experiment, what its result? Is the result what is yielded by the calculation; or the pattern of calculation; or the assent (whatever that consists in) of the person doing the calculation?
>
> But does it make the principles of dynamics, say, into propositions of pure mathematics if we leave their interpretation open, and then use them to produce a system of measurement? [*RFM VII* §20]

This conception of mathematics as the science of the "ultraphysical", a science in the platonic realm of ideal entities such as lines or numbers, contrasts with Wittgenstein's own conception of mathematics, where the paradigmatic nature of a problem depends entirely on the surveyability of the task. Wittgenstein reaffirms one of his central reflections in Ms-122 & Ms-117 / *RFM III*, that a proof must be surveyable (or else it would not be usable as a proof):

> "Der math. Beweis muß übersichtlich sein" – das hängt mit der Übersichtlichkeit jener Figur zusammen. [Ms-124, 87.3 / *BGM VII* §20]

> 'A mathematical proof must be perspicuous' – this is connected with the perspicuousness of that figure. [*RFM VII* §20]

In the context of Gödel's proof, the different conceptions of mathematics discussed here by Wittgenstein lead to entirely different interpretations:

A picture of mathematics as a science in the ultraphysical and platonic realm of ideal entities has the tendency to let Gödel's proof appear as a fundamental and foundational limitative result, which demonstrates clear boundaries to what we are able to do in a formal system. According to such a view, the limits proved by Gödel hold in the ideal realm of mathematics and thus all the more in practice.[50]

50 The most extreme form of this conception then leads almost directly to many of the abuses of Gödel's theorem in the context of the philosophy of mind, where

In contrast, Wittgenstein's conception of mathematics emphasises the surveyability of proofs, which can be understood in two ways in the context of Gödel's first incompleteness theorem: First of all, there is Gödel's proof itself and the rather obvious observation that according to Wittgenstein's view this proof is only a proof if it is surveyable. It is unlikely that Wittgenstein wants to deny the surveyability of Gödel's proof, as this would come close to disagreeing with the validity of the proof and thus commit Wittgenstein to a mathematical critique, which is explicitly not the aim here. Secondly, there is the question of whether the *proof chains inside the formal system* considered by Gödel are surveyable, in other words the question of whether the arithmetical translation of a metamathematical proposition such as 'X is provable' is surveyable. This second form of proof seems to be the proof (with its corresponding notion of surveyability) that Wittgenstein is interested in, as becomes clear in the remark immediately following the remark on surveyability:

> Vergiß nicht: der Satz, der von sich selbst aussagt, er sei unbeweisbar, ist als *mathematische* Aussage aufzufassen, – – – denn das ist nicht *selbstverständlich*.
>
> Es ist nicht selbstverständlich, daß der Satz, die & die Struktur sei: so & so nicht konstruierbar, als mathematischer Satz aufzufassen {sei // ist}.
>
> D.h.: wenn man sagt: "er sagt von sich selbst aus" – so ist das auf eine spezielle Weise zu verstehen. Hier nämlich entsteht leicht Verwirrung durch den bunten Gebrauch des Ausdrucks "dieser Satz sagt etwas von ... aus".
>
> In diesem Sinne sagt der Satz 625 = 25 × 25 auch etwas über sich selbst aus: daß nämlich die linke Ziffer erhalten wird, wenn man die rechts stehenden multipliziert.
>
> Der Gödelsche Satz, der etwas über sich selbst aussagt, *erwähnt* sich selbst nicht. [Ms-124, 87.4–88.4 / *RFM VII* §21]

> Do not forget that the proposition asserting of itself that it is unprovable is to be conceived as a *mathematical* assertion – for that is not *a matter of course*.
>
> It is not a matter of course that the proposition that such-and-such a structure cannot be constructed is to be conceived as a mathematical proposition.
>
> That is to say: when we said: "it asserts of itself" – this has to be understood in a special way. For here it is easy for confusion to occur through the variegated use of the expression "this proposition asserts of something of...".

---

the mathematical theorem is interpreted as demonstrating a limit to what machines could ever hope to achieve, in contrast to humans (see Lucas, 1961 for one of the most egregious examples of such an interpretation and Franzén, 2005, pp. 115–126 for a discussion of Lucas' interpretation and other abuses). It should be noted that Gödel himself was more careful in this regard, which is why Wittgenstein's remarks should be read as targeting not only Gödel's own informal interpretation, but also (and perhaps more importantly) the wealth of philosophical abuses that Gödel's informal introduction gave rise to. Wittgenstein certainly saw that Gödel's way of presenting his purely mathematical result had the tendency to generate these abuses, even though Gödel himself may be only partially to blame for this.

> In this sense the proposition '625 = 25 × 25' also asserts something about itself: namely that the left-hand number is got by the multiplication of the numbers on the right.
>
> Gödel's proposition, which asserts something about itself, does not *mention* itself. [*RFM VII* §21]

Apart from demonstrating that Wittgenstein was, at least at the time of writing Ms-124 in 1941, aware of the indirect form of reference that Gödel's proof uses, the above remarks can help clarify why the notion of surveyability plays a central role in the context of Gödel. Wittgenstein understands how Gödel's proof shows that a particular proof chain is not constructible ("such-and-such a structure cannot be constructed") under the assumption of ω-consistency and he is certainly not disputing the mathematical validity of this aspect of the proof. What he wants to investigate instead is our willingness to conceive this proof of non-constructibility as a mathematical proposition, which says something metamathematical "about itself". Gödel is undoubtedly well aware that the undecidable proposition P does not say something "about itself" in the same way as semantic paradoxes such as the Liar do, after all Gödel himself points out in the introduction that any such translation is only an informal and imprecise reformulation and that the undecidable proposition P is in fact not a paradox.

Wittgenstein's point is more nuanced here, because he is not criticising Gödel on the basis of a rather trivial observation, but rather investigating the 'uniformity' of Gödel's approach. The approach chosen by Gödel shows similarities to Russell's conception of mathematics as logic in one crucial aspect: Exactly how Russell views mathematics as reducible to logic, ignoring the fact that his translation into logic is only understandable against the backdrop of our 'informal' understanding of mathematics with its variety of language games, Gödel views all metamathematical propositions as equally reducible to arithmetic, ignoring the fact that this translation is only understandable against the backdrop of our 'extra-systemic' understanding of metamathematical concepts. For a mathematician, who is usually not interested in the surveyability of their approach as long as it can be shown that a translation exists and can *in theory* be carried out, the variety of language games that give rise to our informal concepts simply do not matter and the recursive method of translation is self-sufficient. But for Wittgenstein, the variety of informal language games and the "variegated use" of our metamathematical expressions is not reducible to a uniform logical treatment.

Gödel's metamathematical translation into arithmetic has sense only because the informal metamathematical concept of provability has sense for us, but the uniform treatment of all propositions through this arithmetical lens does not guarantee that the translation will have sense when it is stretched beyond our informal concepts, as is the case for Gödel's undecidable proposition P. Here we are easily misled to believe that we know what we *must* say, *in analogy to our informal concept of provability*, when we actually have not given the expression any meaning in this particular context and thus have the freedom to either exclude it from the system (as Gödel does) or let the contradictory interpretation stand.

Wittgenstein is not advocating for one option over the other, but rather for an understanding that sees these different options not as ultraphysical laws and instead as logical rules that govern our language game. Once we adopt this view of Gödel's proof, it appears less as a reflection of a fundamental limitation that we have discovered in the realm of logic and mathematics and more as a reflection of rules that have been fixed by a variety of different mathematical language games. If we are able to choose between different options, then only because some expressions do not yet have a fixed meaning in our existing form of life.[51]

Only in this way can there be doubt about the translation of an arithmetical proposition into English: Wittgenstein is not doubting the mathematical validity of Gödel's translation procedure, only its applicability in situations where we are lacking a clear picture in our informal language. In such a situation, the result of translating an arithmetical proposition (such as the undecidable P) into its metamathematical English translation is comparable to a useless but aesthetically pleasing ornament, as Wittgenstein notes in an unpublished remark between remarks of §21:

> 'Der Satz sagt, daß diese Zahl aus diesen Zahlen auf diese Weise nicht erhältlich ist.' – Aber bist Du auch sicher, daß Du ihn recht ins Deutsche übersetzt hast? Ja gewiß, es scheint so. – Aber kann man da nicht fehlgehen?
>
> | Ein Stil, Maschinen zu bauen, in welchem man die wirksamen Räder, Hebel, etc. mit einer Zahl unwirksamer umgibt, die, z.B., nur eines ästhetischen Eindrucks wegen angebracht sind. (Ähnlich wie Scheinfester in einer Fassade.) | [Ms-124, 89.2–89.3 / *BGM VII* §21 (only the first remark)]
>
> ---
>
> 'The proposition says that this number cannot be got from these numbers in this way.' – But are you also certain that you have translated it correctly into English? Certainly it looks as if you had. – But isn't it possible to go wrong here?
> | A style of building machines in which the effective wheels, levers, etc. are surrounded by a number of ineffective ones, which, for example, are only attached for the sake of an aesthetic impression. (Similar to false windows in a façade.) | [*RFM VII* §21 (only the first remark)]

In other words, we might decide to accept the ornament as part of a larger structure, but we cannot point to the usual reasons for justifica-

51 Of course a choice based on the analogy with existing language games and fields of mathematics might very well be a reasonable and perfectly adequate choice, contrary to Wittgenstein's rather dismissive rejection of it as a "stupid reason". Wittgenstein's perspective on Gödel's theorem is certainly tinted by his approach as a non-mathematician, who considers Gödel's proof only on the basis of the proof alone, not in the context of specialised fields of mathematics.

tion, because they stem from a variety of situations that do not apply here.

But what are we to make of this observation? Assuming that we leave open the possibility of 'letting the contradiction stand', can we really use the resulting formal system, given that we cannot *trust* the English translation of the undecidable proposition P (which is provable despite stating that it is not provable)? Wittgenstein considers this question in the next remark:

> Könnte man sagen: Gödel sagt, daß man einem math. Beweis auch {muß trauen können // trauen muß}, wenn man ihn, praktisch, als den Beweis der Konstruierbarkeit der Satzfigur nach den Beweisregeln auffassen will?
>
> Oder: Ein math. Satz muß als Satz einer auf {sich selbst // sein eigenes Zeichen} wirklich anwendbaren Geometrie aufgefaßt werden können. Und tut man das so zeigt es sich, daß man sich auf einen Beweis in gewissen Fällen nicht verlassen kann. [Ms-124, 89.4 / *BGM VII* §21]

> Could it be said: Gödel says that one must also be able to trust a mathematical proof when one wants to conceive it practically, as the proof that the propositional pattern can be constructed according to the rules of proof?
>
> Or: a mathematical proposition must be capable of being conceived as a proposition of a geometry which is actually applicable to itself. And if one does this it comes out that in certain cases it is not possible to rely on a proof. [*RFM VII* §21]

It is not entirely surprising that Wittgenstein marked the expression "in certain cases" in the above remark with a wavy underline, indicating his discontent with this particular phrasing, because these three words strike at the heart of the matter, but without the philosophical clarity of many of his earlier remarks. A philosophical investigation will need to investigate precisely *in which cases* we cannot trust a proof inside the formal system, and more importantly *why* we cannot trust it. Gödel's own interpretation of the proposition P as metamathematically true but unprovable within the system can lead us to believe that unprovability "in certain cases" is a sign of the deficiency of the formal system, which (as Gödel's proof shows) we are unable to resolve inside the formal system itself. But are these cases really deficient, if the proposition in question is inherently useless and (indirectly, when translated into its metamathematical counterpart) reflexive?

In the following remarks, Wittgenstein makes another attempt at succinctly describing the Gödelian situation and offering a philosophically undogmatic clarification. It is worth quoting them in full, because these remarks do better justice to Gödel's mathematical proof (by explicitly describing the construction of the undecidable proposition as an *arithmetical* proposition together with a metamathematical translation) and also manage to clarify its philosophical implications in a very undogmatic and unspectacular fashion:

> 'Nehmen wir an, wir haben einen arithmetischen Satz, der sagt, eine bestimmte Zahl ... könne nicht aus den Zahlen ... , ... , ... , durch die & die Operationen gewonnen werden. Und nehmen wir an, es ließe sich eine

Übersetzungsregel geben, nach welcher dieser arithmetische Satz in die Ziffer jener ersten Zahl – die Axiome {, aus denen wir versuchen ihn zu beweisen, // unseres Beweissystems } in die Ziffern jener andern Zahlen – & unsere Schlußregeln in die im Satz erwähnten Operationen sich übersetzen ließen. – Hätten wir dann *den arithmetischen Satz* aus den Axiomen nach unsern Schlußregeln abgeleitet, so hätten wir *dadurch* seine Ableitbarkeit demonstriert, aber auch einen Satz bewiesen, den man nach jener Übersetzungsregel dahin aussprechen {kann // muß}: dieser arithmetische Satz (nämlich unserer) sei unableitbar.

Was wäre nun da zu tun? Ich denke mir, wir schenken unserer *Konstruktion* des *Satzzeichens* glauben, also dem *geometrischen* Beweis. Wir sagen also, diese 'Satzfigur' ist aus jenen so & so gewinnbar. Und übertragen, nur, in eine andre Notation heißt das: diese Ziffer ist mittels dieser Operationen aus jenen zu gewinnen. Soweit hat der Satz & sein Beweis nichts mit einer besonderen *Logik* zu tun. Hier war jener konstruierte Satz einfach eine andere Schreibweise der konstruierten Ziffer; sie hatte die *Form* eines Satzes aber wir verglichen sie nicht mit andern Sätzen als Zeichen, welches dies oder jenes *sagt*, einen *Sinn* hat.

Aber freilich ist zu sagen daß jenes Zeichen weder als Satzzeichen noch als Zahlzeichen angesehen werden braucht. – Frage Dich: was macht es zu dem einen, was zu dem anderen? [Ms-124, 90.4–92.2 / *BGM VII* §22]

'Let us assume that we have an arithmetical proposition saying that a particular number ... cannot be obtained from the numbers. . . , . . . , . . . , by means of such and such operations. And let us assume that a rule of translation can be given according to which this arithmetical proposition is translatable into the figures of the first number – the axioms from which we are trying to prove it, into the figures of the other numbers – and our rules of inference into the operations mentioned in the proposition. – If we had then derived *the arithmetical propositions* from the axioms according to our rules of inference, then *by this means* should have demonstrated its derivability, but we should also have proved a proposition which, by that translation rule, can be expressed: this arithmetical proposition (namely ours) is not derivable.'

What would have to be done here? I am supposing that we trust our *construction* of the propositional sign; i.e. we trust the *geometrical* proof. So we say that this 'propositional pattern' can be obtained from those in such and such ways. And, merely translated into another notation, this means: this number can be got from those by means of these operations. So far the proposition and its proof have nothing to do with any special *logic*. Here the constructed proposition was simply another way of writing the constructed number; it had the *form* of a proposition but we don't compare it with other propositions as a sign *saying* this or that, making *sense*.

But it must of course be said that that sign need to be regarded either as a propositional sign or as a number sign. – Ask yourself: what makes it into the one, and what into the other? [*RFM VII* §22]

The first half of the first remark is entirely unproblematic and simply describes the result of Gödel's construction of P, namely ω-inconsistency, assuming that we do not exclude P as undecidable.

The second half, however, reiterates Wittgenstein's idea that we might accept the construction (and therefore the provability) of P instead of excluding it, in other words "we trust the *geometrical* proof." As Wittgenstein correctly points out, the correspondence between the arithmetical operations and the metamathematical interpretation as a proved proposition is one of "merely" translating it, indicating that the mathematical validity of Gödel's translation is not at stake here.

But why do we treat the number as a proposition? Only because (thanks to the translation procedure) the propositional sign has the "*form* of a proposition", in other words due to the uniform treatment that we apply to all those propositional signs that fulfil certain formal criteria, irrespective of their particular content, irrespective of what they *say*.

The problem 'only' arises if we then try to translate the constructed number "by means of these operations" back to its metamathematical counterpart, which "says the opposite of what we regard as proved":

> Lesen wir nun den konstruierten Satz (oder die Ziffer) als Satz der mathematischen Sprache (etwa auf Deutsch), so spricht er das Gegenteil von dem, was wir eben als bewiesen betrachtet. Wir haben also den wörtlichen Sinn des Satzes als falsch demonstriert & ihn zu gleicher Zeit *bewiesen* – wenn wir nämlich seine Konstruktion aus den zugelassenen Axiomen mittels der zugelassenen Schlußregeln als Beweis betrachten.
>
> Wenn jemand uns einwürfe, wir könnten solche *Annahmen* nicht machen, da es *logische* oder *mathematische* Annahmen wären, so antworten wir, daß nur nötig ist anzunehmen jemand habe einen Rechenfehler gemacht & sei *dadurch* zu dem Resultat gelangt, das wir 'annehmen', & er könne diesen Rechenfehler vorderhand nicht finden. [Ms-124, 92.3–93.2 / *BGM VII* §22]

> If we now read the constructed proposition (or the figures) as a proposition of mathematical language (in English, say) then it says the opposite of what we regard as proved. Thus we have demonstrated the falsity of the real sense of the proposition and at the same time *proved* it – if, that is, we look on its construction from the admitted axioms by means of the admitted rules of inference as a proof.
>
> If someone objects to us that we couldn't make such *assumptions*, for they would be *logical* or *mathematical* assumptions, then we reply that we need only assume that someone has made a mistake in calculating and so has reached the result we 'assume', and that for the time being he cannot find the mistake. [*RFM VII* §22]

Gödel's own informal introduction has the tendency to present the mathematical result as a surprising discovery about the limits of formal systems, because it is impossible to decide all propositions without running head on into ω-inconsistency, contrary to what we had expected. This is why Gödel's result can appear troubling: We expect to be able to trust proofs, in other words we expect that a proved proposition does not say something false, which is exactly what happens in the case sketched out by Wittgenstein.

Gödel has no other option but to exclude the proposition from the formal system, to avoid inconsistency by accepting incompleteness, because the uniform treatment of propositions and the explosive effect of contradictions in classical logic would render a formal system unusable if we tried to use it 'mechanically' in the presence of a contradictory proposition such as P.

But as Wittgenstein's concept of surveyability makes clear, this approach merely appears to be the only viable option because we lack surveyability of the formal system and as a consequence cannot distinguish between contradictions that threaten the usability of the formal system for practical purposes and potentially harmless contradictions that resemble pointless but ultimately benign language games such as somebody repeatedly drawing conclusions of the form 'I lie, therefore I do not lie, therefore I lie, ...' from a sentence such as the Liar.

Once a formal system is surveyable, which can only happen as a consequence of investigating the use of the system, not as a result of merely considering the formal rules of the system, we might possess criteria which allow us to decide whether to include a diagonalised proposition such as P inside the system. In contrast to Gödel's way of introducing the proof, which presents the result as a fundamental and unexpected limitation of formal systems of a particular form, a surveyable representation can show that a diagonalised construction does not *unexpectedly* lead to a contradiction, but rather obviously as a result of how the rules of the game are laid out. From such a surveyable perspective, the 'limitation' of the formal system is not a law of nature, but only the expected consequence of the rules that we chose to include and therefore entirely unsurprising, similar to how nobody would be surprised that all games which include a rule that allows one player to win immediately in fact do allow one player to win immediately and are thereby limited in a particular sense.

Of course Gödel's proof is far from trivial and the mathematical result much deeper than a mere tautology, but from the perspective of Wittgenstein its value does not lie in showing us a deep discovery about the mathematical world, but rather in the observation that if we follow our usual rules of logic in a certain way, we can diagonalise our way to a contradiction. By bringing this situation to our attention, Gödel makes a particular (intra-systemic) aspect of formal systems surveyable, while simultaneously failing to see that any interpretation of this aspect will depend on its (extra-systemic) use.

Wittgenstein wants to emphasise that this use is not a matter of psychology and therefore irrelevant to a mathematician. Instead, whether or not the geometrical proof by construction of the 'undecidable' proposition P convinces us as a proof depends on "its ratification in the use of what is proved":

> Hier kommen wir wieder auf den Ausdruck "der Beweis überzeugt uns" zurück. Und was uns hier an der Überzeugung interessiert, ist weder ihr Ausdruck durch Stimme und Gebärde, noch das Gefühl, der Befriedigung oder ähnliches; sondern ihre Betätigung in der Verwendung des Bewiesenen. [Ms-124, 94.1 / *BGM VII* §22]

> Here once more we come back to the expression "the proof convinces us". And what interests us about conviction here is neither its expression by

> voice or gesture, nor yet the feeling of satisfaction or anything of that kind; but its ratification in the use of what is proved. [*RFM VII* §22]

With the above remark acting as a preliminary conclusion of Wittgenstein's thoughts on Gödel's theorem, the closing remarks in Wittgenstein's discussion of Gödel in Ms-124 take a more general perspective and reflect on the role of philosophy in regards to such a proof. It is clear that Gödel's proof, although it is itself surveyable (by being a proof) and helps to make a particular aspect of formal systems surveyable (by demonstrating how certain rules can be used to construct an undecidable or contradictory proposition), does not lead to a surveyable presentation of the form that Wittgenstein is interested in, because Gödel's proof ignores the variety of uses that different propositions exhibit. This is not a deficiency of Gödel's particular proof, but simply the result of a difference in method between mathematics and philosophy, as Wittgenstein understands the two fields. Gödel's proof can therefore not solve a philosophical problem, but it can pinpoint a situation that would benefit from greater surveyability (in the philosophical sense):

> Man {könnte // kann} mit Recht fragen, welche Wichtigkeit Gödel's Beweis für unsre Arbeit habe. Denn ein Stück Mathematik {kann nicht Probleme von der Art der unsern // kann Probleme von der Art, die *uns* beunruhigen, nicht lösen. // kann kein Problem von der Art, die *uns* beunruhigt lösen. // kann nicht Probleme von der Art, die uns beunruhigt, lösen.} – Die Antwort ist: daß die *Situation* uns interessiert, in die ein solcher Beweis uns bringt. 'Was sollen {wir // sie} nun sagen?' – das ist unser Thema. [Ms-124, 94.2 / *BGM VII* §22]

> It might justly be asked what importance Gödel's proof has for our work. For a piece of mathematics cannot solve problems of the sort that trouble *us*. – The answer is that the *situation*, into which such a proof brings us, is of interest to us. 'What are we to say now?' – That is our theme. [*RFM VII* §22]

Consequently, a philosopher must focus on details that appear to be completely irrelevant to a mathematician and must ask seemingly trivial questions. Reducing concepts such as the natural numbers or provability to a simple axiomatic system with a small number of rules of inference (such as the Peano arithmetic considered by Gödel in his proof) might be a viable solution for a mathematician, but does nothing to bring us closer to the kind of surveyable representation that describes (without explaining) the variety of uses in our language. Instead, a philosopher must ask what it means to say "Suppose this could be proved", because the answer could be entirely different in the case of Gödel's proposition P than in the case of a 'regular' proposition that we use to infer that it is raining outside:

> Es kommt uns viel zu selbstverständlich vor, daß wir "wieviele?" fragen & darauf zählen & rechnen!
>
> So seltsam es klingt, so scheint meine Aufgabe das Gödelsche Theorem betreffend (bloß) darin zu bestehen, klar zu stellen, was in der Mathematik

> so ein Satz bedeutet, wie: "angenommen, man könnte dies beweisen". [Ms-124, 95.1–95.2 / *BGM VII* §23 & §22]

> We take it much too much for granted that we ask "How many?" and thereupon count and calculate.
>
> However queer it sounds, my task as far as concerns Gödel's proof seems merely to consist in making clear what such a proposition as: "Suppose this could be proved" means in mathematics. [*RFM VII* §23 & §22]

Philosophy in the sense of Wittgenstein must therefore refrain from interfering in mathematical matters and avoid criticising a proof on the mathematical level. A philosophical investigation can surely take a mathematical proof such as Gödel's theorem as a starting point and focus on the way that it might give rise to a particular and potentially misleading interpretation, but the object of any philosophical investigation can only be the *prosaic* interpretation of the proof, acting as a connection point to our extra-mathematical use of certain concepts, never the mathematical formalism itself.

2.7 PHYSICS AND MATHEMATICAL OBJECTS

While the remarks on Gödel in Ms-124 / *RFM VII* end at this point, they continue in the pocket notebook Ms-163, which is the source for the later selection in Ms-124. Given that Wittgenstein himself chose not to transfer the following remarks from Ms-163 into Ms-124 and that Ms-163 (being a pocket notebook) contains coded personal entries and drafts of unfinished remarks, not too much weight should be placed on any interpretation of Ms-163. Accordingly, only a selection of remarks from that notebook will be discussed here.

Wittgenstein continues his remarks on Gödel in Ms-163 by considering the role of a reflexive pronoun, similar to his remarks in Ms-121, while emphasising that such a self-referential word would only find use outside the realm of logical or mathematical propositions. As soon as self-reference is involved, there are different senses of negating such a sentence, because it is not immediately clear how to interpret the 'scope' of the reflexive pronoun (see the list of 6 propositions outlined in Section 2.5) and because there are different ways of deriving a proposition, either directly as the constructed end result of a proof figure or indirectly as the result of a non-constructive provability proof (see Section 2.2). We could thus speak of a different "sense" of negating the same sentence:

> Das auf den Satz *selbst* bezügliche Fürwort des Satzes, der etwas von sich selbst aussagt. Ein solches gibt es in unsrer Sprache nicht, sein Gebrauch, das Sprachspiel, aber kann leicht beschrieben werden, wenn man nur erst sieht daß die Sätze, in denen es vorkommt nicht, vor allem, logische oder math. sein dürfen.
>
> Sagt nun so ein Satz: "ich bin nicht wahr" so habe ich gar keinen Gebrauch für ihn. Es sei denn daß ich das Spiel mit ihm spiele zu sagen: Also ist das Gegenteil dieses Satzes wahr welches lautet: "ich bin wahr."

> Und dies ist in *einem* Sinne das Gegenteil & in einem andern Sinne nicht. [Ms-163, 31r.2–31r.3]

> The pronoun of the sentence referring to the sentence *itself*, which says something about itself. Such a pronoun does not exist in our language, but its use, the language game, can be easily described, if one only sees that the propositions in which it occurs are not allowed to be, above all, be logical or mathematical.
>
> If now such a proposition says, "I am not true," I have no use for it at all. Unless I play the game with it of saying: So the opposite of this proposition is true, which is: "I am true."
>
> And this, in *one* sense, is the opposite, and in another sense it is not.

A few remarks later, Wittgenstein mentions "K provable" or "κ provable" (Ms-163, 32r.1–32r.2, the handwriting is not completely clear), apparently an explicit reference to Gödel's use of the term in his proof. Wittgenstein then introduces the main theme of the following remarks, namely the relation between contradictions in mathematics and physics (although the theme is only alluded to at this point):

> Aber macht nicht dies den Gebrauch solcher Sätze unmöglich daß hier ein Satz & sein Gegenteil wahr sein können?
>
> Z.B.: "ich bin ein Zoll lang" & "ich bin nicht ein Zoll lang".
>
> Man könnte hier sagen es müsse eine äußere & eine innere Negation geben.
>
> Das gleiche gilt natürlich von "ich bin ableitbar" & "ich bin nicht ableitbar", sie können beide wahr & beide falsch sein: Und dennoch nicht sinnlos. [Ms-163, 32r.3]

> But doesn't this make the use of such sentences impossible, that here a sentence & its opposite can be true?
>
> E.g.: "I am an inch long" & "I am not an inch long".
>
> One could say here that there must be an external & an internal negation.
>
> The same is of course true of "I am derivable" & "I am not derivable", they can both be true & both be false: And yet not meaningless.

At first glance, it can appear as if a mathematical contradiction immediately disqualified any contradictory proposition as useless, since this is the case for empirical propositions such as "I am an inch long" and "I am not an inch long". Here, the "use of such sentences [is] impossible", because they cannot both refer to the same physical object and be true at the same time. If both of these propositions were derived somehow, the result would be useless.

The first instinct is to assume that this analogy also holds in mathematics, which would compel us to exclude Gödel's proposition P as undecidable from the formal system instead of accepting its contradictory nature within the system. But in the case of such an undecidable proposition (and in contrast to empirical propositions), P and its negation are derived differently from one another and thus each have a different "sense", which we could distinguish by speaking of "an external & and internal negation". In mathematics, the contradictory propositions can then "both be true & both be false: And yet not meaningless."

As a result, the derivation of such a contradiction loses its predictive power, because if we derive P (saying that P is not derivable), its prediction is restricted to a single sense of derivability, but it does not preclude that we derive the proposition or its negation in the other sense. Under 'normal' circumstances, a proof that a particular proposition cannot be derived can be used to demonstrate that looking for a derivation is "hopeless", but a proposition such as P with its 'double-sense' of negation does not have this use:

> Hättest Du {etwas // einen mathematischen Satz} aus logischen & arithmetischen Grundprinzipien abgeleitet, dessen natürlichste Anwendung zu sein schiene das Ableiten des abgeleiteten Satzes als hoffnungslos darzustellen, dann heißt das, daß der so abgeleitete Satz diese Anwendung eben *nicht* hat, daß die Prinzipien, aus welchen er abgeleitet ist, nicht im Stande sind eine {auf diese Weise // *so*} anwendbare Geometrie zu erzeugen. [Ms-163, 32v.2]

> If you had derived {something // a mathematical proposition} from logical & arithmetical basic principles, the most natural application of which would seem to be the derivation of the derived proposition as hopeless, then this means that the proposition thus derived has simply *not* this application, that the principles from which it is derived are not capable of producing a geometry {applicable in this way // *thus* applicable}.

Wittgenstein then connects this idea to the concept of surveyability, by comparing the contradiction produced by a proposition such as P with a contradiction that results from a mistake made due to calculating with impractically large numbers (such as when calculations are carried out in unary Russellian notation, Wittgenstein's prime example for unsurveyability in mathematics):

> Ist das nun viel anders als gäbe ein allgemeiner arithmetischer Beweis, auf außerordentlich sehr große Zahlen angewandt etwas, was im Widerspruch steht mit dem Resultat der speziellen & ungeheuer langen Berechnung? So könnte ich mir denken, daß Paare ungeheuer langer Multiplikationen $n \times m$, $m \times n$ zu immer verschiedenen Resultaten führten.
>
> Die Jagd nach den Grundlagen der Mathematik {scheint mir auf ein falsches Ideal basiert. // scheint mir erregt durch ein trügliches Ideal. // scheint mir (ganz) getragen von einem trüglichen Ideal.} (Wie eine bestimmte Politik von einer bestimmten Lebensweise.) [Ms-163, 33r.2–33v.2]
>
> [...]
>
> Wenn ich ein Beispiel einer möglichen Verwirrung in der Arithmetik finden will, brauche ich mir nur ein Rechnen mit riesigen Zahlen vorstellen welches unübersehbar & dadurch unzuverlässig wird. [Ms-163, 34v.2]

> Now, is this much different than if a general arithmetic proof, applied to extraordinarily very large numbers, gave something that contradicted the result of the particular & tremendously long calculation? I could think of pairs of tremendously long multiplications $n \times m$, $m \times n$ leading to always different results.
>
> The hunt for the foundations of mathematics {seems to me to be based on a false ideal //seems to me excited by a fallacious ideal //seems to me (entirely) sustained by a fallacious ideal}. (Like a certain policy from a certain way of life.)

> [...]
> If I want to find an example of a possible confusion in arithmetic, I only need to imagine a calculation with huge numbers which becomes unsurveyable & thereby unreliable.

At first this comparison might strike us as odd, because from the perspective of most mathematicians these cases are of a fundamentally different nature: A proposition such as P is *inherently* contradictory and a sign of inconsistency in the formal system itself, whereas an unsurveyable large calculation is only contradictory *in practice*, while the underlying system might be perfectly consistent. But it is exactly this preconception that Wittgenstein wants to attack, because it gives rise to the misleading picture of inconsistency as a deficiency in and of itself. As Wittgenstein has emphasised repeatedly, the inconsistency introduced by a proposition such as Gödel's P occurs only in 'small doses' and could conceivably be contained without leading to logical trivialism. Whether or not the resulting system is useful then depends on how it is used in practice, in our form of life, which is a question that cannot be answered on the basis of purely mathematical reflections.

Contradictions as the result of diagonalisation and contradictions as the result of unsurveyably large numbers are similar in so far as both precisely locate the 'area of unsurveyability' in a way that this area can be 'fenced off' without impacting other areas of the formalism. Although Russellian notation is unsuitable for calculating by hand with large numbers, it could still be used for calculating 'in the small'. Similarly, we might use a formal system containing a proposition such as P for practical purposes as long as we restrict its application to non-diagonalised propositions or employ a paraconsistent logic as our framework. In contrast, a formal system that derives non-diagonalised contradictions is not suitable for such a use until we have found out where and why these contradictions originate, in other words until the formal system becomes *surveyable*.

The importance of surveyability continues to play a role in the remarks immediately following the one quoted above, with Wittgenstein reiterating the difficulty of finding the right perspective for his philosophical investigation (Ms-163, 35r.2–35r.4). He then talks explicitly about how Gödel's proof fits into such a philosophical investigation and why the issue at stake is not the mathematical proof itself, but rather its prosaic interpretation:

> Nicht der Gödelsche Beweis interessiert mich, sondern die Möglichkeiten auf die Gödel durch seine Diskussion uns aufmerksam macht.
>
> Die math. Tatsache daß hier ein arithmetischer Satz ist, der sich in P nicht beweisen noch als falsch erweisen läßt, interessiert mich nicht. – – – [Ms-163, 37v.3–37v.4]

> It is not Gödel's proof that interests me, but the possibilities to which Gödel draws our attention through his discussion.

> The math. fact that here is an arithmetical proposition which cannot be proved in P nor demonstrated to be false does not interest me. - - -

A few remarks later, Wittgenstein continues this general reflection and emphasises that the focus of his philosophical investigation is the "kind of proof" that Gödel introduced, namely the diagonal method applied to metamathematical concepts such a provability:

> Der Gödelsche Beweis {bringt eine Schwierigkeit auf // entwickelt eine Schwierigkeit}, {die sich auch in viel elementarerer Weise zeigen muß // die auch in viel elementarerer Weise erscheinen muß}. (Und hierin liegt, scheint es mir, zugleich Gödels großes Verdienst um die Philosophie der Math., & zugleich der Grund, warum sein besonderer Beweis nicht das ist was uns interessiert.)
>
> Ich könnte sagen: Der Gödelsche Beweis gibt uns die Anregung dazu die Perspektive zu ändern aus der wir die Mathematik sahen. *Was* er beweist, geht uns nichts an, aber wir müssen uns mit dieser mathematischen *Beweisart* auseinandersetzen. [Ms-163, 39v.3–40r.2]

> Gödel's proof {raises a difficulty // develops a difficulty}, {which must also show itself in a much more elementary way // which must also appear in a much more elementary way}. (And here lies, it seems to me, at the same time Gödel's great merit for the philosophy of mathematics, & at the same time the reason why his particular proof is not what interests us).
>
> I could say: Gödel's proof gives us the stimulus to change the perspective from which we saw mathematics. *What* he proves is none of our business, but we have to deal with this mathematical *kind of proof*.

This is why Wittgenstein's remarks on Gödel should be read in the larger context of 'higher-order systems', which appear to resolve philosophical issues, but in fact merely shift the issue to a higher layer. As Wittgenstein has noted repeatedly, this is comparable to 'second-order orthography' (*PI §121*), which is also the example used in Ms-163:

> Es gilt die Gedanken *so* zu ordnen, daß man die Untersuchung an einem beliebigen Punkt abbrechen kann ohne daß, was nach diesem Punkt kommt, wieder das in Frage stellen kann, was {bis dorthin // bis dahin} gesagt wurde.
>
> Hier kommen wir wieder zu dem Gedanken, daß, das Wort "buchstabieren" buchstabieren, nicht ein {Buchstabieren des zweiten // Buchstabieren-höhern} Grades ist. [Ms-163, 40v.4–41r.2]

> It is necessary to arrange the thoughts *in such a way* that one can stop the investigation at any point without what comes after this point being able to call into question again what has been said {until there // until then}.
>
> Here we come again to the thought that to spell the word "spell" is not a spelling of {the second // a higher} order.

The remarks immediately following continue the general reflections on the role of the specific mathematical proof for Wittgenstein's investigation and nicely illustrate that even in the case of Gödel, Wittgenstein's intent is not to interfere with or attack the proof, which he calls "impeccable", while at the same time drawing our attention to the "exceptional position" of Gödel's proposition P in the system:

> Wenn die beiden ω-widersprechenden Beweise wirklich vorliegen, dann wird es problematisch, was wir mit dem so bewiesenen & entkräfteten Satze anfangen können.
>
> Gödel zeigt *einwandfrei*, daß der von ihm konstruierte Satz eine *Ausnahmsstellung* im System der Sätze {hat // einnimmt}. (D.h.,) wie immer man diese Ausnahmsstellung beschreibt, *so bleibt es eine solche*. [Ms-163, 41r.3–41v.2]

> If the two ω-contradictory proofs are really at hand, then it becomes problematic as to what we can do with the thus proven & invalidated proposition.
>
> Gödel shows *impeccably* that the proposition constructed by him occupies an *exceptional position* in the system of propositions. (I.e.,) however one describes this exceptional position, *it remains one*.

Gödel's result can rightfully be called a "mathematical discovery", as long as we interpret it not as the discovery of a fact in the ideal world of mathematical objects, but rather as an *"extension of grammar"* (which comes close to blurring the line between mathematical discovery and invention), in other words as a new rule that governs certain language games by excluding propositional forms such as P from the game of useful propositions. But why do we adopt this extension of grammar? What is the extra-mathematical use for such an exclusion of P from certain language games? By raising these questions, Wittgenstein draws our attention to the fact that the use of Gödel's theorem is similarly unclear as the use of a proposition such as P itself, as we have not given either any use:

> Gödels Entdeckung ist eine mathematische Entdeckung. Wenn nun eine solche sich als Ausbau der Grammatik auffassen läßt, {was // welches} ist die grammatische Bedeutung der Konstruktion.
>
> Könnte man das auch so ausdrücken: {Welches ist die außermathematische Verwendung des Gödelschen Theorems. // Welche, *außermathematische* Verwendung können wir dem Theorem Gödels geben?}
>
> Welche Verwendung haben wir für einen Satz, der seine eigene Unbeweisbarkeit mathematisch behauptet? [Ms-163, 41v.3–42v.2]

> Gödel's discovery is a mathematical discovery. If such a discovery can be understood as an extension of grammar, {what // which} is the grammatical meaning of the construction.
>
> Could it also be expressed like this: {which is the extra-mathematical use of Gödel's theorem. // Which, *extra-mathematical* use can we give to Gödel's theorem?}
>
> What use do we have for a proposition that mathematically asserts its own unprovability?

As Wittgenstein has made clear in several remarks before, this lack of use is what distinguishes both a proposition such as P and Gödel's theorem from other (non-diagonalised) propositions and proofs: The similarity between these cases misleads us into thinking that a contradiction *must* be avoided by excluding P from our language game, because a contradiction in the non-diagonalised case spells trouble. By extending this analogy to the diagonalised case, we fail to see that Gödel's result demonstrates a *logical* impossibility, in other words a rule of grammar.

From the perspective of Wittgenstein, this conceptual confusion is a direct consequence of the careless use of the notion "interpreted according to the meaning of the terms of PM" / "inhaltlich gedeutet"[52], as this idea "is based on the conception of mathematics as {a // the} physics of 'mathematical objects'", in other words the idea of mathematics as *ultraphysics*:

> 'Inhaltlich deuten' müßte heißen: *anwenden*; & zwar, etwa, auf die, durch diese Worte angedeutete, Weise anwenden.
>
> 'Inhaltlich gedeutet besagt diese Formel ... ' heißt also: "diese Formel kann man in die Worte kleiden: ... "
>
> Die ganze Idee des inhaltlichen Deutens beruht auf der Auffassung der Mathematik als {einer // der} Physik der 'mathematischen Gegenstände'.
>
> Ich will doch immer sagen: {Mathematische Wahrheit & Falschheit entspricht in ihrer Anwendung nicht (der) Wahrheit & Falschheit nicht-mathem. Sätze // Wahr & falsch in der Mathematik entspricht in der Anwendung auf Erfahrungssätze nicht dem Gegensatz wahr-falsch}, sondern der Unterscheidung von Sinn & Unsinn.
> Einer math. Unmöglichkeit entspricht die Ausschaltung einer Satzform aus der Klasse der Erfahrungssätze. [Ms-163, 46r.2–46v.2]
>
> ---
>
> 'To interpret according to its meaning in the system' should read: *to apply*; & more precisely, to apply in the way implied by these words.
>
> 'Interpreted according to its meaning in the system, this formula says ...' thus means: "this formula can be clothed in the words: ..."
>
> The whole idea of interpreting according to the meaning in the system is based on the conception of mathematics as {a // the} physics of 'mathematical objects'.
>
> I always want to say: {Mathematical truth & falsity does not correspond in its application to (the) truth & falsity of non-mathematical propositions // True & false in mathematics does not correspond in its application to empirical propositions to the opposition true-false}, but to the distinction of sense & nonsense.
> A math. impossibility corresponds to the elimination of a propositional form from the class of empirical propositions.

The above remarks thus constitute a summary of Wittgenstein's views on Gödel's theorem, clarifying the extension of grammar in terms of an "elimination of a propositional form from the class of empirical propositions" and emphasising the danger of interpreting it as an ultraphysical discovery. Although the remarks that follow in Ms-163

52 Gödel uses the terms "inhaltlich interpretiert" and "inhaltlich gedeutet" on 5 pages (Gödel, 1986, pp. 146, 148, 150, 170, 172), which are not consistently translated using the same expression into English.

continue to investigate the role of "interpreted according to..." in different contexts, they gradually depart from Gödel's theorem and will not be discussed here.

The last remarks from Ms-163 that should be highlighted in this chapter occur 10 pages later and concern diagonal proofs in general. The two remarks show that Wittgenstein's investigation of Gödel's proof is primarily driven by his interest in the *kind of the proof*, namely the diagonal method, and its tendency to "change our concept of the system", notably our *extra-mathematical* concept, while making it appear as if the proof demonstrated an *intra-mathematical (ultraphysical) discovery*:

> Wenn der Diagonalbeweis etwas tut, {so ist es, daß er unsern Begriff vom System *ändert*. // *so ändert* er unsern Begriff vom System.}
>
> Hier muß man aber unterscheiden zwischen dem Begriff *in* der Math. & außerhalb der Math. Nur von *diesem* müssen wir sagen er habe sich geändert.
> Hier darf man nicht dogmatisch sein wollen: Von manchem neuen Beweis wird man zu sagen geneigt sein, er ändere unsern Begriff, von manchem – sozusagen trivialen – nicht. Aber für uns ist gerade der Übergang zwischen der Geneigtheit, das eine, & der, das andere zu sagen, {das Wichtige // wichtig}. [Ms-163, 55r.2–55r.3]

> If the diagonal proof does anything, {it is that it *changes* our concept of the system. // *it changes* our concept of the system.}
>
> But here one has to distinguish between the concept *in* math. & outside of math. Only of *the latter* must we say that it has changed.
> Here one must not want to be dogmatic: Of some new proofs one will be inclined to say that they change our concept, of some - so to speak trivial ones - not. But for us it is precisely the transition between the inclination to say one thing and the other that is important.

When Wittgenstein writes about our freedom to decide what to do with the contradiction arising from Gödel's construction, he is not advocating for any single logical position, because this would come down to making the same error as Gödel in his informal introduction, by prematurely deciding the use of a concept solely on the basis of a mathematical proof, in the absence of a form of life that provides the concept with context. Wittgenstein's suggestion of 'letting the contradiction stand' is not an invitation for trivialism and even less a misunderstanding of Gödel's proof, but only a warning against a one-sided diet in the philosophy of mathematics.

Wittgenstein shows that Gödel's proof is not a fundamental discovery in an 'ultraphysical' sense (that is to say, comparable to a 'natural' law in the platonic realm of logic) and that many of the later interpretations of the proof are philosophical abuses that venture far beyond what the mathematical proof itself (excluding its prosaic and informal introduction) attempts to show. A philosophically clear picture will only emerge if we understand the proof as a valuable observation about a *logical impossibility* and therefore as a reflection of our rules of language.

3 TURING, MACHINES AND DECIDABILITY

> Turings 'Maschinen'. Diese Maschinen sind ja die *Menschen*, welche kalkulieren. Und man könnte, was er sagt, auch in Form von *Spielen* ausdrücken. Und zwar wären die interessanten Spiele solche, bei denen man gewissen Regeln gemäß zu unsinnigen Anweisungen gelangt. Ich denke an Spiele ähnlich dem "Wettrennspiel". Man erhielte etwa den Befehl "Setze auf die gleiche Art fort"; wenn dies keinen Sinn ergibt, etwa weil man in einen Zirkel gerät; denn jener Befehl hat eben nur an gewissen Stellen Sinn. (Watson.[1]) [Ms-135, 59v.2 / *BPP I* §1096]

> Turing's 'Machines'. These machines are *humans* who calculate. And one might express what he says also in the form of *games*. And the interesting games would be such as brought one *via* certain rules to nonsensical instructions. I am thinking of games like the "racing game". One has received the order "Go on in the same way" when this makes no sense, say because one has got into a circle. For any order makes sense only in certain positions. (Watson.) [*RPP I* §1096]

Wittgenstein's above remark is one of only three in the *Nachlass* directly referring to Alan Turing. The other two remarks (Ms-161, 11v.1 and Ms-161, 11v.2, both in English) only appear in a pocket notebook from 1939 and revolve around issues not directly related to Turing machines.[2] Even though Wittgenstein's discussion of Turing's seminal paper "On Computable Numbers, with an Application to the *Entscheidungsproblem*" (Turing, 1936) is limited to the above remark, it is important in so far as it 'survived' the draft stage and does not only appear in one of Wittgenstein's notebooks (Ms-135, 59v.2, where it appears shortly after remarks explicitly dated as "30.7.[1947]"), but was also carried over into two typescripts (Ts-229, 448.1 and Ts-245, 319.3), suggesting that Wittgenstein was at least somewhat satisfied with the quality of the remark.

Given that the above remark is the only remark on Turing that moved beyond the draft stage, it might seem questionable to draw a close connection between the two authors, let alone use it as a starting point to investigate Turing's writings from the perspective of Wittgenstein. However, even though direct references to Turing are scarce in Wittgenstein's *Nachlass* (and vice versa), there are nevertheless a number of similarities and implicit connections between Turing and Wittgenstein that justify a closer examination of Turing's notion of "computing machines" and his diagonal argument from the perspective of Wittgenstein.

---

1 Refers to Alister Watson, who "discussed the Cantor diagonal argument with Turing in 1935 and introduced Wittgenstein to Turing" (Floyd, 2012, p. 26, see also Watson, 1938).

2 They were most likely written in the context of Wittgenstein's "Lectures on the Foundations of Mathematics" from the same year, where Turing was an active participant.

The first and most obvious connection is Turing's participation in Wittgenstein's "Lectures on the Foundations of Mathematics" from 1939, which are better described as discussions between Wittgenstein and his students than traditional lectures, and where Turing played the role of Wittgenstein's most prolific interlocutor. Their most well-known exchange, on the importance of consistency in formal systems for practical applications such as the construction of bridges, is intimately connected with the issues that will be examined in the present chapter (their exchange will be discussed in more detail below). Briefly, Turing's preoccupation with issues of consistency is closely connected to the aforementioned 1936 paper on computing machines, where the assumption of a general procedure for deciding whether or not a particular machine comes to a halt leads to a contradiction and thus sets the stage for the theoretic limitations of these machines. Similar to Gödel's result that consistent systems of a particular structure must be incomplete, in other words unable to prove all true propositions of the system within the system, Turing's seminal paper demonstrates the fundamental limitations of computing machines if the computations made by these machines are assumed to be consistent.

Questions of consistency and contradiction are an important and recurring theme in Wittgenstein's mathematical writings, a theme which forms the second connection between Wittgenstein and Turing. Given that that Wittgenstein never wrote a sustained investigation focused solely on these questions and that remarks on contradictions in mathematics instead appear in a number of different contexts in the *Nachlass*, it might perhaps be objected that consistency is of lesser importance to Wittgenstein than to Turing. But an investigation focused solely or primarily on these issues would have hardly made sense for Wittgenstein in light of his conception of philosophy: As this chapter attempts to show, consistency should not be understood as an abstract requirement that must hold for all of mathematics or for everything that we might call "computable" (contra Turing), but is instead a concept whose importance can, following Wittgenstein, only be judged in the context of the use of a particular formal system. These claims will of course have to be substantiated in the following sections. For now, it is mainly relevant to emphasise that the fragmented nature of Wittgenstein's writings 'on' consistency is not by itself an argument against drawing from a range of different documents while discussing Turing, as this fragmentation is a direct result of Wittgenstein's philosophical method, not a value judgment on Wittgenstein's part.

Third and most importantly, Turing's paper employs at its heart a diagonal argument, as Turing himself points out. Wittgenstein may not have written much about Turing's 1936 paper itself and it is unclear whether he was aware of the finer mathematical details, but he wrote extensively about Cantor's diagonal argument (Chapter 1) and Gödel's application of the diagonal method (Chapter 2). These remarks are for the most part applicable in the context of Turing and in fact become much more illuminating when applied to Turing's computing machines, because Turing's machines provide a more 'practical' backdrop for Wittgenstein's thought than set theory or logic.[3] Applied to Turing's diagonal argument, it is easier to see *why* Wittgenstein raises certain objections or challenges, whereas these remarks often seem to amount merely to misunderstandings of standard mathematical practice in the more intra-mathematical contexts of Cantor and Gödel.[4]

In a nutshell, part of the allure of Turing's computing machines is that Turing draws from well-known concepts by picturing computation as the (idealised) execution of calculating clerks, as Wittgenstein correctly points out in the remark quoted above ("humans who calculate"). The impossibility results derived in Turing's formalism (for example, that no general procedures exist to decide certain well-defined questions such as whether a machine ever comes to a halt) thus seem to hold not only in this *idealised* formal system, but rather as "ultra-physical" limitations (Ts-213, IIr.5; Ts-222, 11.3 / *RFM I* §8, see Section 0.1) that hold *all the more* in any *practical* realisation of computing machines, no matter whether they are implemented by rule-following humans or built on top of the fastest digital computer imaginable. The theoretical results proved by Turing thus seem to carry over into our physical world and put a limit on what we can ever hope to do, similar to how the speed of light presents a hard limit to how fast we can ever hope to travel. A philosophical investigation along the lines of Wittgenstein cannot and does not question the theoretical result itself, neither its consequences inside the idealised formal system, but rather the tendency of the 'prose' of the mathematical result to present a *logical* impossibility as an *(ultra-)physical* impossibility. As this tendency originates in the use and abuse of well-known concepts

3 Floyd, 2019, p. 271 traces this "homespun" aspect, which differentiates Turing from Gödel and other equivalent logical formalisms, back to Wittgenstein's influence. This chapter will not delve into questions of whether and to which degree Turing was influenced by Wittgenstein, but the practicality of Turing's idealised machines cannot be denied and is one of the reasons why Turing's diagonal argument is worth discussing from the perspective of Wittgenstein.

4 Wittgenstein's background might have played a role here: Following the wishes of his father, he originally studied to be an engineer and his education in mathematics was at first driven by practical considerations. Only during his time in Manchester did he turn to pure mathematics and the foundational issues raised by Frege and Russell (Monk, 1991, pp. 3–36). Not too much weight will be placed on these biographical aspects, but his frequent use of mechanical machines as examples even in his later years shows to which degree his thinking remained influenced by such a practical perspective. As this chapter attempts to show, Wittgenstein's 'engineering mindset' is an important factor in understanding his philosophical critique and offers a view of computation that is quite different from Turing's more mathematical perspective.

in new mathematical contexts (which appear to be innocuous applications and generalisations of the more concrete cases), the antidote is a *surveyable representation* of these concepts and their uses, so that it becomes clear where the analogies with the idealised mathematical context hold and where they end.

As the theoretical model of computation with the most obvious connections to practical computations, Turing's computing machines are not just 'a' mathematical calculus, but rather 'the' way to formalise mechanical rule following. Being universal and in terms of theoretical computational capabilities equivalent to any other universal model of computation (via the Church-Turing thesis), Turing's computing machines are a mathematically precise framework for anything that can be done without any kind of mathematical or psychological insight, purely by training a machine or a human to follow a set of rules. The programs executed by a computing machine in Turing's model thus correspond rather directly to those mathematical language games that we can fully codify as precise and unambiguous rules, with obvious connections to Wittgenstein's remarks on the view of mathematics as a formal game and his frequent example of students being trained to do arithmetic.

For all these reasons, the remark on Turing quoted above is a promising starting point for an investigation of Turing's paper from a Wittgensteinian perspective. It should be pointed out that the remark will hereafter be read primarily in the context of the mathematical results of Turing's 1936 paper, not as a more general remark on the similarities between Turing machines and human thought processes that aims to investigate what it would mean to speak of 'thinking' machines, let alone as a comment on Turing's position in the philosophy of mind.[5] An interpretation in the context of the philosophy of mind, which would focus on how well Turing's concept captures the thought processes or actions of human calculators, seems natural in so far as Turing's paper appears most open to philosophical attacks in section §9, where he offers three kinds of arguments that attempt to show how his definition of "computable" corresponds to what "would naturally be regarded as computable" (Turing, 1936, p. 74). All three kinds of arguments include appeals to intuition in some form or other and thus require Turing to leave the more mathematical ground of the rest of his paper at least temporarily. While certainly worthy of attention, the focus hereafter will be primarily on the philosophy of mathematics and deal mostly with Turing's use of the diagonal argument, at least while discussing Turing's 1936 paper. In later sections, the more mathematical reflections will then shed

5 An example of an extensive investigation of these matters is Shanker, 1998, which claims that Wittgenstein's remark should be read as a critique of the philosophical misunderstanding inherent in the Church-Turing thesis and more specifically the confusion between 'Mechanical rule-following' and 'Following a "mechanical rule"' (Shanker, 1998, p. 27).

new light on the similarities between machines and calculating clerks, but a full discussion of the issues raised in the philosophy of mind would go beyond the scope of this chapter.

The outline of this chapter is as follows: The first two sections will introduce the main topic by giving an overview of a crucial part of Turing's seminal paper, his diagonal argument (Section 3.1), and then discuss remarks by Wittgenstein on mechanical rule following that are relevant in the context of Turing's argument (Section 3.2). The next two sections form the core of this chapter and will present a philosophical critique of the concepts leading to the contradiction in the application of the diagonal argument: First by 'sidestepping' the consequences and exploring philosophical reasons why such a contradiction might be harmless (Section 3.3), then 'head-on' by investigating the philosophical consequences of accepting such a contradiction inside the system instead of excluding it (Section 3.4). The next two sections return to Turing's writings and discuss the previous arguments in light of Turing's view on the differences between machines and mathematicians (Section 3.5) and the role of using "new techniques" in mathematics (Section 3.6). The chapter concludes with an attempt to clearly articulate the importance of Turing's notion of computability without falling prey to some common conceptual confusions (Section 3.7).

3.1 Turing's diagonal argument

To give a rigorous definition of "computable numbers", Turing defines "computing machines" operating on an infinite tape of squares of symbols (e.g. the blank symbol, '0', '1', 'x', 'y', 'z') and producing sequences of '0's and '1's on the tape, which are interpreted as the expressions of real numbers in binary notation. A computing machine has access only to the single symbol in the "scanned square" and can read the symbol, erase it, print a different symbol in the square or move left or right to an adjacent square. At any point in time, a computing machine is only in a single state – or "*m*-configuration" in Turing's terminology – and can switch to another *m*-configuration as the result of the instruction determined by the current *m*-configuration and the scanned symbol (Turing, 1936, p. 59). The complete table of instructions fully specifies the machine and ensures that the execution of a machine proceeds in a systematic and mechanised fashion (at least in the case of "automatic machines" or "*a*-machines", which is the only type of machine used by Turing in his paper, cf. Turing, 1936, p. 60). Each machine always starts in a specified *m*-configuration and then proceeds according to its instructions. An example of a simple table with instructions for 4 *m*-configurations is as follows (from Turing, 1936, p. 61, with gothic letters replaced by roman letters):

Configuration		Behaviour	
m-config.	*symbol*	*operations*	*final m-config.*
b	None	P0, R	c
c	None	R	e
e	None	P1, R	k
k	None	R	b

Table 4: Example Turing Machine Instruction Table

Started in the *m*-configuration 'b' and with a blank tape, the machine specified by this table will read the scanned symbol (which is 'None', since the tape is blank) look up the instruction with the configuration corresponding to the pair of *m*-configuration 'b' and symbol 'None' (which will be the instruction in the first row of the table), execute the operations 'P0, R' (thus printing the symbol '0' and then moving to the right) and finally switch to the *m*-configuration 'c', executing the instruction for that m-configuration in the next step and so forth. The result of the execution will thus be an endless sequence of '0', '1', '0', '1', . . . , interspersed by blank squares, which follows Turing's convention of printing the result of a computation "only on alternate squares", called "*F*-squares", leaving each "*E*-square" immediately to the right of each "*F*-square" reserved for temporary computing purposes (Turing, 1936, p. 63). If an *E*-square contains a nonblank symbol, its corresponding *F*-square is said to be "marked" by the symbol in the *E*-square.

Crucial to the following arguments are the concepts of "computing machines", "circular and circle-free machines" as well as "computable sequences and numbers". Turing defines them as follows:

> If an *a*-machine prints two kinds of symbols, of which the first kind (called figures) consists entirely of 0 and 1 (the others being called symbols of the second kind), then the machine will be called a computing machine. If the machine is supplied with a blank tape and set in motion, starting from the correct initial *m*-configuration, the subsequence of the symbols printed by it which are of the first kind will be called the *sequence computed by the machine*. The real number whose expression as a binary decimal is obtained by prefacing this sequence by a decimal point is called the *number computed by the machine*.
>
> [. . .]
>
> If a computing machine never writes down more than a finite number of symbols of the first kind, it will be called *circular*. Otherwise it is said to be *circle-free*.
>
> [. . .]
>
> A sequence is said to be computable if it can be computed by a circle-free machine. A number is computable if it differs by an integer from the number computed by a circle-free machine. [emphasis by Turing, 1936, pp. 60–61]

The significance of these definitions becomes clear once the "universal computing machine" is defined and the diagonal argument is then applied based on these definitions. For now, it will suffice to point out that Turing uses the terms "circular" and "circle-free" somewhat counter-intuitively, since in his terminology a computing machine that never stops and prints '0' or '1' endlessly (such as the simple example considered above) is "circle-free", whereas a machine that comes to an end is "circular", because there is no instruction that could enter a different *m*-configuration or change the scanned symbol. The reason for this lies in Turing's use of computing machines for the purpose of constructing computable sequences, infinite sequences of '0' and '1' as figures in the *F*-squares: only a computing machine that will forever go on printing these figures can be said to construct a computable number with an infinite number of binary decimal places, whereas a "circular" number gets 'stuck in a loop' at some point and will never print more than a finite number of binary decimal places. This is true even for numbers which we would usually consider to have only finitely many binary decimal places, such as 0.1, which is computed as .1000... by a circle-free machine that prints '1', '0', '0', '0', ... endlessly on its tape. In this way Turing's definitions differ from modern usage, where computing machines are considered to "halt" when they enter an end state and to be circular when they enter a loop that can never lead to such an end state.[6]

For Turing's main argument to work, it is necessary for computing machines to be able to operate on computing machines by accepting their definitions as input on the tape. To achieve this, Turing defines an encoding of instruction tables as sequences of symbols and later on as numbers. The details are not essential for the philosophical discussion of his results and other encoding schemes would have lead to analogous results. At a high level, the encoding proceeds by encoding each row in the instruction table as a tuple of 5 components (matched *m*-configuration, matched symbol, replacement symbol, movement operation and next *m*-configuration), and then encodes the whole table as a sequence of these component tuples separated by the special separator symbol ';'. Such an encoding presupposes that all instructions of a table follow the same 5-component tuple structure, which is not true for the simple example table above (where two rows include only a movement operation, the other two rows a movement + print operation). Turing defines a "standard form" for this purpose and consequently calls the encoding of a computing machine in standard form a "standard description" or "S.D", while the straightforward decimal encoding of this standard description is called the "description number" or "D.N" (Turing, 1936, p. 67). This implies that computing machines with their standard descriptions and the computable sequences determined by them are both enumerable. The description

6 Cf. Hopcroft, Motwani, and Ullman, 2001, p. 327, Sipser, 2012, p. 170

number of a circle-free machine is called "satisfactory" in Turing's terminology (Turing, 1936, p. 68).[7]

Turing then goes on to list several subsidiary definitions as abbreviated tables (with variables than can be substituted by concrete *m*-configurations and symbols to obtain full instruction tables), but these are of little interest for a philosophical discussion of his argument and serve mainly to define the universal computing machine $\mathscr{U}$, which, when supplied with the standard description of a machine $\mathscr{M}$ as input on the tape, computes exactly the same sequence as $\mathscr{M}$ would if $\mathscr{M}$ had been started on a blank tape.

Now that the stage is set, Turing applies what he calls the "diagonal process", which is also where Wittgenstein's introductory remark will finally come into play. After first demonstrating an incorrect application of a diagonal argument, which purportedly shows that computable sequences cannot be enumerable, Turing shows that there cannot be any "general process" for "finding out whether a given number is the [description number] of a circle-free machine [. . .] in a finite number of steps" (Turing, 1936, p. 72). Suppose that there is such a machine, called $\mathscr{D}$, which tests whether or not the machine described by a given standard description is circular. It is then possible to enumerate all circle-free standard descriptions (and thus all computable sequences), by simply iterating through all standard descriptions one by one and checking each one for circularity by testing it with $\mathscr{D}$. By incorporating the universal machine $\mathscr{U}$, a combined computing machine $\mathscr{H}$ can be constructed which, for the N-th circle-free computing machine tested by $\mathscr{D}$, employs $\mathscr{U}$ to calculate the first N figures of the sequence computed by the N-th circle-free computing machine (Turing, 1936, p. 73).

The combined computing machine $\mathscr{H}$ must be circle-free: $\mathscr{D}$ is circle-free by assumption and $\mathscr{U}$ only calculates a finite number of figures for each machine (namely N for the N-th circle-free machine), which will finish in a finite number of steps, since the calculation is only performed for those machines that are circle-free and thus guaranteed to always produce figures in a finite number of steps.

But if the combined computing machine $\mathscr{H}$ is circle-free, it will be one of the circle-free standard descriptions iterated by $\mathscr{H}$ at some point N. The universal machine $\mathscr{U}$ will then be employed to calculate the first N figures in the sequence computed by $\mathscr{H}$, which is unproblematic for the figures up to and including figures $N - 1$, for which

7 In the following text, the precise distinction between 'circle-free machines' and 'satisfactory standard descriptions' (of those machines) will, for the sake of brevity, often be ignored in cases where the usage is unambiguous or not relevant to the philosophical argument, because it is usually clear that whenever a computing machine is operated on, it is actually this machine's *standard description* that is operated on by another machine. For example, "the N-th circle-free computing machine tested" should be read as "the N-th satisfactory standard description tested, describing the N-th circle free computing machine".

$\mathscr{H}$ will iterate through the $N-1$ first standard descriptions and calculate their figures as far as necessary using $\mathscr{U}$. But to calculate the N-th figure, $\mathscr{H}$ would need to calculate the first N figures of the N-th circle-free computing machine, which is $\mathscr{H}$ itself, which again would include the calculation of $\mathscr{H}$ up until the N-th figure, and so forth, which means that $\mathscr{H}$ cannot be circle-free. As a consequence, there can be no circle-free machine $\mathscr{D}$ (Turing, 1936, p. 73).

This application of $\mathscr{H}$ to itself is the "Zirkel" that Wittgenstein's introductory remark is referring to. Hence why "Setze auf die gleiche Art fort" does not make any sense for the N-th figure of $\mathscr{H}$, even though it was perfectly clear what to do in the $N-1$ cases before it. This makes the diagonal case akin to a game such as the "Wettrennspiel", which is most likely referring to a board game depicting a circular horse track.[8]

That Wittgenstein refers to this particular part in Turing's argument becomes more clear if the following remark is considered (which follows the Turing remark in all documents, namely in Ms-135, 60r.2 and, in a slightly changed version, in Ts-229, 448.2 and Ts-245, 319.4, a version which Wittgenstein also included in his "Zettel" collection of explicitly selected snippets, as Ts-233b, 61.2 / *Z* §694):

> Eine Variante des Cantor'schen Diagonalbeweises:
>
> $N = F(k, n)$ sei die Form der Gesetze für die Entwicklung von Dezimalbrüchen. N ist die n-te Dezimalstelle der k-ten Entwicklung. Das Gesetz der Diagonale ist dann: $N = F(n, n) \overset{\text{Def.}}{=} F'(n)$.
>
> Zu beweisen ist, daß $F'n$ nicht eine der Regeln $F(k, n)$ sein kann. Angenommen, es sei die 100ste. Dann lautet die Regel
>
> zur Bildung von $F'(1)$ $F(1, 1)$
> von $F'(2)$ $F(2, 2)$ etc.
>
> aber die Regel zur Bildung der 100sten Stelle von $F'(n)$ {wird // lautet} $F(100, 100)$; D.h., sie sagt uns nur, daß die 100ste Stelle sich selber gleich sein soll, ist also für $n = 100$ *keine* Regel.
>
> Die Spielregel lautet "Tu das Gleiche, wie ... !" – und im besondern Fall wird sie nun "Tu das Gleiche, wie das, was Du tust!" [Ts-233b, 61.2 / *BPP I* §1097 / *Z* §694]

> A variant of Cantor's diagonal proof:
>
> Let $N = F(k, n)$ be the form of a law for the development of decimal fractions. N is the n^{th} decimal position of the k^{th} development. The law of the diagonal is then: $N = F(n, n) \overset{\text{Def.}}{=} F'(n)$.
>
> To prove: that $F'(n)$ cannot be one of the rules $F'(k, n)$. Assume it is the 100^{th}. Then the rule for the construction
>
> of $F'(1)$ runs $F(1, 1)$
> of $F'(2)$ $F(2, 2)$ etc.
>
> but the rule for the construction of the 100^{th} position of $F'(n)$ becomes $F(100, 100)$ i.e. it shews only that the 100^{th} place is supposed to be the same as itself, and so for $n = 100$ it is *not* a rule.

8 Floyd, 2012, p. 25: "I presume that Wittgenstein is thinking of a board game in which cards are drawn, or knobs turned so as to move pieces in a simulated horse race."

> The rule of the game runs "Do the same as ..." – and in the special case it becomes "Do the same as what you do". [*RPP I* §1097 / *Z* §694]

At the crucial point on the diagonal, where $k = n$, we have to apply the same rule as we did in all previous cases, but as Wittgenstein points out, the 'rule' for this case merely says to follow *this very rule itself and nothing else*, which means that *it is no rule*. The 100th decimal place of F is *underdetermined*: While in all other cases there is some rule that determines the decimal place (namely the rule describing the k-th sequence), there is no such determining rule in the case of $F(n, n)$, which leaves open how to proceed, since the rule takes the tautological form "Tu das Gleiche, wie das, was Du tust!" / "Do What You Do".[9]

	f_1	f_2	f_3	f_4	...	f_N	...
k_1	[0]	1	0	0	...	1	...
k_2	1	[1]	1	0	...	1	...
k_3	0	1	[0]	0	...	0	...
k_4	0	0	0	[1]	...	1	...
...	...	...	...	...	...	...	...
k_N	0	1	0	1	...	[?]	...
...	...	...	...	...	...	...	...

Table 5: Application of the Diagonal Argument

In contrast to Wittgenstein's variant, diagonal arguments such as Cantor's original version usually construct a diagonal which leads not to a *tautological*, but rather to a *contradictory* rule, where the decimal place at the crucial point of the diagonal must be 1 when assumed to be 0, but 0 when assumed to be 1. Such a rule could be said to be *overdetermined* like a game rule of the form "Do What You Do *Not* Do".[10] Importantly, this would correspond to the "halting problem" for Turing machines, which was not explicitly stated by Turing

9 The wording "Do What You Do" is borrowed from Floyd, 2012, who notes that neither Turing's diagonal argument nor Wittgenstein's variant employ a *paradoxical* rule, in contrast to what is nowadays known as the "Halting Problem". While this observation is certainly correct, it should be pointed out that Turing's argument nevertheless works by *reductio ad absurdum* and thus hinges on producing a contradiction: "It [the decision procedure $\mathscr{D}$] must test whether K is satisfactory, giving a verdict "s" or "u". [...] Thus both verdicts are impossible and we conclude that there can be no machine $\mathscr{D}$." (Turing, 1936, p. 73) For Turing's argument to work, it is unimportant whether the 'rule' used is tautological or contradictory, the result is in both cases paradoxical. In the following text, terms such as "paradoxical" will thus sometimes be used to refer to machines or rules that are, strictly speaking, not paradoxes, but that nevertheless lead to those paradoxical results.

10 Priest, 2006b, p. 15, uses the terms "underdetermine" and "overdetermine" in a very similar way in the context of sentences of the form "This sentence is True" and "This sentence is False".

himself, but only by later commentators[11], and follows rather directly from Turing's diagonal argument. Very briefly, the halting problem results if, starting again from the assumption that a decision procedure machine $\mathscr{D}$ exists for deciding whether or not an arbitrary computing machines halts, a machine $\mathscr{H}$ is constructed that simply halts if $\mathscr{D}$ applied to $\mathscr{H}$ predicts that $\mathscr{H}$ will not halt, but enters an infinite loop if $\mathscr{D}$ applied to $\mathscr{H}$ predicts that $\mathscr{H}$ will in fact halt. The result of the decision procedure $\mathscr{D}$ will therefore always be incorrect in the case of the machine $\mathscr{H}$, which proves that there cannot exist a general decision procedure to solve the halting problem.[12]

Apart from a minor change of the function name[13], the version in the notebook differs from the typescript version quoted above by including an additional paragraph before "Die Spielregel lautet", marked with the curved 'S', frequently used by Wittgenstein to indicate his dissatisfaction with the quality of a remark:

> Ich habe nämlich immer das Gefühl gehabt, der Cantorsche Beweis tue zwei Dinge, scheine aber bloß eines zu tun.
>
> ---
>
> I have always had the feeling that Cantor's proof does two things, but appears to do only one.

In light of the section mark, not too much interpretative weight should be placed on this phrase, but it might shed a light on Wittgenstein's interest in the matter: The above diagonal argument appears to apply the word 'rule' uniformly, because at all points in the proof we seemingly proceed by merely applying the rules of the game. But in the case of $k = n$, the 'rule' of $F(n, n)$ is not a rule in the same way as all the other rules before it, since it does not give us a concrete rule that we could follow. Seen from this angle, it would be natural to say that the diagonal argument actually uses *two different concepts of rules*, non-tautological 'regular' rules as well as the tautological 'exceptional' rule $F(n, n)$. But of course all the rules were assumed to be concrete, non-tautological, or else they would not determine sequences of decimal places, which leads to the contradiction that the 'rule' $F(n, n)$ cannot be the same kind of rule that was assumed in the beginning. In other words, the proof seems to use the word 'rule' uniformly throughout the proof, but in reality a different use of the word 'rule' sneaks in and by 'applying a rule' the proof does two different things. F is both applied as a concept and also used as an object, but the proof slightly obscures these different uses.[14] If this interpretation of Wittgenstein's "tue zwei Dinge, scheine aber bloß

---

11 See Turing and Copeland, 2004, p. 40.

12 For a more detailed and rigorous description of the problem, see Davis, 1985, p. 70.

13 Wittgenstein uses 'ϕ' in Ms-135, but 'F' in the typescripts.

14 While discussing Wittgenstein's remark on Turing's diagonal argument, Floyd, 2019, p. 286 points this out in the context of the "*positive* Russell set" $S = \{x|x \in x\}$, which works analogously to Wittgenstein's variant of the diagonal argument: "In the above argument an apparently unproblematic way of thinking is applied, but two different ways of thinking about S are involved. For there is the thinking of S as an object or

eines zu tun" is correct[15], it is not surprising that Wittgenstein discarded it: The way it is phrased, it might be read as suggesting that only the non-uniform-use perspective, the "zwei Dinge" view, is *the right one* and that Wittgenstein was trying to criticise the diagonal proof, which cannot be the intention if his claim of non-interference in purely mathematical matters is to be taken seriously.[16]

In the end, it comes down to either emphasising the similarities or the differences between the different uses: Compared to Turing's application of the diagonal argument, Wittgenstein's variant places a stronger emphasis on the differences and points out that a 'rule' with no steps to follow can hardly be called a rule. Turing, on the other hand, certainly still considers the hypothetical machine $\mathscr{H}$ to be a computing machine (with its characteristic operating mode of mechanically following steps and thus rules), simply one for which the circular/circle-free question cannot be decided. We are here in a 'border region' between rules and non-rules, which is one of the reasons why it is easy to be lead astray by conceptual confusions.

3.2 CALCULATING CLERKS

In addition to reframing Turing's proof in terms of a language game with an underdetermined rule, Wittgenstein's variant of the diagonal argument also offers an interesting perspective on the "humans who calculate", as it might call into question whether the reaction of human calculators to the "Do What You Do" instruction would be as clear cut as it is in Turing's argument. Let us imagine several different scenarios for these "humans who calculate", with only the first corresponding to the behaviour exhibited by Turing machines:

---

element that is a member of other sets, and the thinking of S in terms of a concept, or defining condition. Similarly, in Turing's [Turing, 1936] proof, there is the unproblematic characterisation of a particular machine, and then there is the difficulty that it must, at one precise point or another, get stuck in a loop, confronted with the command to do what it does."

15 There are other possible interpretations. Mühlhölzer, 2020, pp. 144–45, interprets the "two things" to be first the construction of the *unaltered* diagonal itself and second the *alteration* of the decimal expansion by changing the decimal places. Wittgenstein's variant could then be seen as a 'cleaner' version of Cantor's argument that only does the first thing and thus shows that two different actions are involved. This reading has its merits and would perhaps be unproblematic in the context of one of Wittgenstein's more 'thorough' discussions of the diagonal argument. In the remark at hand, however, the alteration of the decimal expansion is not explicitly mentioned, it thus seems more plausible that the "two things" refer to the dual use of the rule: "die Regel [...] ist also für $n = 100$ keine Regel" and "Die Spielregel lautet [...], im *besondern* Fall *wird* sie nun [...]", suggesting that in the special case of the diagonal the rule does *something else* than in the general case.

16 See *LFM I*, p. 13: "[I]t will be most important not to interfere with the mathematicians. I must not make a calculation and say, 'That's the result; not what Turing says it is.'", but also Ts-227a, 89.2 / *PI* §124 and Ms-124, 82.2–82.3 / *RFM VII* §19.

1. The human calculator might not notice the tautological nature of the rule and simply carry on in a loop, exactly like Turing's computing machines would. This is perhaps hard to imagine in the case of a rule as simple as "Do What You Do", but we could certainly picture an analogous scenario with a longer and less 'surveyable' chain of rules leading to the same tautological situation, such as: "Rule A: Do what rule B says. Rule B: Do what rule C says. Rule C: Do what rule D says. . . . Rule Z: Do what rule A says".

2. The human calculator might notice the senseless instruction and terminate the whole operating procedure, alert a supervisor, or in some other way refuse to continue with the current calculation.

3. The human calculator might notice the senseless instruction, but just shrug and say: "Either way of proceeding from here seems to be fine, since there is nothing that determines the next number, so I will just write 0 and be done with this step."

Before discussing these alternatives in detail, it should be pointed out that the aim here is not to cast doubt on the validity of Turing's argument, which of course follows logically from his definition of computing machines and circularity. The purpose of the alternatives listed here is to show that while Turing's definition of computing machines originally draws its motivation from the stepwise application of simple rules by "humans who calculate", it is not immediately obvious whether this analogy still applies at the crucial point of the diagonal argument, or at the very least whether it might not deserve a closer investigation of the similarities and differences with the practical example that lent it a large part of its persuasiveness. The convincing aspect of Turing's machines, their practical angle (or perhaps, to borrow Wittgenstein's words, their "homespun" / "hausbacken" use[17]) is, after all, one of the most important differentiating factors in comparison to other formalisms used in earlier negative resolutions of the *Entscheidungsproblem* that were later shown to be logically equivalent, such as Church's λ-calculus or "general recursiveness in the sense of Herbrand-Gödel-Kleene"[18], and partly explains why Turing machines have enjoyed such an enthusiastic reception, reaching far beyond the narrow confines of mathematical logic.

The issue with Turing's proof, if there is any, is certainly not mathematical, but only philosophical: Instead of criticising the proof on mathematical grounds, we might ask whether his use of ordinary

17 Floyd, 2019, p. 271: "Our argument is that it is Turing who showed that analysis in the sense of formal logic, the very idea of "simplicity" of formal steps, their transparency and gap-free character, *must* have a "homespun" use."
18 See Turing and Copeland, 2004, p. 45.

terms in the technical context of the proof corresponds to the ordinary uses that motivated the proof. If not, we need to be mindful of the danger of interpreting the results of the proof outside of its mathematical setting, in a way that leads us to mistake a *logical* impossibility inside the proof for a *physical* impossibility that applies outside the proof as well.

If we only consider scenario 1, where the hypothetical machine $\mathscr{H}$ would be thrown into a loop and a human calculator would equally fail to see the futility of the tautological rule, the logical contradiction arises from the assumption that the rule F is a productive, non-tautological rule which always tells us what to do in a stepwise fashion and the fact that $F(n, n)$ is actually tautological and does not provide concrete steps to follow. This logical contradiction thus forecloses the possibility of ever finding a general decision procedure which tells us whether or not a rule can be followed stepwise without ending up in a loop.

As a result, Turing's diagonal argument seems to have practical implications in the context of human calculators that could reach far beyond the immediate proof with its mathematical interest. For example, if some people had been in the process of setting up a (pre-computer era) company offering calculation services by employing a myriad of clerks that exactly follow precise computing machine rules without any other external oversight, the owners of such a company might be worried that the productivity of the company would over time grind to a halt, as more and more clerks 'get stuck' in a loop. They might have noticed that some calculations never finish and thus decided to search for a rule to distinguish between 'productive' and 'futile' calculations. In the face of Turing's proof, their search appears altogether hopeless, as the diagonal argument seems to put a limit on *possibilities in the physical world*, the consequences of which would then doom such a calculation-as-a-service company.

But what the proof actually does is showing the inconsistency between the concept of a rule as used in the proof and the assumption that there is a general decision procedure (for deciding whether a rule can be applied stepwise without ending in a loop) that is *rule-based according to the assumed concept of a rule*. The alternative to giving up the search for such a procedure, then, is to revise the concept of a rule. Such a concept change would of course require a completely new proof and leave the old proof perfectly unimpaired, which is why such a philosophical investigation will always remain conceptual and can never interfere with the work of mathematicians.

But does this not miss the point of Turing's argument? Turing's definition of computing machines ensures that calculations can be performed *mechanically*, without the kind of oversight or insight necessary in the other scenarios described above. Only then does the contradiction produced by machine $\mathscr{H}$ become disastrous, as it seems to put a limit on what can be done by machines or human calculators without oversight. Wittgenstein investigates the role of such a contradiction in Ms-124:

> 'Wir machen lauter legitime – d.h. in den Regeln erlaubte – Schritte, & auf einmal kommt ein Widerspruch heraus. Also ist das Regelverzeichnis, wie es ist, nichts nutz, denn der Widerspruch wirft das ganze Spiel um.' Warum läßt Du ihn es umwerfen?
>
> Aber ich will, daß man nach der Regel soll *mechanisch* weiter schließen können, ohne je zu widersprechenden Resultaten zu gelangen. Nun, welche Art der Voraussicht willst Du? Eine, die Dein gegenwärtiger Kalkül nicht zuläßt? Nun, dadurch ist er nicht ein schlechtes Stück Mathematik – oder: nicht im vollsten Sinne Mathematik. Der Sinn des Wortes "mechanisch" verführt Dich. [Ms-124, 56.3 / *BGM VII* §11]

> 'We take a number of steps, all legitimate – i.e. allowed by the rules – and suddenly a contradiction results. So the list of rules, as it is, is of no use, for the contradiction wrecks the whole game! 'Why do you have it wreck the game?
>
> But what I want is that one should be able to go on inferring *mechanically* according to the rule without reaching any contradictory results. Now, what kind of provision do you want? One that your present calculus does not allow? Well, that does not make that calculus a bad piece of mathematics, – or not mathematics in the fullest sense. The meaning of the word "mechanical" misleads you. [*RFM VII* §11]

Evidently, Wittgenstein is aware of the main reason why a contradiction in a given logical system could be seen as a problem: the requirement of *mechanical* calculability. This requirement can lead us to give up on a game once we see that some of its rules can lead to a contradiction, because it is incompatible with the kind of "provision" / "Voraussicht" we expected. But as Wittgenstein points out, this does not mean that the contradictory calculus necessarily is a bad piece of mathematics. A possible alternative to giving up the game is to accept the contradiction and instead give up the expectation of a particular kind of certainty while playing the game. Leaving aside for now how this would look like for Turing's computing machines, it is important to note that the requirement of total mechanical calculability alone does not compel us to abandon a contradictory calculus. Later in Ms-124, Wittgenstein briefly returns to the issue of mechanical calculations:

> Man folgt der Regel '*mechanisch*'. Man vergleicht sich also mit einem Mechanismus.
>
> "Mechanisch", das heißt: ohne zu denken. Aber *ganz* ohne zu denken? Ohne *nachzudenken*. [Ms-124, 164.4–5 / *BGM VII* §60]

> One follows the rule *mechanically*. Hence one compares it with a mechanism.
>
> "Mechanical" – that means: without thinking. But *entirely* without thinking? Without *reflecting*. [*RFM VII* §60]

If we, as humans, follow the same rules as a computing machine in the same way that a computing machine would, then by following Turing's diagonal argument or Wittgenstein's variant we will of course end up in the same predicament. The general decision procedure assumed by Turing as part of his diagonal *reductio ad absurdum* is then just as unattainable for us humans as it is in the case of Turing machines operated by computers.

3.3 FALLING BRIDGES

Can a contradictory calculus ever be *useful*? We might grant Wittgenstein the objection that a contradictory calculus could still be considered a mathematical calculus, but is such an objection more than a philosophical quibble that completely overlooks the use we make of logical calculi for practical calculations? This is the point raised by Turing in Wittgenstein's *Lectures on the Foundations of Mathematics*:

> The sort of case which I had in mind was the case where you have a logical system, a system of calculations, which you use in order to build bridges. You give this system to your clerks and they build a bridge with it and the bridge falls down. You then find a contradiction in the system. – Or suppose that one had two systems, one of which has always in the past been used satisfactorily for building bridges. Then the other system is used and the bridge falls down. When the two systems are then compared, it is found that the results which they give do not agree. [*LFM XXII*, p. 212]

According to Turing, the motivation for consistent logical calculi is not just an intra-mathematical desire, but rather of a very practical nature. Put crudely, contradictions can make bridges fall down. Sure, we can allow contradictions in a logical system, but then the system cannot be applied to practical tasks, since according to Turing: "You cannot be confident about applying your calculus until you know that there is no hidden contradiction in it" (*LFM XXII*, p. 217). With stakes this high, what are we to make of Wittgenstein's nonchalance regarding contradictions? Does the observation that we humans are usually not petrified by a contradiction not simply boil down to an appeal to "common sense"?

> *Turing:* You seem to be saying that if one uses a little common sense, one will not get into trouble.
>
> *Wittgenstein:* No, that is *NOT* what I mean at all. – The trouble described is something you get into if you apply the calculation in a way that leads to something breaking. This you can do with *any* calculation, contradiction or no contradiction. [*LFM XXII*, p. 219]

This exchange between Turing and Wittgenstein is important, because Wittgenstein's attitude towards contradictions in the *Nachlass* can often appear as appealing to common sense as a way of resolving contradictions. But here, Wittgenstein explicitly emphasises that this is not the case. An appeal to common sense would of course fall flat in the face of Turing's ('mechanically' calculating) computing machines, where the notion of common sense is not applicable in the same way as it is for human clerks. Rather, Wittgenstein questions the privileged position that contradictions seem to enjoy when it comes to questions of 'getting into trouble'. Turing treats the application of a calculus as an all-or-nothing situation: Either the calculus is consistent and can be applied or it is inconsistent and thus worthless for practical applications. However, whether or not a bridge falls down *in practice* depends on many different factors and will involve engineering assumptions that might turn out to be insufficient in extreme events such as a particularly strong earthquake. Viewed from this perspective, a logical contradiction is just one of these different factors and can perhaps even be ignored if the resulting calculation errors are rare enough in practice, just like how small calculation errors in consistent calculi can sometimes be ignored or mitigated. And even if a bridge might fall down due to a contradiction, this does not necessarily need to call the whole application of the calculus into question:

> – Hier ist ein Widerspruch: Aber wir sehen ihn nicht & ziehen Schlüsse aus ihm. Etwa auf mathematische Sätze; & auf falsche. Aber wir erkennen diese Schlüsse an. – Und bricht nun eine von uns berechnete Brücke zusammen, so finden wir dafür eine andere Ursache, oder sagen, Gott habe es so gewollt. War nun unsre Rechnung falsch; oder war es keine Rechnung?
>
> Gewiß, wenn wir als Forschungsreisende nun die Leute {betrachten // beobachten}, die es so machen, werden wir vielleicht sagen: diese Leute rechnen überhaupt nicht. Oder: in ihren Rechnungen sei ein Element der Willkür, welches das Wesen ihrer Mathematik von dem der unsern unterscheidet. Und doch würden wir nicht leugnen können daß die Leute eine Mathematik haben. [Ms-124, 117.4 / *BGM VII* §34]

> – There is a contradiction here. But we don't see it and we draw conclusions from it. E.g. we infer mathematical propositions; and wrong ones. But we accept these inferences. – And now if a bridge collapses, which we built on the basis of these calculations, we find some other cause for it, or we call it an Act of God. Now was our calculation wrong; or was it not a calculation?
>
> Certainly, if we are explorers observing the people who do this we shall perhaps say: these people don't calculate at all. Or: there is an element of arbitrariness in their calculations, which distinguishes the nature of their mathematics from ours. And yet we should not be able to deny that these people have a mathematics. [*RFM VII* §34]

Wittgenstein then gives an example of a king's contradictory order, a situation originally described in Ms-130:[19]

> Wer durch seine Regeln zum Widerspruch geleitet wurde, kann sagen: "Ich habe falsche Regeln gegeben". [Aber was sind falsche Regeln?] Es sind Regeln, deren Konsequenzen ich desavouiere.
>
> Der König sagt zum Henker: "Hänge den nicht, der richtig errät, was wir mit unsern Gefangenen tun." Der Gefangene sagt nun zum Henker: "Ich

19 This does not refer, as the footnote of the editors in *RFM VII* §34 suggests, to "the king who made the law that all who came to his city must state their business and be hanged if they lied", although such situation could also be made compatible with the remark.

> werde gehenkt werden." Etc. Er hat den Henker mit dieser Antwort überrascht & verwirrt (confounded). Der Henker weiß nicht, wie er den Königs Befehl ausführen soll. [Ms-130, 81.2–81.3]

> He who has been led to contradiction by his rules can say: "I have given wrong rules". [But what are wrong rules?] They are rules whose consequences I disavow.
>
> The king says to the executioner, "Do not hang him who guesses correctly what we are to do with our prisoners." The prisoner now says to the executioner, "I will be hanged." Etc. He has surprised the executioner with this answer & confounded him. The executioner does not know how to carry out the king's order.

...and in Ms-124:

> Was für Regeln muß der König geben, damit er der unangenehmen Situation von nun an entgeht, in die ihn sein Gefangener gebracht hat? – Was für eine Art Problem ist das? – Es ist doch ähnlich diesem: Wie muß ich die Regeln dieses Spiels abändern, daß die & die Situation nicht eintreten kann. Und das ist eine mathematische Aufgabe.
>
> Aber kann es denn eine mathematische Aufgabe sein, die Mathematik zur Mathematik zu machen?
>
> Kann man sagen: "Nachdem dies mathematische Problem gelöst war, begannen die Menschen eigentlich zu rechnen"? [Ms-124, 118.2–119.3 / *RFM VII* §34]

> What kind of rules must the king give so as to escape henceforward from the awkward position, which his prisoner has put him in? – What sort of problem is this? – It is surely like the following one: how must I change the rules of this game, so that such-and-such a situation cannot occur? And that is a mathematical problem.
>
> But can it be a mathematical problem to make mathematics into mathematics?
>
> Can one say: "After this mathematical problem was solved, human beings began really to calculate"? [*RFM VII* §34 (only for the remarks from Ms-124)]

The prisoner's answer that he will be executed is clearly contradictory: If it is true that he will be executed, he will have guessed correctly what happens to him, but then by the order of the king he cannot be executed. If it is false that he will be executed, his prediction is false, but then he will meet the same fate as all prisoners who guessed incorrectly and will be executed.

These royal orders could be followed flawlessly for a long time if, for whatever reasons, no prisoner ever answers in such a contradictory way. But *if the goal is to prevent this situation from ever happening*, the rules of the game need to be changed, which is then the task of a mathematician and not of a philosopher. However, mathematics cannot justify the need for this change, because the usefulness of the contradictory language game has to be judged on the basis of its role in the form of life of the king and his executioner. The contradiction might then turn out to be utterly disastrous, completely benign, or somewhere in between.

Wittgenstein thus reframes the situation by moving away from the dichotomy of 'mechanical calculation' and 'common sense' and instead emphasising *different degrees of reliability*. Turing's argument gives rise to the belief that a mechanical calculation can only be reliable if it is consistent and that inconsistency immediately renders it completely unreliable. For Wittgenstein, however, a calculation can be reliable, unreliable or somewhere in between, irregardless of whether it is consistent or inconsistent, both in the case of mechanical calculation and also when applying common sense:

> 'Solange die Widerspruchsfreiheit nicht bewiesen ist, kann ich nie ganz sicher sein, daß mir jemand, der gedankenlos, aber gemäß den Regeln, rechnet, nicht irgend etwas Falsches {herausrechnet. // herausrechnen wird.}' So lange also jene Voraussicht nicht gewonnen ist, ist der Kalkül unzuverlässig. – Aber denke, ich fragte: {"*Wie* unzuverlässig?"– // '*Wie* unzuverlässig ist er?' –} Wenn wir von Graden der Unzuverlässigkeit redeten, könnten wir ihr dadurch nicht den metaphysischen Stachel nehmen?
>
> Waren die ersten Regeln des Kalküls nicht gut? Nun, wir gaben sie nur, *weil* sie gut waren. – Wenn sich später ein Widerspruch ergibt, – haben sie *nicht* ihre Pflicht getan? Nicht doch, sie waren für diese Anwendung nicht gegeben worden.
>
> Ich kann meinem Kalkül eine bestimmte Art der Voraussicht geben wollen. Sie macht ihn nicht zu einem *eigentlicheren* Stück Mathematik, aber, etwa{, // –} zu gewissem Zweck {brauchbarer. // brauchbarern.}
>
> Die Idee des Mechanisierens der Mathematik. Die Mode des axiomatischen Systems. [Ms-124, 58.2–59.3 / *BGM VII* §12]

> 'So long as freedom from contradiction has not been proved I can never be quite certain that someone who calculates without thinking, but according to the rules, won't work out something wrong.' Thus so long as this provision has not been obtained the calculus is untrustworthy. – But suppose that I were to ask: "'*How* untrustworthy?" – If we spoke of degrees of untrustworthiness mightn't this help us to take the metaphysical sting out of it?
>
> Were the first rules of the calculus not good? Well, we gave them only '*because* they were good. – If a contradiction results later, – have they '*failed* in their office? No, they were not given for this application.
>
> I may want to supply my calculus with a particular kind of provision. This does not make it into a '*proper* piece of mathematics, but e.g. into one that is more useful for a certain purpose.
>
> The idea of the mechanization of mathematics. The fashion of the axiomatic system. [*RFM VII* §12]

What is philosophically problematic is thus not mechanisation in and of itself, but rather the misleading idea that mechanical calculations commit us to a situation where a calculus is either completely reliable or completely unreliable. In the case of computing machines, it can be tempting to think that the mechanical nature of these machines demands 'stronger' standards of reliability than our ordinary non-mechanical language games, in which we can apply "common sense" when we encounter contradictions. But such a view overlooks that 'ordinary' language games also follow rules that can be described and taught, possibly even to machines, although the effort required to describe certain games rigorously enough for a completely mechanical application is certainly not trivial and sometimes entirely impractical. What separates computing machines from humans, at least concerning the issue at stake here, is not "common sense", but rather the acceptance of *degrees of unreliability* in the case of humans.

This might even hold in the presence of a *hidden contradiction*. As Wittgenstein provocatively asks, if our method of calculating has held up in practice until now, why would the discovery of a previously hidden contradiction sink the whole endeavour and make the entire system worthless? As long as a contradiction is hidden well enough not to matter in practice, the usefulness of a system will not necessarily be impacted by such a discovery. In practice, we might add, the dose makes the poison:

> Wenn der Widerspruch wirklich so gut versteckt ist, {daß wir ihn nicht merken // daß ihn niemand merkt}, warum sollen wir nicht das, was wir jetzt tun, das eigentliche Rechnen nennen?
>
> Wir sagen, der Widerspruch würde den Kalkül *vernichten*. Aber wenn er nun sozusagen in winzigen Dosen aufträte, gleichsam blitzweise, nicht als ein ständiges Rechenmittel, würde er da {das Spiel // den Kalkül} auch vernichten? [Ms-124, 65.4–65.5 / *BGM VII* §15]

> If the contradiction is so well hidden that no one notices it, why shouldn't we call what we do now proper calculation?
>
> We say that the contradiction would '*destroy* the calculus. But suppose it only occurred in tiny doses in lightning flashes as it were, not as a constant instrument of calculation, would it nullify the calculus? [*RFM VII* §15]

Even principles as fundamental as *reductio ad absurdum* or the law of non-contradiction are then up for debate. In his language game, Turing presupposes these rules, which is not surprising in the context of a logical system used to resolve the *Entscheidungsproblem*. But if the goal is to present a formalisation of the notion of mechanical calculation that actually corresponds to "humans who calculate", the 'normal' rules involved in dealing with contradictions cannot simply be taken for granted. As Wittgenstein notes, the "ideal certainty" (Ms-124, 119.4–120.3 / *RFM VII* §35) of a non-contradictory logical system is not an end in itself, but instead dependent on the use of the system. If the system proves useful even in the face of a contradiction, what is the justification for a principle like the law of non-contradiction in such a case?

3.4 USEFUL INCONSISTENCY

Given how Turing has set up the definitions of the concepts used in his argument, a contradiction follows from the assumption of the general decision procedure $\mathscr{D}$. But while a contradiction is usually a sign of an 'error' and thus leads us to reject the 'faulty' assumption, we need to realise that this attitude towards contradictions *stems from the simple, familiar case*, where a 'feedback loop' (of computing machines operating on themselves) is simply not part of the game that is being played. In the more complex case of machines that can operate on their own descriptions and simulate their own behaviour, the contradictions that follow from underdetermined or overdetermined situations are *inherent to this more complex case*. Why then would it follow without any further argument that the only option is to reject the assumption, instead of properly exploring the implications of the contradiction?[20]

Of course it *does follow in classical or intuitionistic logic*, because *reductio ad absurdum* is a valid and even necessary proof technique in a logical system where a contradiction is explosive and any proposition could be inferred from a contradiction, *ex contradictione quodlibet*. The question, then, is not whether Turing's proof is valid in classical or intuitionistic logic (of course it is), but whether such an explosive logic is the right framework to describe the language game that is being played, or rather, whether the language game played by Turing according to the principle of explosion is the only or even the most obvious one that Turing could be playing if the technical terms employed in the proof are to correspond to their ordinary meaning. It is crucial to remember that this correspondence to concepts that have a *use in our non-mathematical activities* lends Turing's proof its importance and standing. We could of course imagine many other logical games that would lead to situations with underdetermined rules such as "Do What You Do", Wittgenstein's variant of the diagonal argument being just one example, so if Turing's proof is to be more than an amusing recreational activity, it must convince us of its applicability to our use of words such as 'decision', 'calculation' and 'computation' outside of the proof. Are these concepts consistent? Or can a philosophical investigation show that some of the language games played with these concepts are inherently inconsistent?

A starting point for such an investigation is the observation that Turing's machine $\mathscr{H}$ *has no use* outside the language game played

20 Even though much of the following argument is influenced by Priest, 2006b and Priest, 2006a, it should not be read as an endorsement of "dialetheism", the notion "that there are true contradictions" (Priest, 2006b, p. 4). The use of contradictions and the application of paraconsistent logics that these situations necessitate are *tools* that might prove 'useful' to clarify our understanding of terms that lie at the heart of computation, but this chapter will deliberately not attempt to extend these observations to theses about the nature of logic or mathematics as a whole. The question will thus not be whether there "are" true contradictions, but merely whether the application of contradictions inside (non-explosive) inconsistent systems can adequately describe and illuminate a certain kind of important language game played at the foundation of the theory of computation, which consistent logical systems would struggle to describe.

by Turing with his application of the diagonal argument. Preventing contradictions is *useful* if the propositions in question stand for contradictory premises that cannot both be true at the same time. But as Wittgenstein notes in Ms-178d, a short fragment most likely dating from before Ms-121 (see Section 1.4), it is hard to see the usefulness of preventing a contradiction at all costs even in the case of propositions such as "436. Der {Satz // Abschnitt} 436 dieses Buches ist nicht beweisbar." (Ms-178d, 2.3), because the use of these propositions is undetermined:

> 'Aus seinem Gegenteil läßt sich ein Widerspruch ableiten.' – Nun, vielleicht macht er hier nichts.
>
> Den Widerspruch zu vermeiden ist eine mathematische Methode. Sie führt zu brauchbaren {Gebilden // Sätzen} & brauchbar ist hier ähnlich unbestimmt wie eine Pointe haben.
> Ist aber die Funktion eines {Satzes // irgendwie satzähnlichen Gebildes} gänzlich unbestimmt, warum soll er nicht ein Widerspruch sein? Warum sollte sich ein Mathematiker prinzipiell vor {*jedem* // dem} Widerspruch bekreuzigen. (Man {könnte // möchte} sagen: hab keine Angst er beißt nicht!) [Ms-178d, 1.2–1.3]

> 'A contradiction can be derived from its opposite.' - Well, maybe it doesn't cause any trouble here.
>
> Avoiding the contradiction is a mathematical method. It leads to usable {constructions // propositions} & usable here is similarly undetermined as having a punch line.
> But if the function of a {proposition // proposition-like construction} is completely undetermined, why should it not be a contradiction? Why should a mathematician in principle cross himself before {*every* // the} contradiction. (One {could // might be inclined to} say: don't be afraid, it doesn't bite!)

To give contradictory decision procedures such as $\mathscr{D}$ a different use than in Turing's argument does not imply that the status of all contradictions suddenly needs to change. To prevent contradictions in mathematics is a *technique*, its general usefulness does not stand or fall with a specific case such as $\mathscr{D}$, as Wittgenstein notes in Ms-127, written in March of 1944:

> "Warum soll es in der Mathematik keinen Widerspruch geben dürfen?" – Nun, warum darf es in unsern einfachen Sprachspielen keinen geben? (Da besteht doch gewiß ein Zusammenhang.) Ist das also ein Grundgesetz, das alle denkbaren Sprachspiele beherrscht?
>
> Angenommen ein Widerspruch in einem Befehl z.B. bewirkt Staunen & Unentschlossenheit – & nun sagen wir: das eben ist der Zweck des Widerspruchs in diesem Sprachspiel.
>
> Es ist *eines* eine mathem. Technik zu gebrauchen, die darin besteht, {den Widerspruch zu vermeiden // dem Widerspruch zu entgehen}, & ein anderes gegen den Widerspruch in der Mathematik überhaupt zu philosophieren. [Ms-127, 80.3–81.3]

> "Why should no contradiction be allowed in mathematics?" - Well, why should none be allowed in our simple language games? (Surely there is a connection.) Is this, then, a basic law that governs all conceivable language games?
>
> Suppose a contradiction in a command, for example, causes astonishment & indecision - & now we say: that is precisely the purpose of the contradiction in this language game.
>
> It is *one thing* to use a mathematical technique that consists in avoiding the contradiction, & another to philosophise against the contradiction in mathematics in general.

If we were to suspend this technique in the case of the computing machine $\mathscr{D}$ and its application as part of $\mathscr{H}$, what use could the contradictory verdict of the machine possibly have? As we have seen, Wittgenstein's variant of the diagonal argument reframes the crucial point of the diagonal as an underdetermined rule, a "Do What You Do" instruction. But why should such a rule not be interpreted as an instruction that leaves the next step up to the person or machine following the rule? A rule that can be followed, by doing whatever we wish to do? The rule would then mean "Do as you like" and could be seen as a "hint from the gods":

> Der Widerspruch. Warum grad dieses eine Gespenst? Das ist doch sehr verdächtig.
>
> Warum sollte eine Rechnung zu einem praktischen Zweck ausgestellt die einen Widerspruch ergibt mir nicht sagen: "Tu wie Dir beliebt, ich die Rechnung entscheide darüber nicht."?
>
> Der Widerspruch könnte als Wink der Götter aufgefaßt werden, daß ich handeln soll & nicht überlegen. [Ms-127, 83.2–83.4]
>
> ---
>
> The contradiction. Why this particular spectre? That is very suspicious.
>
> Why shouldn't a calculation issued for a practical purpose that results in a contradiction be telling me: "Do as you like, I the calculation don't decide about it."?
>
> The contradiction could be taken as a hint from the gods that I should act & not deliberate.

Turing's unwillingness to open the floodgates to inconsistency is understandable, given that foundational issues in (consistent) logic and more specifically the *Entscheidungsproblem* were the impetus for his reflections on computability. But especially in the wake of his negative resolution of said problem, the question might be raised why the systematic procedures executed by computing machines should be held to the same standard of consistency as (classical) logic. If we understand machines from a Wittgensteinian angle, as uneducated, sometimes even 'idiotic' clerks who are taught the rules of particular language games through a form of "training" / "Abrichtung", it becomes hard to see why these machines should not be inconsistent at times and consequently could be reasoned about using inconsistent formal systems. After all, many of the games that could possibly be played might involve inconsistent premises or lead to inconsistent rules, if only out of amusement over the nonsensical rules.

To give up the law of non-contradiction would be to give up a principle that we take for granted, and which has been enshrined as a standard of measurement. Turing needs to presuppose this whole system of measurements as a foundation for his mathematical proof, which then leads us to accept a new proposition as a standard of measurement, namely that there cannot be something that we would call the "general decision procedure" $\mathscr{D}$. However, this is not the only possible modification of our system of measurements in the face of the diagonal argument. We could imagine that people had constructed a general decision procedure which returned a contradictory result for machines such as the paradoxical $\mathscr{H}$. We would reject such a construction, because it clashes with our standard of measurement, but if these people had no standard to compare this result to, they might accept the contradictory result of the machine as the 'correct' result, because they would *define the correct result as the result of the machine*. These people first use the decision procedure as an *experiment*, since they have no other standard that would define the result, but could then adopt the decision procedure as a new standard if the results hold up in practice. When used as an experiment, the practical implementation of their machine would have an influence on what they consider correct: If, for example, a cog jams, the result still is 'correct', because the empirical result is the only criterion for correctness. As soon as they abstract from these empirical factors and use a *rule* as their criterion for correctness, they stop treating the machine as an experiment and adopt a standard of measurement instead. As Wittgenstein emphasises, the standard of measurement is not the *result* of experimental measurements, but it could *follow* from such measurements:

> Man kann auf Grund eines Experiments – oder wie man es sonst nennen will – manchmal die Maßzahl des Gemessenen, manchmal aber auch das geeignete Maß bestimmen.
>
> So ist also die Maßeinheit das Resultat von Messungen? Ja & nein. Nicht das Messungsresultat, aber vielleicht die *Folge* von Messungen. [Ms-124, 95.5–96.1 / *BGM VII* §23]

> An experiment – or whatever one likes to call it – can be what we go on, sometimes in determining the measurement of the thing measured, and sometimes even in determining the appropriate measure.
>
> Then is the unit of measurement in this way the result of measurements? Yes and no. Not the result reached in measuring but perhaps the '*consequence* of measurements. [*RFM VII* §23]

That the distinction between experiment and calculation is not always clear cut is made evident a few pages later, in a remark that was written about three years after the remarks above. However, it should be considered closely related to the remarks from 1941, as it is nearly identical to the pocket notebook remark Ms-163, 27v.3, dating from the same day in 1941 as the remarks quoted above and following them immediately in Ms-163:

> Man könnte sagen: Experiment – Rechnung sind Pole, zwischen welchen sich menschliche Handlungen bewegen. [Ms-124, 114.2 / *BGM VII* §30]
>
> ---
>
> It might be said: experiment – calculation are poles between which human activities move. [*RFM VII* §30]

After including the remark three years later in Ms-124, Wittgenstein then adds another remark about a month later on the shift from empirical proposition to rule, also included in Ms-127 around the same date:

> Jeder Erfahrungssatz kann als Regel dienen wenn man ihn – wie einen Maschinenteil – feststellt, unbeweglich macht, so daß sich nun alle Darstellung um ihn dreht & er zu einem Teil des Koordinatensystems wird & unabhängig von den Tatsachen. [Ms-124, 199.5 / *BGM VII* §74]
>
> Jeder Erfahrungssatz kann als Regel dienen wenn man ihn feststellt, ich meine unbeweglich macht, so daß sich nun alle Darstellung um ihn dreht & er ganz zur Methode der Darstellung gehört & unabhängig von den Tatsachen {wird // ist}. [Ms-127, 224.1]
>
> ---
>
> Every empirical proposition may serve as a rule if it is fixed, like a machine part, made immovable, so that now the whole representation turns around it and it becomes part of the coordinate system, independent of facts. [*RFM VII* §74]
>
> Every empirical proposition may serve as a rule if it is fixed, made immovable, so that now the whole representation turns around it and it belongs entirely to the method of representation, independent of facts.

The above remark shows a remarkable similarity to Wittgenstein's remarks on 'hinge propositions', written much later in 1951 (notably Ms-175, 48r.2, Ms-175, 48v.2 and Ms-177, 5v.2, corresponding to *OC* §341, §343 and §655, respectively). Not only does it shed light on the origin of some of the ideas that were only later fully developed by Wittgenstein, it also underlines what is at stake here: If we give up the assumption that the whole logical system must be consistent, we use certain propositions 'flexibly' that were previously assumed to be 'fixed hinges'. Turing's diagonal argument can then potentially be seen as showing where and why consistency is insufficient and in which area the demand for absolute consistency has to make room for a small dose of inconsistency. Philosophy cannot offer a justification for such a flexible use, as the justification has to come from the use in our form of life, but it can help to show that these hinges are not fixed as the result of discoveries in the world of mathematics that leave no other way out, but rather as the consequence of a conceptual impossibility that could conceivably be cast aside by people with a different form of life.

But even if we were prepared to consider a different language game without *reductio ad absurdum* at the crucial point of the diagonal argument, a question remains: What would the concrete ramifications be in the context of Turing's computing machines, which seem to be a very apt model of computation and certainly comparable to the 'ordinary' computers used by us every day? Let us assume, only for the sake of argument, that we do not immediately reject the assumption based on the contradiction of the diagonal argument, *then what?*

Well, why not relax the requirement of the decision procedure $\mathscr{D}$ and allow an answer of 'circular *and* circle-free' for a case such as $\mathscr{H}$? This does not need to have an impact on any of the 'ordinary' computing machines, in fact the verdict of 'circular and circle-free' will be reserved exclusively for cases such as $\mathscr{H}$, where we can show by the same analysis as in the diagonal argument that the two cases 'circular' and 'circle-free' alone would invariably lead to a contradiction.

But what does it mean for a computing machine to be both circular and circle-free? In contrast to (paraconsistent) logic, where we might be willing to call a paradoxical proposition both true and false, Turing's proof is framed in terms of *numbers*, more specifically binary expansions. A computable number either corresponds to a particular binary expansion or it does not, but there seems to be little room for something that both is and is not circular *at the same time*. There are many ways to extend Turing's own model to allow for such a scenario, for example by considering concurrent 'threads' of execution,[21] but in fact Turing's own description of computing machines already allows for such a possibility:

Turing's machines are able to revisit 'old' squares to the left of the currently scanned symbol and might erase or overwrite them with other symbols. Turing makes ample use of this capability to implement the universal computing machine $\mathscr{U}$, but restricts such a use to the 'temporary' symbols of *E*-squares. The computable sequences that he has in mind thus seem to be restricted to computing machines for which it is not possible to 'rewrite' already written figures after they have been written down. This corresponds to a view of computable sequences and computable numbers as infinite developments of numbers where each execution step increases the 'precision' of the result in the same way a manual calculation of 1/3 in decimal could be said to increase the precision by going from 0.3 to 0.33 to 0.333 etc. *ad infinitum* or when we calculate increasingly more digits of π.

If we allow a machine to revisit and change previously written figures, however, this view breaks down once we consider a computing machine that alternates between printing '0' and '1' (and which would thus correspond to the number 0.010101..., 1/3 as a fraction) but which then always flips all previously written '0' and '1' after printing another decimal place, so that the tape would go from '0' to '10' to '010' to '1010' to '01010' and so forth. (We could alternatively

21 It should be pointed out that these 'extended' models can usually be simulated by Turing's own model, thanks to the computational universality of Turing's computing machines.

implement this behaviour by 'shifting' the tape to the right, copying each symbol one square to the right until there is an empty square at the beginning of the tape, then print a new symbol in the empty square.) What kind of number would this machine compute? It will 'oscillate' between the number '0.01010101...', 1/3 expressed as a fraction, and '0.10101010...', 2/3 in other words. But can the sequence generated by such a machine even be considered a computable number? Turing's definitions at least do not seem to explicitly preclude this possibility:

> A sequence is said to be computable if it can be computed by a circle-free machine. A number is computable if it differs by an integer from the number computed by a circle-free machine. [Turing, 1936, p. 61]

The computing machine is circle-free, since it certainly "writes down more than a finite number of symbols of the first kind" (the figures '0' and '1'). It can therefore be said to describe a computable number, although it should be pointed out that Turing's own implementation of the universal computing machine $\mathscr{U}$ enforces stricter requirements, including the convention never to erase or change a previously written figure to the left of the scanned symbol. However, since according to Turing's own definitions such a behaviour can only be considered a 'convention', we might ask whether a general decision procedure $\mathscr{D}$ is feasible for 'unconventional' computing machines and if so, why Turing's conventional versions are given preference without further argument.[22]

Analogously to the special case of 0.010101..., we can construct such an oscillating variant for *any* 'normal' computable sequence by flipping all the figures between each step of printing 0 or 1. We could then imagine different uses for these oscillating numbers: Perhaps they are used in situations where people do not care what they end up with or want to indicate politeness by being flexible, they might

22 This discrepancy between Turing's definition of computable numbers and his 'implementation' as part of the universal computing machine was noted by Emil Post in Post, 1947, p. 98, (emphasis added): "Primarily as a matter of *practice*, Turing makes his machines satisfy the following *convention*. Starting with the first square, alternate squares are called *F*-squares, the rest, *E*-squares. In its action the machine then never directs motion left when it is scanning the initial square, *never orders the erasure, or change, of a symbol on an* F*-square*, never orders the printing of a symbol on a blank *F*-square if the previous *F*-square is blank, and, in the case of a computing machine, never orders the printing of 0 or 1 on an *E*-square. *This convention is very useful in practice*. However the actual performance, described below, of the universal computing machine, coupled with Turing's proof of the second of the two theorems referred to above, strongly suggests that Turing *makes this convention part of the definition* of an arbitrary machine. We shall distinguish between a *Turing machine* and a *Turing convention-machine*." It is understandable that neither Turing nor Post considered this critique to be more than a minor correction aimed at including the implicit convention used by Turing in a more rigorous definition. Philosophically, though, the passage is very interesting, as it possibly opens the door to language games closely related to Turing's own, but with different conventions, in other words corresponding to Turing machines, but not Turing convention-machines.

for example say: "Give me either a third or two thirds", similar to Wittgenstein's example of the tribe not bothering with coins that fell to the ground (Ms-124, 52.2–53.3 / *RFM VII* §11), or perhaps merely to teach students ways to imagine how something could be said to be in two states at the same time. We might then think of these oscillating numbers as being 'in motion' and distinguish them from 'static', 'motionless' numbers that we usually calculate with. Wittgenstein uses a similar example:

> Aber Du kannst doch einen Widerspruch nicht gelten lassen! – Warum nicht? Wir gebrauchen {ihn // diese Form} ja manchmal in unsrer Rede, freilich selten – aber man könnte sich eine Sprachtechnik denken, in der er ein ständiges Implement {ist // wäre}.
>
> Man könnte z.B. von einem Objekt in Bewegung sagen, es existiere & es existiere nicht an diesem Ort; Veränderung könnte durch den Widerspruch ausgedrückt werden. [Ms-124, 54.4 / *BGM VII* §11]
>
> ---
>
> But you can't allow a contradiction to stand! – Why not? We do sometimes use this form in our talk, of course not often – but one could imagine a technique of language in which it was a regular instrument.
>
> It might for example be said of an object in motion that it existed and did not exist in this place; change might be expressed by means of contradiction. [*RFM VII* §11]

He considers further examples of useful applications that involve contradictions on the following pages of Ms-124 and also returns to the idea more than 60 pages later:

> Ein reflexives Fürwort, das sich auf den Satz in dem es steht bezieht. Gebrauchen wir das Wort "ich" – so daß "Ich bin 5 cm lang" dadurch zu prüfen ist, daß man diesen Satz mißt. Eine solche Form wird meines Wissens nie gebraucht; könnte {aber unter Umständen eine // aber auch eine} wichtige {Rolle spielen. // {Satzform // Form von Sätzen} sein.} Oder: "Ich bestehe aus 5 Wörtern." [Ms-124, 60.2]
>
> [...]
>
> Man könnte sich fragen: Welche Rolle kann ein Satz, wie "Ich lüge immer", im menschlichen Leben spielen? Und da kann man sich Verschiedenes vorstellen. [Ms-124, 126.3 / *BGM VII* §37]
>
> ---
>
> A reflexive pronoun that refers to the sentence in which it is placed. Let us use the word "I" - so that "I am 5 cm long" is to be verified by measuring this sentence. Such a form is never used, as far as I know; it could, however, {play an important role // be a sentence form } in some circumstances. Or: "I consist of 5 words."
>
> [...]
>
> We might ask: What role can a sentence like "I always lie" have in human life? And here we can imagine a variety of things. [*RFM VII* §37]

The same example is used as late as 1947, where a remark on such a usage of the word "ich" appears in Ts-229, 205.6 and Ts-245, 143.7, first written down in this form in Ms-130, 125.5 and clearly influenced by the remark in Ms-124. These examples go further than simply investigating the mathematical "fear" of contradictions, they imagine (more or less) practical uses for contradictions that could then convince us to give up our absolute rejection of contradictions in favour of a more nuanced view that looks at individual language games. This approach is summed up in a later remark in Ms-124, from 1944:

> Das Überraschende, Paradoxe ist paradox nur in einer gewissen, gleichsam mangelhaften, Umgebung. Man muß diese Umgebung so ergänzen, daß, was paradox schien nicht länger so erscheint. [Ms-124, 141.3 / *BGM VII* §43]
>
> ---
>
> Something surprising, a paradox, is a paradox only in a particular, as it were defective, surrounding. One needs to complete this surrounding in such a way that what looked like a paradox no longer seems one. [*RFM VII* §43]

In the case of Turing, it is clear that his convention of never changing a previously written figure is not merely a practical afterthought, but crucial for his reflections on the enumerability of computable sequences and the ability to apply the diagonal argument. After all, the sequence computed by the diagonal can only be said to differ from every sequence in the enumeration if these sequences are 'stable' and do not retroactively change the very figure that was their contribution to the diagonal. However, if the requirement not to change previously written figures does not arise from Turing's definitions but only from his implementation, then perhaps we could imagine an 'unconventional' language game that respected Turing's definitions but also included a decision procedure $\mathscr{D}$ for deciding circularity that we would be willing to call "general"?

For example, in the case of the paradoxical machine $\mathscr{H}$ the decision of $\mathscr{D}$ could be 'oscillating' between the symbols for 'circular' and 'circle-free'. This corresponds quite well to how $\mathscr{H}$ itself 'oscillates' between being one or the other depending on the verdict of $\mathscr{D}$. This shows that there is a decision procedure that we conceivably *could* call a "general decision procedure" for deciding circularity, but which is of course not *the* "general decision procedure $\mathscr{D}$" that Turing's proof works with.

We would then have changed the *technique* of our use of truth functions, which, however, does not mean that the use were somehow *arbitrary*. Rather, we are bound by the logical "must", not because these rules were eternal laws of thought, similar to laws of physics but in the ideal realm of mathematics, but because these are the techniques that we use and their use is embedded in our way of life:

> Das Nicht-Geltenlassen des Widerspruchs charakterisiert die Technik {der // unserer} Verwendung {der // unserer} Wahrheitsfunktionen. {Lassen wir den Widerspruch gelten, so {heißt // bedeutet} das {daß wir die Verwendung der Wahrheitsfunktionen ändern // eine Änderung der Auffassung der Wahrheitsfunktionen}; als faßten wir z.B. eine doppelte Verneinung nicht mehr als Bejahung auf. // Lassen wir den Widerspruch in unsern Sprachspielen gelten, so {bedeutet // ist} das eine Änderung jener Technik. // Lassen wir den Widerspruch in unsern Sprachspielen gelten, so ändern wir jene Technik – so, als gingen wir davon ab, eine doppelte Verneinung als Bejahung anzusehen.} Und diese Änderung wäre von Bedeutung, da die Technik unserer Logik ihrem Charakter nach zusammenhängt mit – – –

> "Die Regeln zwingen mich zu etwas", nun das kann man schon sagen, weil, was mir mit der Regel übereinzustimmen scheint ja nicht von meiner Willkür abhängt. Daher kann es ja geschehen daß ich die Regeln eines Brettspiels ersinne & nachträglich herausfinde daß in diesem Spiel wer anfängt gewinnen *muß*. Und so ähnlich ist es ja, wenn ich finde, daß die Regeln zu einem Widerspruch führen [Ms-124, 104.3–105.2 / *BGM VII* §27]
>
> ---
>
> Not letting a contradiction stand is something that characterises the technique of our employment of our truth-functions. It we do let the contradiction stand in our language-games, we alter that technique – as, if we departed from regarding a double negative as an affirmative. And this alteration would be significant, because the technique of our logic is connected in its character with the conception of the truth-functions.
>
> "The rules compel me to. . . " – this can be said if only for the reason that it is not all a matter of my own will what seems to me to agree with the rule. And that is why it can even happen that I memorize the rules of a board-game and subsequently find out that in this game whoever starts '*must* win. And it is something like this, when I discover that the rules lead to a contradiction. [*RFM VII* §27]

We are then forced to admit that such a game with contradictory rules is not a game ("Ich bin nun gezwungen anzuerkennen, daß das eigentlich kein Spiel ist.", Ms-124, 105.3 / *RFM VII* §27) if our notion of game excludes games that do not give each player the same fair chance of winning the game. But not due to a "spell", in fact such an idea would be exactly the kind of conceptual confusion, the "plain nonsense, and bumps..." that philosophy must try to dispel (PI §119):

> Was zwingt mich denn? – Der Ausdruck der Regel? – Ja; wenn ich einmal so erzogen bin. Aber kann ich sagen, er zwingt mich, ihm zu folgen? Ja; wenn man sich hier die Regel nicht als Linie denkt, der ich nachfahre, sondern als Zauberspruch der uns im Bann hält.
>
> ["schlichter Unsinn, & Beulen . . . "] [Ms-124, 107.3 / *BGM VII* §27]
>
> ---
>
> What is it that compels me? – the expression of the rule? – Yes, once I have been educated in this way. But can I say it compels me to follow it? Yes: if here one thinks of the rule, not as a line that I trace, but rather as a spell that holds us in thrall.
>
> (("plain nonsense, and bumps. . . ")) [*RFM VII* §27]

In the case of Turing's paradoxical machine $\mathscr{H}$, such a "spell" leads us to believe that our usual standard of measurement, the demand that there be no contradictions in our logical system, *must* apply even to machines such as $\mathscr{H}$ and as a result we give up the concept of a "general decision procedure" for circularity. But Turing's argument obscures that his formalism goes far beyond 'logical' computing machines and allows for the construction of paradoxical machines like $\mathscr{H}$. Why then should the standard of measurement that originated in our use of *logically consistent and non-paradoxical* calculations apply to machines such as $\mathscr{H}$? Why not revise the standard of measurement to accommodate machines that can operate on their own encoded representation, which can of course lead to paradoxical results?

Turing shows that computation encompasses more than what is usually considered 'logical' or even 'mathematical' and his diagonal argument proves that we will run into a contradiction if we assume that computation could be reduced to these more restrictive language games. But Turing then excludes this contradiction from his conceptual language game, instead of revising the language game to find a role for contradictory decision procedures. What is missing is a "serious investiture" / "ernsthafte Einkleidung" [23] of the logical paradox:

> Ich bestimme ein Spiel & sage: "Machst Du diese Art Zug, so ziehe ich *so*, machst Du jene, so ziehe ich *so*. – Jetzt spiele!" Und nun macht er einen Zug, oder etwas, was ich auch als Zug anerkennen muß, & wenn ich nach meinen Regeln {weiterspielen // daraufhin ziehen} will, so erweist sich, was immer ich tue, {als unrichtig // als {den // meinen} Regeln nicht gemäß}. Wie konnte das geschehen? Als ich Regeln aufstellte, da *sagte* ich etwas. Ich folgte einem gewissen Brauch. Ich sah nicht voraus, was wir weiter tun würden, oder sah nur eine bestimmte Möglichkeit. Es war nicht anders als hätte ich Einem gesagt: Gib das Spiel auf; mit diesen Figuren kannst Du nicht mattsetzen" & hätte eine bestehende Möglichkeit des Mattsetzens übersehen.
>
> Die verschiedenen, halb scherzhaften, Einkleidungen des logischen Paradoxes sind nur in sofern interessant als sie einen daran erinnern, daß eine ernsthafte Einkleidung des Paradoxes von Nöten ist, um seine Funktion eigentlich zu verstehen. Es fragt sich: Welche Rolle kann ein solcher 'logischer Irrtum' in {einem Sprachspiel // einer Sprachanwendung} spielen? [Ms-124, 109.4–110.2 / *BGM VII* §29]
>
> ---
>
> I am defining a game and I say: "If you move like this, then I move like '*this*, and if you do that, then I do '*this*. – Now play." And now he makes a move, or something that I have to accept as a move and when I want to reply according to my rules, whatever I do proves to conflict with the rules. How can this have come about? When I set the rules up, I '*said* something: I was following a certain use. I did not foresee what we should go on to do, or I saw only a particular possibility. It was just as if I had said to somebody: "Give up the game; you can't mate with these pieces" and had overlooked an existing possibility of mating.
>
> The various half joking guises of logical paradox are only of interest in so far as they remind anyone of the fact that a serious form of the paradox is indispensable if we are to understand its function properly. The question is: what part can such a logical mistake play in a language-game? [*RFM VII* §29]

The reason Turing can simply exclude the contradiction is that machines such as $\mathscr{H}$ operate in a 'vacuum', they (or rather the contradictory general decision procedure machine $\mathscr{D}$ that they depend on) can be excluded because there is no use for them. But exactly *because* there is no use for them (yet), we could revise the language game so that the contradiction produced by these machines need not bother us and thus *give them a use*. This is what distinguishes the contradiction in Turing's diagonal argument from a contradiction such as "$2+2=4$ and $\neg(2+2=4)$", which would immediately clash

23 Translated by Anscombe simply as "serious form".

with a multitude of our techniques of calculation. We are led to believe that opening the door to the contradictory decision procedure $\mathscr{D}$ would open the floodgates to all sorts of contradictions and to the end of logic and mathematics. But machine $\mathscr{H}$ is uncharted territory, its use is restricted to Turing's diagonal argument, which is exactly what gives us the conceptual freedom to investigate other uses philosophically, with the aim of pointing out the difference in use between machines such as $\mathscr{H}$ and 'ordinary' computing machines.

3.5 machines as mathematicians

Let us come back to the 'limits' established by Turing's argument and their consequences for what we, as human calculators, can and cannot do once we commit ourselves to a stepwise execution of instructions corresponding to the capabilities of Turing's universal computing machine.[24] To avoid the possibility of getting trapped too early in questions relating to the philosophy of mind, we will only indirectly consider computing machines and instead distinguish between two human professions: 'clerks', who are exactly what Wittgenstein calls a "human calculating machine" and who we can picture as being "completely idiotic" (Ms-126, 33.4)[25], and 'mathematicians', who may use insight to go beyond a purely mechanistic and 'idiotic' sequence of instructions. The question to be investigated will then be this: Could we possibly teach clerks how to act like mathematicians, through the sheer use of instructions that are to be followed in an 'idiotic' way? Or is there an unsurmountable boundary between clerks and mathematicians that can only be crossed by employing 'non-idiotic' insight which cannot be codified as stepwise instructions?

There are certainly many ordinary mathematical tasks carried out by mathematicians which can be delegated to clerks and Turing says as much in his lecture on the "Automatic Computing Engine", describing one of the first attempts to build a working computing machine as a practical embodiment of the principles underlying Turing's theoretical universal computing machine:

24 In the following quotations, Turing usually refers to the negative resolution of the *Entscheidungsproblem* when talking about the limits of machines. This is not equivalent to his diagonal argument, but heavily depends on it. A description of how Turing's diagonal argument leads to the negative resolution of the *Entscheidungsproblem* would go beyond the scope of this text, but it should be mentioned that the limits demonstrated by Turing are very similar to Gödel's incompleteness results, which are discussed in more detail in Section 2.1. Turing's negative resolution of the *Entscheidungsproblem* manages to adapt Gödel's results to the conceptual framework of computing machines.

25 The connection between the remark in Ms-126 and the discussion of Turing's diagonal argument is underscored by the fact that the only other mention of such an 'idiotic' calculation process occurs in the context of Wittgenstein's main treatment of a diagonal argument, his discussion of Cantor's proof in Ms-117, 104.3.

> As regards mathematical philosophy, since the machines will be doing more and more mathematics themselves, the centre of gravity of the human interest will be driven further and further into philosophical questions of what can in principle be done etc. [Turing, 1947, p. 392]

What is it that cannot "in principle be done" by clerks operating like machines? The most important difference between clerks and mathematicians in that regard seems to be their handling of decision procedures in logical systems:

> It might be argued that there is a fundamental contradiction in the idea of a machine with intelligence. [. . .] It has for instance been shown that with certain logical systems there can be no machine which will distinguish provable formulae of the system from unprovable, i.e. that there is no test that the machine can apply which will divide propositions with certainty into these two classes. Thus if a machine is made for this purpose it must in some cases *fail to give an answer*. On the other hand if a mathematician is confronted with such a problem he would search around a[nd] find new methods of proof, so that he ought *eventually to be able to reach a decision about any given formula*. This would be the argument. [Turing, 1947, pp. 393–94, emphasis added]

The same sentiment is echoed in Turing's article "Intelligent Machinery":

> Recently the theorem of Gödel and related results [...] have shown that if one tries to use machines for such purposes as determining the truth or falsity of mathematical theorems and one is not willing to tolerate an *occasional wrong result*, then any given machine will in some cases be unable to give an answer at all. On the other hand the human intelligence seems to be able to find methods of ever increasing power for dealing with such problems 'transcending' the methods available to machines. [Turing, 1948, pp. 410–11, emphasis added]

The most detailed description, however, is found in Turing's "Computing Machinery and Intelligence":

> There are a number of results of mathematical logic which can be used to show that there are limitations to the powers of discrete-state machines. [. . .] The result in question [of Turing, 1936] refers to a type of machine which is essentially a digital computer with an infinite capacity. It states that there are certain things that such a machine cannot do. If it is rigged up to give answers to questions as in the imitation game, there will be some questions to which it will either give a *wrong* answer, or fail to give an answer at all however much time is allowed for a reply. There may, of course, be many such questions, and questions which cannot be answered by one machine may be satisfactorily answered by another. [. . .] The questions that we know the machines must fail on are of this type, "Consider the machine specified as follows . . . Will this machine ever answer 'Yes' to any question?" The dots are to be replaced by a description of some machine in a standard form, which could be something like that used in §5. When the machine described bears a certain comparatively simple relation to the machine which is under interrogation, it can be shown that the answer is *either wrong or not forthcoming*. This is the mathematical result: it is argued that it proves a disability of machines to which the human intellect is not subject. [Turing, 1950, pp. 450–51, emphasis added]

According to this argument, negative resolutions of the *Entscheidungsproblem*, such as Turing's proof on the basis of the application of the diagonal argument, show that a calculation by machines or clerks acting like them must "fail to give an answer" at the crucial point of the diagonal. What is meant by the word "fail" is left open, presumably this is exactly the 'getting stuck in a loop' situation that Turing's diagonal argument inevitably leads to. The important point is that it appears that mathematicians will be able to give a meaningful answer, not just for specific formulae, but for "any" formula: They reach a "decision" (a yes/no answer for provable/unprovable or circular/circle-free, since the latter forms the basis for Turing's negative resolution of the *Entscheidungsproblem*), and they reach it "eventually", meaning in a finite number of steps.

This is a tall order! If mathematicians are, as Turing says, able to solve these questions under the same conditions available to machines (clearly defined decision procedure, general for any formula, finite number of steps) and we accept Turing's argument in "On Computable Numbers", then surely we must admit that mathematicians can complete practical tasks that clerks simply cannot hope to ever complete? And would this then not be a physical impossibility, a discovery about our empirical world and the limits of mechanical rule-following, an argument in favour of our eternal superiority over machines, at least in the area of mathematics?[26]

Such a conclusion would be premature, because, as Turing remarks himself, we have overlooked an important difference in the rules of the game between clerks and mathematicians. We first need to level the playing field:

> Against it I would say that fair play must be given to the machine. Instead of it sometimes giving no answer we could arrange that it gives *occasional wrong answers*. But the human mathematician would likewise make *blunders* when trying out new techniques. It is easy for us to regard these *blunders as not counting* and give him another chance, but the machine would probably be allowed no mercy. In other words then, if a machine is expected to be *infallible*, it cannot also be intelligent. [Turing, 1947, p. 394, emphasis added]

Similarly, in "Intelligent Machinery" Turing writes:

> "The argument from Gödel's and other theorems [. . .] rests essentially on the condition that the machine must not make *mistakes*. But this is not a requirement for intelligence. [Turing, 1948, p. 411, emphasis added]

26 See e.g. Lucas, 1961, p. 116: "The Gödelian formula is the Achilles' heel of the cybernetical machine. And therefore we cannot hope ever to produce a machine that will be able to do all that a mind can do: we can never, not even in principle, have a mechanical model of the mind." The paper also contains argumentative gems such as the following (p. 120): "If a machine were wired to correspond to an inconsistent system, then there would be no well-formed formula which it could not produce as true; and so in no way could it be proved to be inferior to a human being. Nor could we make its inconsistency a reproach to it – are not men inconsistent too? Certainly women are, and politicians; and even male non-politicians contradict themselves sometimes, and a single inconsistency is enough to make a system inconsistent." Leaving aside the very questionable second part of the argument here, Lucas tacitly equates inconsistency with logical trivialism, as pointed out in Priest, 2006b, p. 42 (footnote). Other arguments similar to the one by Lucas can be found in Penrose, 1991, pp. 416–18 and Penrose, 1994.

And in 'Computing Machinery and Intelligence', Turing reframes the situation in terms of multiple machines:

> The short answer to this argument is that although it is established that there are limitations to the powers of any particular machine, it has only been stated, without any sort of proof, that no such limitations apply to the human intellect. But I do not think this view can be dismissed quite so lightly. Whenever one of these machines is asked the appropriate critical question, and gives a definite answer, *we know that this answer must be wrong*, and this gives us a certain feeling of superiority. Is this feeling illusory? It is no doubt quite genuine, but I do not think too much importance should be attached to it. We too often give wrong answers to questions ourselves to be justified in being very pleased at such evidence of fallibility on the part of the machines. Further, our superiority can only be felt on such an occasion in relation to the one machine over which we have scored our petty triumph. *There would be no question of triumphing simultaneously over all machines. In short, then, there might be men cleverer than any given machine, but then again there might be other machines cleverer again, and so on.*" [Turing, 1950, pp. 450–51, emphasis added]

That mathematicians sometimes make mistakes is undeniably a fact of life in their profession, but it does not lead us to immediately discard their results or question mathematics as a whole. We thus need to extend the same courtesy to our mechanical clerks if we want to treat both sides equally. This reminder by Turing seems innocuous enough, but it obscures an implicit assumption, because the mistakes that clerks might make appear to be of a different sort than the mistakes made by mathematicians: Clerks have to give "wrong answers" when calculating answers to a decision procedure, because the alternative would be to "fail to give an answer", but importantly *there is a correct answer*, and mathematicians will usually be able to find it, unless they make "blunders". This wording, "wrong answers" in comparison to "blunders", makes it seem as if the mistakes committed by the clerks were more serious than the blunders of the mathematicians, since the former are the results of theoretical limits that hold even for idealised machines under perfect conditions, whereas the latter are the results of practical realities, an unfortunate side effect of a mathematician's biological hardware (or rather 'wetware'), similar to a grain of sand causing a disturbance in a complex system of cogs, or an ink blot accidentally changing a digit written down on paper. Turing seems to rely on this asymmetry between "wrong answers" and (harmless) "blunders" to establish who has the last word: Otherwise, how do we know that the (supposedly wrong) answer given by the machine is inferior to the proof by the mathematician using a new technique, if the two methods are in disagreement?[27]

---

27 Note that the same distinction between serious and minor mistakes is equally crucial for Lucas' argument that the mind cannot be modelled by machines operating mechanically, Lucas, 1961, p. 120: "The fact that we are all sometimes inconsistent cannot be gainsaid, but from this it does not follow that we are tantamount to inconsistent systems. Our inconsistencies are mistakes rather than set policies. They correspond to the occasional malfunctioning of a machine, not its normal scheme of

It is important to note here that blunders occur only rarely and that mathematicians often trust new techniques, though in some cases perhaps only after some cross-checks were made by the wider mathematical community, and that even if blunders do happen, these incidents are not disastrous, but can be remedied.[28] A fitting example is Turing, 1936, which contains a number of such blunders, subsequently corrected by Turing and others.[29]

But while these blunders by mathematicians are treated very nonchalantly, the discovery of an inconsistency in a formal system is treated as a much more serious matter, even if it had not led to problems before. The reason for this is of course the missing oversight in the case of mechanical clerks, who would happily deduce anything from an inconsistent system according to their calculation rules, whereas mathematicians have the ability to distinguish between serious and incidental mistakes and may even gloss over the latter kind. But what leads us to assume that this distinction could not be formalised and taught to the mechanical clerks? It is of course not obvious how this would be done and could turn out to be entirely impractical, but for the current discussion of "what can be done in principle" it is only fair to set these purely practical matters aside, in the same way that Turing's computing machines is allowed to assume an infinite tape.

At this point, it becomes clear that Turing's mention of "wrong answers" applies a double standard. At first glance, it might appear as if mathematicians were not bound by the limitations that apply to mechanical clerks and computing machines. But mathematicians are unimpeded by these limitations only in so far as they are allowed to play a *different language game* than the mechanical clerks. In a way, this is a trivial observation, since we are merely pointing out that mathematicians, if they follow the same calculation rules as mechanical clerks and act as a computing machine, will also run into the same troubling situations as a computing machine following these same rules. The only way to decide that mechanical clerks have made a

operations. Witness to this that we eschew inconsistencies when we recognize them for what they are. If we really were inconsistent machines, we should remain content with our inconsistencies, and would happily affirm both halves of a contradiction. Moreover, we would be prepared to say absolutely anything – which we are not. It is easily shown that in an inconsistent formal system everything is provable, and the requirement of consistency turns out to be just that not everything can be proved in it – it is not the case that "anything goes"." Lucas conflates inconsistency with logical trivialism. If his notion of consistency requires only that "not everything can be proved", many paraconsistent systems of logic would fit the bill.

28 Priest, 2006b, p. 40 (footnote), notes how rarely there is a dispute among mathematicians about what constitutes a valid proof and that "[a]s Wittgenstein stressed, without consensus the whole "language game" of proof would break down." To this we might add that without this consensus the distinction between serious mistake and rectifiable blunder would break down as well.

29 See Turing and Copeland, 2004, pp. 91–93 for a discussion of different corrections, both by Turing himself and others.

"mistake", given a "wrong answer", is in another language game that is external to the process of calculating the particular answer we are interested in at that particular moment.[30]

This double standard is made evident once we ask what the correct answer would be. To speak of a "mistake" usually implies that there is a correct alternative, otherwise the usage of the word "mistake" must be considered quite unorthodox and should be explicitly called out. What, then, would the correct answer be in the case of the calculation performed by a mechanical clerk tasked with providing a general decision procedure for the halting problem? The clerk is faced with a binary choice and can either answer with 'yes' or 'no', '0' or '1', but we know from the start that we will accept none of these options as the correct answer. However, we also do not give the clerk the chance to answer in any other way. Seen from this perspective, the verdict ("wrong answer") is already reached before the trial even begins.

To call such an answer a "mistake" makes sense *only in a different language game*, where we *can* give a correct answer. The answer in this different language game is not 'yes' or 'no', '0' or '1', however, but rather the realisation that *no answer can be given in the more restricted language game* played by the mechanical clerks.

This distinction between different ways of answering what might at first appear to be the same question in the same language game is important. The problem with Turing's proof and its subsequent interpretations is not that the proof in any way contradicts this way of stating the situation, in fact, the proof is perfectly fine as long as we restrict our interpretation to the purely mathematical results of it. Rather, the proof *together with its ordinary language interpretations* has the tendency to obscure the differences of these two language games, since we are led to use the concepts of a decision procedure, of questions and answers in multiple conflicting ways, but fail to see the differences in their usages, simply because the same word is used in every case.

Seen in this light, Turing's reflections on the superiority of humans over machines (or vice versa) need to be reexamined carefully. The passage on the fallibility of machines from 'Computing Machinery and Intelligence', quoted in full above, posits this superiority in terms of being "cleverer" than a given human or machine:

> We too often give wrong answers to questions ourselves to be justified in being very pleased at such evidence of fallibility on the part of the machines. Further, our superiority can only be felt on such an occasion in relation to the one machine over which we have scored our petty triumph. *There would be no question of triumphing simultaneously over all machines. In short, then, there might be men cleverer than any given machine, but then again there might be*

30 Lucas, 1961, pp. 113, 117, 120, repeatedly states that "we [humans], standing outside the system, can see it [the undecidable proposition] to be true", but rejects the possibility of a mechanical and inconsistent system playing the same game as an adequate model for the human mind.

other machines cleverer again, and so on." [Turing, 1950, pp. 450–51, emphasis added]

This word choice appears rather peculiar if we keep in mind that different language games are being played in these situations. Being "cleverer" than a machine means only that the direct 'yes'/'no' answer expected from the machine is sidestepped in favour of an answer in a different and indirect language game. But such an avenue is in principle open to "cleverer" machines as well, if we consider the behaviour of entering an indirect language game to be characteristic of this cleverness described by Turing. Of course human mathematicians will sometimes make mistakes, but unless there are convincing reasons to believe that these mistakes are somehow necessary for playing the more indirect language game, these mistakes can be ignored in the same way that the errors inherent in any practical implementation of computers are ignored in theoretical arguments. In contrast to these issues of practical machines, the fundamental undecidability results in the wake of Gödel and Turing seem to be concerned with more fundamental mistakes: They give rise to the idea that machines are *inherently limited* compared to humans because these limitations concern *fundamental* mistakes that *must* occur in every machine, no matter how much we try to reduce any accidental hardware glitches. This is why such a result appears to be an important discovery in the ideal realm of mathematics, as it shows us the limitations of a machine even in a perfect operating environment without friction or other physical factors.[31]

That machines make logical mistakes as shown by Turing in "On Computable Numbers", even in an idealised operating environment without friction, grains of sand or other such disturbances, does not somehow 'balance out' a mathematician's human inability to stay perfectly focused all the time and avoid blunders in calculations that they 'actually' know the answer to. These two cases are fundamentally different in so far as only the latter case is a "calculation" in Wittgenstein's sense, where the correct result of the calculation is defined by some "timeless" measure independent of the particular result of calculation at the given point in time. This means most notably that in the case of "blunders", when the mathematician is distracted or a

31 In reply to Turing's remark on the "petty triumph", Lucas, 1961, p. 118 contests that "this is irrelevant. What is at issue is not the unequal contest between one mind and all machines, but whether there could be any, single, machine that could do all a mind can do. For the mechanist thesis to hold water, it must be possible, in principle, to produce a model, a single model, which can do everything the mind can do. [...] To succeed, [the mechanist] must be able to produce some finite mechanical model of the mind – any one he likes, but one he can specify, and will stick to. But since he cannot, in principle cannot, produce any mechanical model that is adequate, even though the point of failure is a minor one, he is bound to fail, and mechanism must be false." As an objection to Turing, this falls flat, because Lucas fails to argue convincingly why an inconsistent (but not logically trivial) system must in principle be unable to circumvent these limits.

grain of sand jams the cogs of their mechanical counterpart, the result, changed by this blunder, does not become the *timeless result of the calculation*, but is instead declared "wrong", an exception to the rule. In the case of the machine calculating an answer and then making a "mistake", however, the machine's calculation should more appropriately be called an "experiment", since no correct answer is defined by some measure *inside the system of the machine* that could be considered the "timeless" result of the calculation. The measure of what constitutes a correct or incorrect answer is only supplied by the indirect and external language game observing the calculation and *only then can the experimental calculation be called a mistake*.

3.6 blunders and new techniques

Let us revisit the point that a "human mathematician would likewise make blunders when trying out new techniques", a sentiment echoed in "Intelligent Machinery, A Heretical Theory":

> By Gödel's famous theorem, or some similar argument, one can show that however the machine is constructed there are bound to be cases where the machine fails to give an answer, but a mathematician would be able to. On the other hand, the machine has certain advantages over the mathematician. Whatever it does can be relied upon, assuming no mechanical 'breakdown', whereas the mathematician makes a certain proportion of mistakes. *I believe that this danger of the mathematician making mistakes is an unavoidable corollary of his power of sometimes hitting upon an entirely new method*. This seems to be confirmed by the well known fact that the most reliable people will not usually hit upon really new methods." [Turing, 1951, p. 472, emphasis added]

It is notable that Turing mentions these blunders mainly in the context of "trying out new techniques". In his view, the work of mathematicians can be separated into two different activities: On the one hand they work with established techniques, where they presumably do not make any mistakes after having learned these techniques thoroughly. But on the other hand they will sometimes leave the confines of established techniques to try out new techniques. It is safe to assume that the reason for this is not just boredom on the part of the mathematical practitioner, but the need to use previously unknown or unexplored techniques to answer questions which *could not be answered through the use of established techniques alone*. The work of the mathematician thus includes an element of creative exploration and the set of techniques is not set in stone, but rather constantly evolving. To imitate such a behaviour would then require some form of training, as Turing points out in his lecture on the Automatic Computing Engine, again in the context of the theoretical limits of machines as implied by his negative resolution of the *Entscheidungsproblem*:

> There are several mathematical theorems which say almost exactly that. But these theorems say nothing about how much intelligence may be displayed if a machine makes no pretence at infallibility. To continue my plea for 'fair

> play for the machines' when testing their I.Q. A human mathematician has always undergone an extensive *training*. This training may be regarded as not unlike putting instruction tables into a machine. One must therefore not expect a machine to do a very great deal of *building up of instruction tables on its own*. No man adds very much to the body of knowledge, why should we expect more of a machine? Putting the same point differently, the machine must be allowed to have *contact with human beings in order that it may adapt itself to their standards*. The game of chess may perhaps be rather suitable for this purpose, as the moves of the machine's opponent will automatically provide this contact. [Turing, 1947, p. 394, emphasis added]

Why would it be necessary for machines to build up instruction tables on their own? If this is not done simply for reasons of efficiency or 'programmer convenience', where a machine 'bootstraps' itself and generates a more complex instruction table on the basis of a simpler 'meta-table' (similar to Turing's own use of subsidiary "skeleton tables" in his definition of the universal computing machine), then the reason for such a learning process, at least when considered on a purely theoretical level with no regard for practical efficiency, can only be that the training of the machine introduces an element that *could not have been known and formalised at the time of the initial creation of the program*, because the requirements are evolving. If this is what is meant, and it is hard to see what other explanation could apply here in the context of discussing the theoretical limits of a machine implied by "mathematical theorems", the choice of chess as an example is rather strange, exactly because the positions in chess *could* in theory be completely enumerated and the 'only' reason why learning is required at all is the impracticality of enumerating these positions. Turing's example makes more sense if the goal is to imitate human styles of play, but there is no theoretical reason to include this human and evolving component if the ultimate goal is to create a theoretically optimal chess playing machine.[32]

If training is necessary even in the idealised and theoretical case, then the elements that could not be formalised during the creation of the machine's instruction table must be instructions that could not have been derived by the machine on its own. For example, if the machine's task is to prove mathematical propositions, these new elements could be new axioms or new inference rules, but it would be superfluous to train the machine with propositions that can already be inferred from the axioms and inference rules in its 'untrained' state. This implies that there can be no justification for the inclusion of such a new axiom or inference rule *based on the logical system of the machine*, the machine must simply accept this new way of acting as a result

32 Turing himself mentions this rather obvious fact, Turing, 1953, p. 570: "If the machine could calculate at an infinite speed, and also had unlimited storage capacity, a comparatively simple rule would suffice, and would give a result that in a sense could not be improved on. [...] Such a rule is practically applicable in the game of noughts and crosses, but in chess is of merely academic interest."

of its training, it will thus be a form of "training" / "Abrichtung", to borrow Wittgenstein's wording.[33]

This is further elucidated in Turing's PhD dissertation, "Systems of Logic Based on Ordinals", and is closely related to Turing's distinction between "intuition" and "ingenuity" in mathematics:

> Mathematical reasoning may be regarded rather schematically as the exercise of a combination of two faculties, which we may call *intuition* and *ingenuity*. The activity of the intuition consists in making spontaneous judgments which are not the result of conscious trains of reasoning. These judgments are often but by no means invariably correct (leaving aside the question what is meant by "correct"). Often it is possible to find some other way of verifying the correctness of an intuitive judgment. We may, for instance, judge that all positive integers are uniquely factorizable into primes; a detailed mathematical argument leads to the same result. This argument will also involve intuitive judgments, but they will be less open to criticism than the original judgment about factorization. I shall not attempt to explain this idea of "intuition" any more explicitly.
>
> The exercise of ingenuity in mathematics consists in aiding the intuition through suitable arrangements of propositions, and perhaps geometrical figures or drawings. It is intended that when these are really well arranged the validity of the intuitive steps which are required cannot seriously be doubted. [Turing, 1938, p. 192, emphasis by Turing]

The hope of Hilbert's program was to replace informal intuitive judgements through the use of formal and unambiguous logical systems, to such a degree that "in pre-Gödel times it was thought by some that it would probably be possible to carry this programme to such a point that all the intuitive judgments of mathematics could be replaced by a finite number of these rules" (Turing, 1938, pp. 192–93), but of course Gödel and Turing himself demonstrated the limitations of such an approach. No matter how a (consistent) formal system (that is sufficiently expressive to formalise basic arithmetic) is defined, it will always be necessary to add 'external' intuitive judgements if the goal is to prove all results that can be proved by mathematicians, which in the case of both mathematicians and machines will require the "extensive training" to learn new techniques.

But "ingenuity", the "suitable arrangements of propositions" necessary to properly move from (intuitive) step to (intuitive) step on

33 However, it is relevant in this context that the trained axioms or inference rules are not arbitrary and that there usually exists a consensus in the mathematical community about what constitutes an intuitively correct assertion. Priest, 2006b, p. 40 (footnote, already partly quoted above), points out the importance of this consensus for the "notion of mathematical proof" and emphasises the Wittgensteinian aspect of this observation:

> There may, from time, to time, be disputes over whether a mathematical assertion is a basic statement, and, more, generally, over canons of proof. Still, perhaps the amazing thing about mathematics (in virtue of its non-empirical nature) is the unanimity of the mathematical community at any one time about what constitutes a legitimate proof. (Witness the fact that with very few exceptions intuitionism has made hardly any inroads into mathematics departments of universities.) [. . .] One might even say that this consensus is necessary for there to be a notion of mathematical proof at all. As Wittgenstein stressed, without consensus the whole "language game" of proof would break down.

the way to the mathematical result, can in theory be completely eliminated through the use of a mechanical process, in the way already outlined above in the context of chess:

> We are always able to obtain from the rules of a formal logic a method of enumerating the propositions proved by its means. We then imagine that all proofs take the form of a search through this enumeration for the theorem for which a proof is desired. In this way ingenuity is replaced by patience. [Turing, 1938, p. 192]

The need for these intuitively correct extensions on the way to more logical completeness is Turing's motivation for his logical systems based on ordinals, an infinite hierarchy of increasingly more complete logical systems, where each step in the hierarchy is a new logical system including a new intuitive judgement that allows the system to prove a result unprovable by the weaker systems. The logical details are outside the scope of the current discussion, the relevant point is that each such logical system corresponds directly to a computing machine, so that "by choosing a suitable machine one can approximate 'truth' by 'provability' better than with a less suitable machine, and can in a sense approximate it as well as you please." (Turing, 1940, p. 215) This further explains Turing's view, already quoted above, that "there might be men cleverer than any given machine, but then again there might be other machines cleverer again, and so on."

At this point, the situation as portrayed by Turing might seem to be agreeably compatible with many views held by Wittgenstein. Post-Gödel, and especially post-Turing, the concept of mathematics as a single formal system, amenable to calculation by a single powerful and universal computing machine, appears to give way to a heterogeneous activity carried out in a multitude of logical systems, where no single logical system could possibly subsume all the others. In the vocabulary of Wittgenstein, mathematics turns out to be a multitude of language games, often overlapping and sharing certain family resemblances, but without an essential defining characteristic that could be stated as a formalised logical system, however complex this formalisation may be.

Such an emphasis of the commonalities between Turing and Wittgenstein is certainly appealing, but it runs the risk of obscuring conceptual differences not just between these two authors, but more importantly between related notions in Turing's own comparison of computing machines with mathematicians:

> [I]t has been shown that there are machines theoretically possible which will do *something very close to thinking*. They will, for instance, test the validity of a formal proof in the system of Principia Mathematica, or even tell of a formula of that system whether it is provable or disprovable. *In the case that the formula is neither provable nor disprovable* such a machine certainly does not behave in a very satisfactory manner, for it continues to work indefinitely without producing any result at all, *but this cannot be regarded as very different from the reaction of the mathematicians*, who have for instance

> worked for hundreds of years on the question as to *whether Fermat's last theorem is true or not*. [Turing, 1951, p. 472, emphasis added]

It is interesting that in Turing's view, these two situations "cannot be regarded as very different", as there is a rather obvious and quite fundamental difference: In the case of Fermat's last theorem, mathematicians *did not know whether the theorem was true or not*, it still needed to be proved and it was not clear how to arrive at a proof, whereas in the case of the impossibilities shown by Turing for computing machines, *we already know that no answer given by the particular machine will be considered correct, ever*. If the search of a proof of Fermat's theorem is to be compared to the mechanistic calculation of a clerk, then the better picture would be a very complicated and long-running calculation, for which it was for a long time unknown what result the calculation would arrive at or if the calculation could even arrive at a result of "true" or "false" at all without adding new instruction tables and thus teaching the clerk how to apply new techniques. Such a situation is different from the one presented by Turing's diagonal argument, however, because Turing has given us a schema that can be applied to any universal computing machine and which always ensures that the machine will fail to give an answer. The fact that a computing machine "continues to work indefinitely without producing any result at all" is thus different from not knowing what the result will be in so far as we know that *there will never be a result*. In other words, the difference is one between *not knowing what to expect* (in the case of *unsolved* mathematical problems) and *knowing not to expect anything* (in the case of *unsolvable* mathematical problems). In contrast to Fermat's last theorem, where the result was unknown for a long time, the fallibility of computing machines stems from results that are *tautological and underdetermined* (or in 'inverse' cases such as the halting problem, *contradictory and overdetermined*). Furthermore, it is easy to imagine that new techniques might be required to solve Fermat's last theorem, but it is not obvious why new techniques would necessarily be involved in the case of underdetermined or overdetermined calculations or could help in such a case at all. In so far as these cases follow a schema and can be applied to any particular computing machine through the application of a diagonal argument, the technique is exactly what Turing's proof already provides and is itself applicable to any computing machine, which will then lead to the definite result that the machine must fail to give a correct answer, no matter which new technique it employs.[34]

34 That Turing is far from the only one to equate the undecidability of paradoxical machines with undecided mathematical problems can be seen in Hopcroft, Motwani, and Ullman, 2001, pp. 308–313, where Fermat's last theorem is presented as the intuitive introductory example that then motivates an undecidability proof using the halting problem as a more rigorous treatment of the intuitive notion. (See also the connection with Chaitin's Ω explored in Appendix A.)

It might appear to be of little importance whether or not the logical limitations of computing machines can be compared to the mathematical community's difficulty of finding a proof of Fermat's last theorem. However, a crucial differentiating factor not mentioned by Turing is the mathematical interest in these matters: While Fermat's last theorem presumably drew its mathematical interest from its status in number theory and the connections to multiple concepts investigated in this mathematical field, the interest of computing machines such as $\mathscr{H}$ (the combination of the universal computing machine $\mathscr{U}$ and the decision procedure machine $\mathscr{D}$ that is assumed to exist in Turing's diagonal argument) lies *solely in the application of the diagonal argument*. Put more concretely, the interest in the decision procedure machine $\mathscr{D}$ was motivated by the usefulness that such a decision procedure would have by checking whether a given machine is indeed circle-free, but the decision procedure $\mathscr{D}$ was shown to be impossible to implement not for any *constructible* computing machine, but for the *non-constructible and paradoxical* machine $\mathscr{H}$. However, $\mathscr{H}$ was not one of the machines that motivated the search for a general decision procedure to check a machine's circularity to begin with, it could not even have been known to us *prior* to Turing's proof, as it is a machine that is only constructible *once $\mathscr{D}$ is assumed to exist*. Of course $\mathscr{H}$ cannot be constructible even *after* we haves seen and understood Turing's proof, since the assumption of $\mathscr{D}$ turns out to lead to a contradiction. If Turing had constructed a concrete computing machine that had been overlooked in our search for $\mathscr{D}$ and showed us that no computing machine $\mathscr{D}$ could ever decide the question for this concrete case, we might have given up our search and concluded that the concrete obstacle was unsurmountable. But Turing's proof is non-constructive in this sense, he convinces us to give up our search for $\mathscr{D}$ even though none of the concrete computing machines that motivated the search for the decision procedure were negatively impacted by the proof, based solely on an entirely artificial computing machine that could not even exist outside the sterile environment of Turing's diagonal argument, because its construction depends on the assumption of the general decision procedure $\mathscr{D}$ that is then led *ad absurdum*. Similarly, the unprovable but intuitively correct statements shown by Gödel are never concrete mathematical statements that interest mathematicians in their day to day work, but always artificial sentences that *only draw their interest from their application in Gödel's proof*, similar to how the liar's paradox only troubles us if we consider it from a logical standpoint, without negatively impacting our everyday language.

This is why, even though Turing's view of mathematics is certainly closer to Wittgenstein than a Hilbertian belief in a promised land where mathematics is replaced by a single formalism, his reasoning masks a philosophical confusion that runs afoul of Wittgenstein's views of mathematics and logic. Mathematics can never be reduced to a single language game that would capture its essence, but not because of theoretical results by Gödel or Turing. The impossibilities demonstrated in those arguments are situations that arise when mathematicians restrict the language game of their logical calculus to a consistent set of rules that excludes questions for which 'yes'/'no' answer would lead to a contradictory answer. But then the mathematicians afford themselves the possibility to step outside of this calculus and into *another* consistent calculus which is able to give a consistent 'yes'/'no' answer to the same question that the other calculus was unable to answer. The sleight of hand occurs when the question is called the 'same', even though it has a completely different role and standing in both calculi. The option of stepping out of the calculus and answering the 'same' question in a different calculus is not afforded to the machine, even though it is far from clear why this 'meta-game' could not be 'mechanically' formalised at the cost of accepting inconsistency.

Gödel and Turing provide us with a schema that can be followed to 'outsmart' machines: Given a clearly specified and consistent logical system or a machine that implements it, we can always construct a contradiction. But this means that Gödel and Turing have shown us a *game* that can be taught and followed. The rules of this game are not arbitrary, in fact the rules are convincing enough that we can point someone to these proofs if asked for a justification of our belief that we will *always* be able to construct such a contradiction. Why not teach this game to a machine? Would we then still want to say that we are smarter than this machine? How so, if the machine can point out its own 'mistakes' in exactly the same way that we could? And can we even call these contradictions 'mistakes', if the machine's incompleteness is already exposed by the machine itself? What is it that we would consider 'correct' but the machine cannot do?

In contrast to the impossibility results of Gödel and Turing, the search for a proof of Fermat's last theorem or unsolved problems in mathematics is marked by the *absence of a result*. Either such a proof can be found on the basis of a fixed set of "intuitive judgements" and the proof search is simply incredibly long-running and complex, or the search for a proof requires new "intuition", "new techniques". In the former case, a machine could be able to tackle the task just as well as a mathematician and the difference in efficacy comes down to questions of practicability and implementation. In the latter case, a machine can solve these problems only if it is allowed to learn these new intuitive judgements from the outside world, but then no logical justification can be given, it will be a form of "training" / "Abrichtung". The end result is then again a multitude of mathematical language games that can find their justification only in our form of life, but *not* in any theoretical result by Gödel or Turing.

3.7 STEERING CLEAR OF UNDECIDABILITY

What is philosophically problematic about Turing's proof is not the proof itself and not even the application of the proof inside of mathematics. It is rather the feeling that Turing's proof leaves no alternative, that Turing's notion of "computation", "computable machines" and "decision procedure" is the only possible way to use these words if they are to apply to our empirical reality and not just in the context of a formal game devoid of any purpose. The danger lies in mistaking the logical impossibility shown by Turing, which we accept as the basis of our actions if we accept his language game as a fitting description of our 'ordinary' understanding of computation based on 'idiotic' clerks, with a physical impossibility, because we cannot see how there could possibly by any other way to understand computation and its related concepts. This, then, is where the work of the philosopher begins, by 'imagining tribes' of people who use these concepts in a different way, similar enough to our own use to be able to see similarities, but different enough to contrast with our form of life. In most, perhaps all cases we will declare these other people "mad" (*LFM XXI*, pp. 201–203), because their use of these concepts diverges too much from our understanding to be able to still call it "computation" in our words. The philosophical aim is not to change our usage of such concepts, but merely to point out that we could conceivably adopt different concepts if our form life changed. As a result, we will often see just where the impossibility lies: It is not an impossibility comparable to laws of physics, but a conceptual impossibility tied to our use of these words and their associated language games.

The practical consequences of Turing's proof can vary wildly, depending on whether we see it as a physical or logical impossibility. In the former case, we are being convinced by the proof to give up certain attempts, exactly as we would if someone pointed out that a machine proposed by us would require one of its components to operate at a speed faster than light. In such a case, Turing's proof is the end of our endeavour. However, in the case of a logical impossibility, the proof demonstrates why and how some of our concepts contradict each other and lead to a nonsensical situation. It is then up to us to decide whether this nonsense is and must be useless in the form of life that motivated these concepts and their surrounding language games (as in the case of nonsensical games played for the purpose of amusement) or whether we want to revise our concept in light of the proof so that we might *give a use* to these nonsensical results. In such a case, the proof is not the end, but rather the *beginning of a conceptual investigation* that will include detailed descriptions of how we use or could use these concepts in our life. The proof itself will always be left untouched by these philosophical investigations, it can thus never lose its validity as long as it is mathematically correct, but it might very well lose its status and its usefulness if the use of concepts that motivated the proof change based on their role in our form of life.

As a result, we might start to see contradictions in a logical system not as a "catastrophe", but only as a step on our way to clarify concepts, a process which might lead to language games that give contradictions a useful role instead of banning them. Philosophy cannot say which language games are the 'right' ones, but only strive for a 'surveyable representation' of the role and standing of a particular contradiction, as Wittgenstein notes in a series of remarks Ms-130, a continuation of his remarks in Ms-124 and Ms-127 and which were also included in several typescripts:[35]

> Der Widerspruch ist nicht als Katastrophe aufzufassen, sondern als eine Mauer, die uns anzeigt, daß wir hier nicht weiter können.
>
> Die bürgerliche Stellung des Widerspruchs, oder seine Stellung in der bürgerlichen Welt: das ist das philosophische Problem.
>
> Ich möchte nicht so sehr fragen "was müssen wir tun, um einen Widerspruch zu vermeiden", als "was sollen wir tun, wenn wir zu einem Widerspruch gelangt sind."
>
> Warum ist ein Widerspruch mehr zu fürchten, als eine Tautologie?
>
> Unser Motto könnte sein: "Lassen wir uns nicht behexen!"
>
> Zu meiner Bemerkung: die Philosophie lasse alles, wie es ist, sie lasse auch die Mathematik, wie sie ist: Es ist nicht Sache der Philosophie, den Widerspruch durch eine mathematische, logisch-mathematische, Entdeckung zu lösen. Sondern den Zustand der Mathematik; der uns beunruhigt, den Zustand *vor* der Lösung des Widerspruchs, übersehbar zu machen. (Und damit geht man nicht etwa einer Schwierigkeit aus dem Wege.) [Ms-130, 13.4–14.5]

> The contradiction is not to be understood as a catastrophe, but as a wall that indicates to us that we cannot go any further here.
>
> The civic status of the contradiction, or its status in the civic world: that is the philosophical problem.
>
> I don't want to ask so much "what must we do to avoid a contradiction" rather than "what shall we do when we have arrived at a contradiction?"
>
> Why is a contradiction to be feared more than a tautology?
>
> Our motto could be, "Let us not be bewitched!"
>
> As to my remark: philosophy leaves everything as it is, it also leaves mathematics as it is: it is not the task of philosophy to solve the contradiction by a mathematical, logico-mathematical, discovery. But to make the state of mathematics which worries us, the state *before* the solution of the contradiction, comprehensible. (And in doing so, one does not sidestep a difficulty).

35 See Ts-228, 159.6–160.4, Ts-230a/b/c, 35.6–36.4, Ts-233b, 60.3–60.6 (*Z* §§687–690, which does not include the second and last remark of the remarks in Ms-130) and also Ts-227a, 88a.1 (*PI* §125, which includes only the second and last remark of the remarks in Ms-130).

The aim of 'imagined tribes' and examples of a different use of contradictions is not to solve the 'problem' caused by a contradiction, as this would be a task for mathematicians. Instead, these examples can help to show how a contradiction can fit into the fabric of interrelated language games that are part of our form of life. To find a new place for a contradiction in this fabric, even if we never go on to use it in such a way in practice, can help to give it a place in our "filing-system" or "filing cabinet", a picture that Wittgenstein uses first in 1941 in Ms-124 and then in 1944 also in Ms-130, where it occurs shortly after Wittgenstein's remark on the contradictory royal orders already quoted earlier:

> Es ist unglaublich, wie eine neue Lade, an geeignetem Ort in unserem filing cabinet, hilft. [Ms-124, 25.2]
>
> Das Ergebnis einer philosophischen Untersuchung ist manchmal ein neues 'filing-system'. [Ms-130, 82.3]
>
> ---
>
> It is incredible how a new drawer, in a suitable place in our filing cabinet, helps.
>
> The result of a philosophical investigation is sometimes a new 'filing-system'.

The most detailed use of this picture can be found in Ms-132 and explains how such a new place in our filing cabinet can lead us to change our attitude towards things that we previously only saw as "filth" or "vermin", such as contradictions in mathematics, as we might add in view of Ms-124 and Ms-130:

> Wenn Du Dir die Welt schön geordnet denkst, für alles ist eine Lade vorhanden, alles ist schön & reinlich, – nur eine Sache paßt in keine der Laden hinein – {man // Man} hat nur *ein* Gefühl: "Oh, wäre doch *das* nicht da! Es verunziert die schöne Ordnung der Dinge." Man verhält sich {bloß // einfach} ablehnend zu {diesem // dem} Ding. Man sagt nicht "Es hat auch {einen // seinen} Platz in der Welt", sondern "Es ist Schmutz, Ungeziefer, Unkraut".
>
> Wenn wir unser schönes, reinliches filing-cabinet haben, & nur ein Ding paßt nicht hinein, & bleibt draußen liegen so möchten wir es am liebsten einfach los werden. Gibt uns Einer aber ein anderes System von Laden & das Ding, das früher heimatlos war, findet nun seinen Platz, so verändert sich unsere Stellung zu ihm {gänzlich // ganz}. [Ms-132, 57.2]
>
> ---
>
> If you picture the world beautifully ordered, with a drawer for everything, everything is neat & clean, - only one thing doesn't fit into any of the drawers - you have only *one* feeling: "Oh, if only *that* weren't there! It spoils the beautiful order of things." One behaves {only // simply} disapprovingly of {this // the} thing. One does not say "It also has {a // its} place in the world," but rather "It is filth, vermin, weeds".
>
> If we have our neat, clean filing-cabinet, & only one thing does not fit in, & remains outside, we would like to get rid of it. But if one gives us another system of drawers & the thing that was homeless before now finds its place, our position towards it changes {completely // entirely}.

This chapter is an attempt to give a seemingly contradictory machine such as $\mathscr{D}$ a place in our filing cabinet. The critique here is not that

Turing's concept of computability and the supposed capabilities of his machines would go *too far*. On the contrary, the critique of this chapter is that *Turing does not go far enough* and that he restricts his concept of computation to a limited area that does not include many activities of mathematicians that could conceivably be carried out mechanically by machines and thus called computable, which would require inconsistency to accommodate contradictory decision procedures and paradoxical machines.[36] This creates a situation where we think that Turing's usage of terms such as "calculate", "verdict", "decision", etc. corresponds well enough to our ordinary usage to be able to draw conclusions from Turing's argument about the limits of *practical* machines. After all, if these limits apply to idealised theoretical machines, then we would expect them to apply all the more so to practical machines.[37] In the same way that the proof of the impossibility of trisecting an angle using compass and straightedge might convince us to give up our search for a practical method of trisecting an angle with these tools at hand, Turing's proof has the ability to convince us to give up the search for certain general decision procedures. But if there is a mismatch between Turing's usage of certain terms and our ordinary understanding of them, then there is the danger of giving up the search for such a decision procedure prematurely, even though we might be able to find one that corresponds to our expectations of generality. Of course in case we ever do find such a 'general' decision procedure, it would have no direct relation with Turing's proof, because we would use terms such as "calculate", "verdict" and "decision" differently from Turing. Such a philosophical investigation can therefore never impact the validity of the proof, but it might shift the focus in such a way that the proof becomes less important to us if we start to question its applicability to our usage of the terms that initially lent the proof its motivation.

Seen in this light, Turing's proof could then lead to an investigation of exactly the area of reasoning that his diagonal argument seemed to exclude, which encompasses contradictory decisions and logical inconsistency in small doses. This will require us to revise our concepts, but it need not lead to the collapse of bridges.

---

36 Of course there is no guarantee that there will not be other reasons for believing that these activities cannot be carried out mechanically, but in this text we are only concerned with the limits demonstrated by Turing's line of argument.

37 See Hopcroft, Motwani, and Ullman, 2001, p. 307: "[T]he Turing machine long has been recognised as an accurate model for what any physical computing device is capable of doing."

CONCLUSION

> Das Überraschende, Paradoxe ist paradox nur in einer gewissen, gleichsam mangelhaften, Umgebung. Man muß diese Umgebung so ergänzen, daß, was paradox schien nicht länger so erscheint. [Ms-124, 141.3 / *BGM VII* §43]

> Something surprising, a paradox, is a paradox only in a particular, as it were defective, surrounding. One needs to complete this surrounding in such a way that what looked like a paradox no longer seems one. [*RFM VII* §43]

Before attempting to draw general conclusions from the preceding investigations of Wittgenstein's remarks on three different diagonal arguments, it must be pointed out that any attempt to do so is bound to violate precisely those philosophical convictions that led Wittgenstein to embark on the different investigations discussed here. Wittgenstein's philosophy is, at least in the later years of his life, decidedly undogmatic and focused on concrete uses of language, requiring a variety of investigations for a variety of concepts. A general discussion (or even worse, an explanation) of concepts such as enumerability, provability and decidability will thus only fall prey to the dangers that Wittgenstein warned about and produce at best agreeable trivialities and at worst philosophical nonsense.

One reason for this, which has been discussed at different points in the three preceding chapters, is a mistaken desire for uniformity, arising as a byproduct from an overly formal and mathematical treatment of our informal concepts. Inspired by a translation of our informal concepts of numbers, theorems and machines into rigourous mathematical systems, we are led to believe that these formal translations could supplant the seemingly primitive variety of informal language games. From such a perspective, the informal descriptions of these concepts will appear deficient and in need of a formal explanation. The three diagonal arguments discussed in this text all demonstrate limitations of formal concepts (namely the impossibility of enumerating all real numbers, of deciding all propositions, of deciding in general whether a machine will come to a halt), but their diagonalised constructions are then interpreted as transcending the formal systems, which enshrines the diagonal arguments as fundamental results that demonstrate limits in our mathematical foundations. These results can easily appear to us as limits that *must* hold, as if they were laws of nature that govern the ideal world of platonic numbers and logic, giving them an "ultraphysical" appearance of rigidity and hardness.

This explains Wittgenstein's interest in these seemingly overly specialised pieces of mathematics: All three diagonal arguments can be seen as advocating for a treatment of mathematical concepts as requiring *higher-order systems*, since a single system (an enumeration of the real numbers, a formal system for theorems, a formal system for computing machines) is seemingly insufficient to accommodate all objects in question, so that a higher-order system becomes necessary (and then again systems of increasingly higher orders, to hold the diagonalised objects that escape each higher-order system). Wittgenstein is suspicious of those meta-systems (a tendency which goes back as far as the *Tractatus* with its distaste for Russell's theory of types), because they only seemingly explain a concept, while clinging to the idea that a uniform and formal treatment has *explanatory* power, whereas from Wittgenstein's perspective the only way to clarify concepts is to *describe* them in light of their different uses, without reducing them to a single 'foundational' formal system.

What Wittgenstein shows in his investigations is that no formal system can decide for us how to proceed at the crucial point of the contradiction that results from diagonalisation in each of the three different arguments (by constructing a real number that would not occur in the enumeration of all real numbers, by constructing a proposition that would, if it were provable, also say that it is not provable, by constructing a computing machine that would not be a computable number if it occurred in the enumeration of all computable numbers). At this crucial point, when faced with a contradiction, all three diagonal proofs proceed by excluding the contradictory construction and thereby manage to preserve consistency, at the cost of moving the contradictory construction to a higher level: The results are transfinite numbers (which can enumerate even the real numbers that 'escaped' through diagonalisation), hierarchies of proof systems (which capture all theorems that 'escaped' through diagonalisation on a lower level), and uncomputability (which is the name for decision procedures that 'escaped' through diagonalisation).

In all three cases, Wittgenstein critically examines the formal ideal of consistency and points out that the conclusions of the diagonal arguments only seem inevitable if we are not prepared to accept the contradictory result of the diagonalisation as an object in the formal system. In the case of Cantor and the enumeration of the real numbers, the result would be a real number that is contradictory in its diagonalised digit (which through diagonalisation is defined to be different from itself). In the case of Gödel, the result is a proposition that is provable and yet says that it is not provable, if it is translated from its arithmetic representation into a meta-mathematical statement. In the case of Turing, the result is a computing machine that both halts and does not halt.

As the different chapters have demonstrated, Wittgenstein's intent is not to advocate for a trivialist or paraconsistent treatment of inconsistency, since such an interpretation of the mathematical results would be just as dogmatic and philosophically one-sided as the interpretations that Wittgenstein is critically examining. Instead, his remarks need to be read as emphasising that there is a *decision* to be made when we arrive at the diagonalised construction and that we are, at least on the basis of the mathematical proof, free to choose either consistency (and therefore exclude the diagonalised construction from the system) or inconsistency (and accept the paradoxical nature of the diagonalised construction).

From the perspective of Wittgenstein, the mistake of most interpretations of the three diagonal arguments is to suggest that the mathematical argument could on its own provide the justifications for a decision in favour of consistency, when this actually goes far beyond what any purely mathematical argument can accomplish. As Wittgenstein points out, consistency is not an ideal in and of itself, but merely a principle that has proven itself so useful in a large variety of language games that we accept it as an unquestioned rule even in cases where the situation is radically different.

The situations introduced by the three diagonal arguments are all characterised by the uselessness of their diagonalised constructions: We do not use Cantor's diagonalised number for anything practical, nor Gödel's undecidable proposition, nor Turing's paradoxical computing machine. All of these constructions have sense only within their formal systems and are almost parasitically dependent on them. As Wittgenstein notes, our usual reasons for requiring consistency do not apply here, as it is not clear why inconsistency would lead to trouble, apart from perhaps violating purely formal demands. In regular situations (that is to say in situations where the objects of the formal systems are not diagonalised constructions that depend on the precise formulation of the system in question) we can easily explain why inconsistency would make the formal system useless. But in the case of diagonalised constructions, this analogy breaks down and the primary reason for excluding inconsistency is our desire for a uniform treatment of all objects in the system.

Distinguishing between (extra-systemically) useful and useless constructions is beyond the capabilities of the mathematical proofs. For Wittgenstein, this is not due to a deficiency of these proofs, but rather a direct consequence of what mathematics can and cannot do, which constitutes the border between mathematics and philosophy. The task of describing how a mathematical concept is used (which can include both intra-mathematical and extra-mathematical applications) falls to philosophy, precisely because mathematics is concerned with *surveyability* in a different sense than philosophy: As Wittgenstein emphasises frequently, a mathematical proof must be surveyable to be considered a proof, and mathematics is concerned with surveyability in this sense. But from the perspective of a mathematician, it is perfectly acceptable to consider arithmetical concepts purely in terms of a recursive definition such as Peano arithmetic, which (with its definition of numbers as 0 and the repeated application of a successor function) is completely impractical for arithmetic calculations of any substantial size and *unsurveyable* in this philosophical sense.

In Wittgenstein's philosophy of mathematics, the actual language games that might appear to act only as examples or as motivation for their later formalisation are not merely primitive secondary stimuli for the primary formal system, they are instead essential for an understanding of the formal system to begin with, because the actual language games in all their variety lead to a surveyable representation of our concepts in a way that is not possible by merely considering the uniform treatment in the formal system.

As Wittgenstein shows, Cantor's proof can stipulate what we *call* a real number and that we exclude the diagonalised number instead of extending our concept of numbers to include such a construction. We are certainly free to adopt this conceptual stipulation, but Cantor's proof cannot on its own justify *why* we adopt it. The same holds for Gödel's first incompleteness theorem, which can lead us to exclude Gödel's proposition P as undecidable, but cannot justify this exclusion (instead of choosing inconsistency) on purely mathematical grounds, that is to say not on the basis of the mathematical proof alone. In Turing's case, the diagonal argument can show how the assumption of a general decision procedure for deciding whether a machine halts will lead to a contradiction, but it cannot justify the exclusion of the resulting paradoxical machines from our concept of such a decision procedure. The limits demonstrated by the different proofs are 'merely' logical impossibilities, which does not imply that they would be of a different 'hardness' than physical impossibilities, but only that they play a different role in our language, because they reflect our rules of language and are not directly dependent on empirical discoveries.

This then leads from the distinction between physical and logical impossibilities to an outlook on the hinge propositions of the later Wittgenstein. Incorporating texts such as *On Certainty* would have been outside the scope of this thesis, but the connection with Wittgenstein's late writings should nevertheless be noted here and could be expanded upon in future publications. In the context of *On Certainty*, a philosophical investigation in the sense of Wittgenstein can show us not only that the impossibilities demonstrated by diagonal proofs are logical impossibilities, but more importantly that these impossibilities could play a less important role if considered in the context of a different form of life. In contrast to propositions such as '2 + 2 = 4', which are fixed as hinges by a multitude of activities, the hinge propositions at play in the diagonal proofs are much less firmly fixed by non-mathematical use than basic arithmetic (if they are hinge propositions at all), which makes it much easier to picture alternative forms of life and, in the case of diagonal proofs, to possibly accept inconsistency “in small doses”.

While the three diagonal arguments discussed here could at first glance appear to be of interest only to mathematicians, Wittgenstein’s remarks demonstrate why that is not the case. All three arguments are either considered to be foundational results in their respective fields or deal with fundamental concepts that we frequently use informally in our lives: The interpretation of Cantor’s proof deals with the concept of (real) numbers and of (different notions of) infinity; Gödel’s proof is one of the most important results in mathematical logic and has led (whether intended by its author or not) to uses and abuses of the mathematical result in the philosophy of mind; Turing’s proof still forms one of the fundamental formalisations of the concept of computation and shapes our view of what computers can do in theory. A misleading interpretation of these results as demonstrating “ultraphysical” impossibilities can therefore have serious implications and influence our understanding of much more than just a specific mathematical proof.

Wittgenstein’s investigation of the three diagonal arguments thus shows what is at stake in these particular proofs and why Wittgenstein is interested in them: Although the mathematical proofs themselves are perfectly valid, we have the tendency to interpret them not as merely demonstrating logical impossibilities, which are reflections of our rules of language, but as “ultraphysical” impossibilities, comparable to laws of nature, governing the ideal realm of mathematics. Such a misleading picture is the result of a “one-sided diet”, because we lack surveyability: We fail to grasp the concepts in all their various uses and in the context of how they fit into our form of life. The antidote is to describe them in a surveyable representation, sometimes by imagining different forms of life.

A

BEYOND WITTGENSTEIN: KOLMOGOROV COMPLEXITY

> Wenn man einen Taschenspielertrick {entlarven // aufdecken} will, muß man alles mißtrauisch prüfen, was der Mensch für gewöhnlich ohne Prüfung hinnimmt. *Will* man dies hinnehmen, so ist es unmöglich auf den Trick zu kommen. [Ms-159, 23r.1]

> If one wants to {expose // uncover} a sleight of hand, one has to examine all that suspiciously which a person usually accepts without examination. If one *wants* to accept this, it is impossible to find the trick.

In the decades after the original diagonal argument by Cantor, the diagonal method was employed in a number of different mathematical arguments, with the proofs by Gödel and Turing being the most famous examples. Wittgenstein's level of engagement with these different diagonal arguments varied wildly, however: He devoted more than a whole notebook to Cantor's original diagonal argument, only a short typescript and scattered remarks to Gödel's proof and merely a single specific remark to Turing's application of the diagonal method. At first glance, it could seem as if Wittgenstein's interest in the diagonal method had waned over time and that in his view there remained little to say about the specific uses by Gödel and Turing. It must then appear questionable to write about Wittgenstein in the context of an application of the diagonal argument that first appeared in the 1970s, decades after Wittgenstein's death in 1951, and which can be seen as a continuation of the diagonal arguments by Gödel and Turing.[1] The incompleteness results discussed here, due to Gregory Chaitin, form the backbone of the field known as *Algorithmic Information Theory*, which combines aspects of *Information Theory* as inaugurated by Claude Shannon with the computational foundations pioneered by Turing. Contrary to its appearance as a purely theoretical and rather specialised field, algorithmic information theory offers some of the most practical and easily comprehensible applications of all diagonal arguments, with its central notion of transmitting a minimal amount of information from a source to a receiver. While some of the results in algorithmic information theory are highly technical and beyond the scope of the current discussion, there is also a wealth of elegantly simple proofs with accompanying metamathematical and philosophical

1 In fact, most of the incompleteness results considered in the following text are not usually described as diagonal arguments, because they build upon the results of Turing but do not directly apply the diagonal method. As will become clear, however, these incompleteness results show enough similarities to the more traditional diagonal arguments by Cantor, Gödel and Turing to be discussed along with them and can be reformulated in a way that more directly emphasises their diagonal nature.

issues, which are of more immediate interest to non-mathematicians than for example Cantor's application of the diagonal method with its backdrop of the real numbers and set theory.

Chaitin's incompleteness results form an interesting object of study for a philosophical investigation in the spirit of Wittgenstein, not only due to their comparatively practical embedding in information theory, but also because they stand at the end of a series of diagonal arguments that were all of interest to Wittgenstein and which, each in their own way, raise important philosophical questions. That Wittgenstein himself did not write about the incompleteness results discussed here is not so much a hindrance as an opportunity to apply his philosophy in a way that goes beyond an exegetical interpretation, while at the same time staying faithful to Wittgenstein's remarks. If we take Wittgenstein's aim of not advocating theses seriously, we cannot expect his remarks to provide eternally valid insights, only clarifications of concepts that are tied to their particular time and culture. New developments thus require new philosophical investigations, in areas that only recently became an important part of our way of life. The following investigation is one such attempt, focused on some of the most relevant concepts in our digital age: information and complexity.

CHAITIN'S DIAGONAL ARGUMENT

Before considering the incompleteness results arising in algorithmic information theory, it can be helpful to lay some of the groundwork by introducing the concept of *information*, which stands at the root of information theory and was first given its current mathematical definition by Claude Shannon in 1948 in his seminal paper "A Mathematical Theory of Communication" (Shannon, 1948). In it Shannon considers how the communication of messages can be described mathematically and proposes a definition for the amount of information contained in a particular message. Let us assume that two people want to communicate over a long distance at night with the help of torches or flashlights that can be switched on or off. If they wish to transmit a yes or no message, they can agree to simply turn on the flashlight for one second to mean yes and turn it off for one second to mean no, but if they plan to transmit one of 10 possible messages (a single digit from 0 to 9, for example) then they need to agree on a more complex code that is formed by a sequence of on/off signals over the course of multiple seconds (for example 4 seconds off = 0; 3 seconds off and 1 second on = 1; 2 seconds off then 1 second on and then 1 second off = 2, etc.). It is not hard to see that a sequence of on/off signals can be understood as a sequence of *bits* (a term coined by Shannon as a short form for "binary digit", Shannon, 1948, p. 380) and that a sequence of N bits can represent one of 2^N distinct messages. $\log_2 N$ is then the number of bits necessary to describe a particular message from a set of N messages and is defined as the "measure of the information produced when one message is chosen from the set, all choices being equally likely" (Shannon, 1948, p. 379). This definition of the information of a message, $I = \log_2 N$, captures our intuitive understanding of what is less and what is more informative: If there is only a single possible message, the information of the message is 0, because there is no need to wait for the message to know what it will be, but if there are 4 equally likely messages, every message contains 2 bits of information (since the 4 messages can be encoded in 2 bits as '00', '01', '10' and '11') and to communicate such a message and distinguish it from the other possible messages at least 2 bits need to be transmitted. The number of bits necessary to represent the information in a book of 200 pages is thus usually roughly twice the number of bits necessary to represent the information in a book of 100 pages, which corresponds to our intuitive notion of information content.

While the examples so far assumed that all messages are equally likely, this is rarely the case in practice. If we consider the letters of the alphabet as individual messages, for example, and intend to transmit a sequence of them to communicate an English text, then the frequency of the different letters will vary quite substantially, with the vowels occurring much more frequently than 'q', 'x' and 'z'. A probabilistic definition of information, which does not treat each message as equally likely but incorporates the probability of the occurrence of a particular message, can adequately model this situation and considers rare messages such as 'q', 'x' and 'z' as more informative than more common messages such as 'a' and 'e' (Shannon, 1948, pp. 384–89). This is a straightforward extension that is equivalent to the non-probabilistic definition in cases where all messages are equally likely, but allows a compression of information in cases where some messages are more likely than others. It is no accident that the term 'bit' has proliferated thanks to the ubiquity of digital computers, because such a probabilistic concept of information lies at the root of many practical applications and is involved whenever digital data is compressed and transmitted.

The next important concept, *complexity*, marks the transition from classical to algorithmic information theory. Equivalent definitions of information-theoretic complexity were published independently in the 1960s by Ray Solomonoff (Solomonoff, 1964b, Solomonoff, 1964a), Andrey Kolmogorov (Kolmogorov, 1968a, Kolmogorov, 1968b) and Gregory Chaitin (Chaitin, 1966), with the concept known as "Solomonoff-Kolmogorov-Chaitin complexity" or more simply "Kolmogorov complexity". The basic idea is quite elegant and simple: Instead of considering information from the perspective of an *ensemble* of messages (and the information of a single message against the backdrop of the probability of all other possible messages), Kolmogorov complexity focuses on the information content of "individual objects" (Chaitin, 1982b, p. 118, Chaitin, 1982a, p. 75) and defines the complexity of an individual object as the minimal length of a program that produces this object (Kolmogorov, 1968a, p. 662), which is why Kolmogorov complexity is sometimes called program-size complexity. A program of the form 'print the first 1000 decimal places of π in binary' will produce a binary sequence of length 1000, while itself having a particular program size, which is the length of the sequence that acts as the binary representation of the program in the chosen computing model, for example in 500 bits. The sequence of the first 1000 bits of π thus has a Kolmogorov complexity of at most 500 bits, because the program with a length of 500 bits acts as a compressed description of the full sequence of 1000 bits and could be transmitted whenever we would normally transmit the full 1000 bits with no loss of information. In algorithmic information theory, programs represent sequences of bits, because programs produce sequences of bits as their output (for example as a series of '0' and '1' on the tape of a Turing machine or via a statement such as 'print "01011100101"' in a programming language) and are themselves encoded as a sequence of bits (as a series of '0' and '1' on a tape to be executed by a universal Turing machine or as the compiled binary representation of a programming written in a particular programming language). As a consequence, the size of programs and the length of binary sequences can be used interchangeably, and algorithmic information theory is concerned with minimal programs that fully describe larger binary sequences.

This definition of complexity with its focus on computable programs might seem overly theoretical, but should be understood in the more general context of minimal *descriptions* for data of any kind, for example scientific observations, the context in which Solomonoff first proposed such a definition. Let us assume that a scientist has observed a sequence of 100 data points that are all of the form on/off or 0/1 and wants to describe these findings by formulating a scientific theory. If the sequence is a random series such as '0111001001010011...' without any discernible pattern, all the scientist can do to describe it is to publish it in full, which will require at least the 100 data points in the sequence. But if the sequence is of the form '01010101...', it is possible to describe it as '01 repeating', which will be a much shorter description with a suitable encoding than the full list of 100 bits. This difference in description size becomes even more pronounced for large finite or infinite sequences, with 'π' as a very succinct description for the infinite sequence '3.14159...'. Kolmogorov complexity is nothing else than the formalisation of this idea using unambiguous programs such as Turing machines instead of English language descriptions and thus presents a further generalisation of the concepts of information theory. It builds on the idea that data can be compressed if the data contains less information than its size theoretically allows, so that 100 bits of data that contain only 20 bits of information can theoretically be compressed down to 20% of its original size. But in contrast to classical information theory with its probabilistic approach, the algorithmic approach can theoretically exploit any pattern using a suitable description, even in cases which look probabilistically random (such as the digits of π, where each of 0 to 9 are equally likely).

Complexity in algorithmic information theory is thus closely connected to *randomness*, in fact complexity can be understood as a mathematical measure of the randomness of a (possibly infinite) sequence of data points. A sequence like π might look random but actually contains only a small and easily formalisable amount of information, so that the Kolmogorov complexity (measured as the program size) of the first billion decimal places of π is only insignificantly larger than the complexity of the first 1000 decimal places, because the programs 'compute π up to its billionth decimal place' is only slightly longer than the program 'compute π up to its 1000th decimal place'. From the perspective of algorithmic information theory, π is not very complex, because it can be described and calculated by very short programs. But for most binary sequences produced by repeatedly tossing a coin it will not be possible to describe them in any way shorter than simply listing the whole sequence. These incompressible sequences are called *random* in algorithmic information theory (Chaitin, 1975, p. 14), though it has to be pointed out that this definition is only concerned with the resulting object, independent of whether or not it was produced by a random process. Randomly flipping a coin a thousand times can certainly result in a sequence of 1000 consecutive zeros, but this sequence would not be random from the perspective of algorithmic information theory, because it can easily be described by the much shorter program '0 repeated 1000 times' (cf. Chaitin, 1975, p. 12). Algorithmic information theory can at first glance seem like an overly specialised subfield of information theory, but in fact the focus on programs as descriptions is explained by the fact that programs can be seen as "the most general decoder for compressed messages":

> In summary, information theory teaches us that messages from an information source that is not completely random (that is, which does not have maximum entropy) can be compressed. The definition of randomness is merely the converse of this fundamental theorem of information theory; if lack of randomness in a message allows it to be coded into a shorter sequence, then the random messages must be those that cannot be coded into shorter messages. A computing machine is clearly the most general possible decoder for compressed messages. We thus consider that this definition of randomness is in perfect agreement and indeed strongly suggested by the coding theorem for a noiseless channel of information theory. [Chaitin, 1970, p. 48]

Here we come to the question of how these programs are specified. It is clear that all the examples so far have only been 'pseudo-code', while a 'real' program in a practical programming language would usually be larger and a program in form of a Turing machine would be larger still, by a considerable factor. Even a comparatively simple program for the calculation of π can grow quite large on something as impractical as a Turing machine, so much so that such a program description can quickly dwarf the supposedly more complex approach of simply listing all the decimal places in the program one by one instead of calculating them, even for large sequences. On more practical computers, a program for the calculation of π might be considerably shorter. Which of these computing models are we then to choose as the underlying measure of program-size complexity? More importantly, how valuable are the results of algorithmic information theory if they depend in large parts on something as specific as the choice of the computer used to hypothetically run these programs?

This apparent dependency on the particular computing platform is why algorithmic information theory is only concerned with asymptotic results, which hold for infinitely long sequence independently of which theoretical computing model is used. If we consider arbitrarily long sequences, the differences between different theoretical computing models cease to matter for very long sequence, simply because each computing model can *simulate* any other model (at the cost of speed, which is of no importance in the current theoretical discussion, cf. Chaitin, 1975, p. 16). We can simply prefix the program on the more cumbersome Turing machine with a program that simulates the more ergonomic programming language and then run the more ergonomic program that uses the constant π on the more cumbersome Turing machine. The first part of this program, the simulation of the ergonomic computer, stays of constant length no matter how large the sequence in question grows and will therefore be insignificant for extremely large sequences. That computers can simulate each other is a direct consequence of Turing's proof of the universality of a Turing machine and the possibility to express such a universal Turing machine as a program on a Turing machine (see Section 3.1), but can also be intuitively explained by our ability to implement any programming language in any other (Turing-complete) programming language.[2]

2 To be more precise, this is true only if all imaginable models of computation are equivalent to Turing machines, in other words if the Church-Turing thesis holds. This 'thesis', which lacks the mathematical rigour necessary to definitely prove or disprove it, can certainly be attacked philosophically, see for example Shanker, 1998 for a philosophical critique from the perspective of Wittgenstein. At least for now, this will not be the aim of the current text, not only since it would go beyond the scope of this discussion (Chapter 3 touches on the issue in passing), but also because in the context of algorithmic information theory the status of the Church-Turing thesis is of secondary importance. In the context of the current discussion, the Kolmogorov complexity for a given sequence will asymptotically be the same on all the computing models known to us, disregarding the constant factor necessary to simulate a particular model.

With the choice of the computing model out of the way, we can return to information-theoretic randomness. What are examples of truly random sequences, we might ask, in other words examples of sequences that cannot be compressed? This is where it gets interesting and where we finally enter the realm of the incompleteness results in algorithmic information theory. To show that a sequence is random, we have to show that its Kolmogorov complexity is approximately equal to the length of the sequence itself, which is the case if there is no shorter program that would produce the same sequence. Let us assume that we have a very short program with a size of 100 bits. Its Kolmogorov complexity can be at most 100 bits, since the program itself is a description of itself in 100 bits, but the Kolmogorov complexity could be less than 100 bits if there is a shorter program that produces the same output as our example program. To calculate the Kolmogorov complexity of this program, we could systematically check all programs with a size of less than 100 bits, going from smallest to largest program and stopping if we find a program with the same output as our example program. In case we find such a shorter program, it will be the minimal program and the size of this program is the Kolmogorov complexity of our example program. Otherwise our example program is already minimal. Since a minimal program cannot be compressed any further or it would not be minimal, the sequence of bits of a minimal program is random in the sense of algorithmic information theory.

But how do we check the Kolmogorov complexity systematically for all shorter programs? To know whether a particular program produces the same binary sequence as our example program, we need to wait for the particular program to produce an output, which might never happen if the program enters an infinite loop. We thus need to first check whether the particular program being considered at each step in the search will halt, which is undecidable in general as a consequence of Turing's halting problem (see Chapter 3) and as a result Kolmogorov complexity must be undecidable as well.[3] Such a proof is somewhat unsatisfying, however, as it depends entirely on Turing's diagonal argument and presupposes all of its philosophical assumptions (see Section 3.4), which makes it difficult to properly examine the philosophical underpinnings that are unique to Kolmogorov complexity and algorithmic information theory. A more self-contained proof can show paradoxical aspects of Kolmogorov complexity without depending on Turing's diagonal argument, by demonstrating that we cannot calculate the Kolmogorov complexity for sequences whose length exceeds a particular threshold, as follows: Let us assume that there exists a systematic way to calculate the Kolmogorov complexity

3 An indirect proof via halting undecidability, as sketched, is indeed the proof chosen in many text books focused on computer science, since they need to introduce the halting problem anyway. See for example Cover and Thomas, 2006, p. 483.

of sequences of any length and that we have a program $\mathscr{K}$ for computing it, with a size of k bits. Then we can construct a program $\mathscr{D}$ which does the following, with the constant x chosen so that $k + x$ is greater than the size of the whole program $\mathscr{D}$:

1. It systematically runs through all possible sequences, going from smallest to largest ('0', '1', '00', '01', '10', '11', '000', etc.) and for each one does the following:

2. It calculates the Kolmogorov complexity of the sequence in question using the program $\mathscr{K}$ and based on the result does the following:

3. If the complexity of the sequence in question is greater than $k + x$, the program stops and produces the sequence in question as its output. Otherwise it continues with the next sequence in step 1.

Choosing x so that $k + x$ is greater than the size of the whole program $\mathscr{D}$ is easy enough in practice, because the bulk of the program size of $\mathscr{D}$ will come from $\mathscr{K}$ (whose size is already covered by k) in step 2, so that x only needs to cover the additional program size required for the steps 1 and 3, which are quite simple. Let us assume that these steps will add less than a million bits to the total program size and set $k + x$ to $k + 1\,000\,000$, which is then greater than the program size of $\mathscr{D}$.

What happens if the program $\mathscr{D}$ tries to find a program with Kolmogorov complexity greater than $k + 1\,000\,000$? There must be a program with at least such a complexity, because there are infinitely many possible binary sequences but only finitely many programs with a size less than or equal to $k + 1\,000\,000$ and each of these finitely many programs can only be the minimal program of one of the infinitely many possible binary sequences. Eventually, the program $\mathscr{D}$ will have run through all of the binary sequences that are represented by one of the finitely many programs with a size less than or equal to $k + 1\,000\,000$, it will then test a binary sequence with a Kolmogorov complexity greater than $k + 1\,000\,000$ and produce this sequence as its output. But this would mean that the program $\mathscr{D}$, with a program size of *less than or equal to* $k + 1\,000\,000$, acts as a minimal or at least shorter program for a binary sequence which, according to the result of $\mathscr{K}$, has a minimal program-size of *more than* $k + 1\,000\,000$. Our assumption of a program $\mathscr{K}$ for deciding the Kolmogorov complexity for arbitrary sequences thus leads to a contradiction, because once we cross a certain threshold we could simply enumerate all programs and search for one with a complexity greater than the complexity of the searching program and return it as the result of the search. As a consequence, we know that there must be infinitely many incompressible and thus random binary sequences after a threshold of $k + x$, but we cannot find a single example of such a large and random binary sequence, because if we were to find it we would have immediately given it a description of the form 'the first large random binary sequence found by the program ...', which would make the sequence compressible:

> This is the surprising result that we wished to obtain. Most strings of length n are of complexity approximately n, and a string generated by tossing a coin will almost certainly have this property. Nevertheless, one cannot exhibit individual examples of arbitrarily complex strings using methods of reasoning accepted by Hilbert. The lower bounds on the complexity of specific strings that can be established are limited, and we will never be mathematically certain that a particular string is very complex, even though most strings are random. [Chaitin, 1974, p. 63]

Large and random binary sequences seem to escape our reach as soon as we try to name them: We know they must be there, but we cannot "exhibit a specimen of a long series of random digits" (Chaitin, 1975, p. 19), which appears to be a deep discovery about the randomness of numbers and holds (for large enough bit strings) independently of the particular computing model being used. As Chaitin notes, this "enigma" limits what is "possible in mathematics":

> Although randomness can be precisely defined and can even be measured, a given number cannot be proved to be random. This enigma establishes a limit to what is possible in mathematics. [...] This limitation is not a flaw in the definition; it is a consequence of a subtle but fundamental anomaly in the foundations of mathematics. [Chaitin, 1975, pp. 11–12]

Before examining the philosophical issues of such an apparently astounding discovery, it makes sense to discuss the analogies of the above proof with diagonal arguments in general and certain other paradoxes in particular. The proof as described here does not use diagonalisation in the same way as Cantor's original argument or even Turing's application of the diagonal method, but it does show important parallels to these more obviously diagonal proofs: It proceeds by considering an enumerably infinite sequence of objects (namely programs for which the program $\mathscr{K}$ computes their Kolmogorov complexity, ordered by their size in bits) and then constructs an object of the same form (because the object is a program and so $\mathscr{K}$ should be able to compute its Kolmogorov complexity) which however can be shown not to belong to the infinite sequence, contrary to the assumption (or $\mathscr{K}$ would not be general). In the same way that Cantor's diagonalised number is able to 'escape' the real numbers by depending on the (enumerable) totality of the real numbers or that Turing's diagonalised machine $\mathscr{H}$ 'escapes' the halting machines (or rather the "circle-free" machines in Turing's original argument) by depending on the (recursively enumerable) totality of the halting machines, the program $\mathscr{D}$ 'escapes' the programs for which the Kolmogorov complexity is decidable by depending on the (recursively enumerable) totality of these programs. In all of these cases the contradiction at the end of the argument leads to a rejection of the assumption and thus to a rejection of the possibility of a mechanical way to enumerate exactly those objects that are the real numbers, the halting machines and the programs with a computable Kolmogorov complexity, respectively.

It should be pointed out that the most obvious proofs of the undecidability of Turing's halting problem, which proceed almost exactly like the proof of the undecidability of Kolmogorov complexity described here and construct a Turing machine for which any halting decision leads to a contradiction, are not immediately recognisable as diagonal either and are in fact not due to Turing, whereas Turing's own diagonal argument proceeds somewhat differently and is overtly diagonal. The halting problem is seen to be close enough to Turing's own diagonal proof to follow immediately from it, which is why the analogous proof of the undecidability of the Kolmogorov complexity merits a discussion in the context of these other diagonal arguments, even if it is not itself overtly diagonal.[4]

The resemblance to these other foundational proofs is further emphasised by the similarity of the above proof with the Berry paradox, which exists in many versions, for example as "The first positive integer not definable in under 11 words", which is an English definition of 10 words that, by being a definition of a number, defines the number in under 11 words, in contradiction to its own definition.[5] The analogy between Berry's paradox and incompleteness proof for Kolmogorov complexity has been repeatedly noted by Chaitin (Chaitin, 1974, p. 61, Chaitin, 1995), in fact the above sketch of the incompleteness proof is due to Chaitin (Chaitin, 1974, pp. 62–63) and can be understood as a more rigorous version of the Berry

4 The discussion of Chaitin's incompleteness results in the context of diagonal arguments is further justified by the universal scheme described in Yanofsky, 2003, which "encompasses the semantic paradoxes, and how they arise as diagonal arguments and fixed point theorems in logic". The scheme is discussed, among other examples, for Cantor's theorem (with its diagonal set), Russell's paradox, Gödel's first incompleteness theorem and Turing's Halting problem. The connection to Chaitin's incompleteness results is explicitly called out: "Many of Chaitin's algorithmic information theory arguments seem to fit our scheme" (Yanofsky, 2003, p. 385).

5 Russell, who attributed it to the Bodleian librarian G. G. Berry and was the first to publish it, gave the following more precise version (Russell, 1908, p. 153):

> The number of syllables in the English names of finite integers tends to increase as the integers grow larger and must gradually increase indefinitely, since only a finite number of names can be made with a given finite number of syllables. Hence the names of some integers must consist of at least nineteen syllables, and among these there must be a least. Hence "the least integer not nameable in fewer than nineteen syllables" must denote a definite integer; in fact, it denotes 111,777. But "the least integer not nameable in fewer than nineteen syllables" is itself a name consisting of eighteen syllables; hence the least integer not nameable in fewer than nineteen syllables can be named in eighteen syllables, which is a contradiction.

The context of the first appearance of this paradox, which is followed in Russell's text by clearly diagonal paradoxes such as Richard's paradox and which motivates Russell's theory of types, emphasises the close connection between Berry's paradox (and subsequently Kolmogorov complexity) and issues that lie at the heart of Wittgenstein's interest in Cantor's, Gödel's and Turing's diagonal arguments.

paradox, which turns the paradox into an incompleteness result, similar to how Gödel's and Turing's fundamental results show analogies with paradoxes such as the liar's paradox (Chaitin, 1975, pp. 22–27, Chaitin, 2002).[6] As Chaitin explains:

> What can the meaning of this paradox be? In the case of Berry's original paradox, one cannot arrive at a meaningful conclusion, inasmuch as one is dealing with vague concepts such as an English phrase's defining a positive integer. However our version of the paradox deals with exact concepts that have been defined mathematically. Therefore, it cannot really be a contradiction. It would be absurd for a string not to have a program of length less than or equal to n for calculating it, and at the same time to have such a program. Thus we arrive at the interesting conclusion that such a string cannot exist. For all sufficiently great values of n, one cannot talk about "the first string that can be proven to be of complexity greater than n," because this string cannot exist. [Chaitin, 1974, p. 63]

While Chaitin's diagonal argument might not be strictly diagonal in the same sense as Cantor's, Gödel's or Turing's argument, it certainly shares a *family resemblance* with these more obviously diagonal proofs that justifies its discussion in light of and in connection with the other arguments. Even in the case of the other three more clearly diagonal arguments it is doubtful as to what could even be called their shared 'essence' if this essence is to be more than the trivial observation that they all apply a diagonal method. A philosophical investigation that restricted itself to the essential characteristics of these different proofs would be unable to examine the use of these arguments in their very heterogeneous contexts and thus fail to describe the philosophical grammar of diagonal proofs with all of their unique facets. Furthermore, the non-diagonal aspects of Chaitin's argument can help to clarify not only its own standing in comparison to Cantor, Gödel and Turing, but also illuminate the aspects that bind these proofs together or set them apart.[7]

---

6 The similarity of the different logical and semantic paradoxes thus helps to link together their incompleteness-proof-counterparts, but also highlights the dividing line between paradoxes and incompleteness proofs. On the contradictions of the paradoxes, Priest, 1995, p. 4 writes: "In each of the cases, there is a totality (of all things expressible, describable, etc.) and an appropriate operation that generates an object that is both within and without the totality. I will call these situations *Closure* and *Transcendence*, respectively." While such a general characterisation of the variety of paradoxes is certainly questionable from the perspective of Wittgenstein, as it risks overlooking the specific uses in favour of an overly broad generalisation, it nevertheless serves as an interesting distinction between Berry's paradox and Chaitin's more rigorous information-theoretic incompleteness result: While in the case of Berry's paradox it is not immediately clear how the transcendence is to be resolved and thus leads to the paradox, the incompleteness result can simply reject the assumption of the decidability of Kolmogorov complexity to exclude the transcendent operation from the system of programs. This is similar to what happens in Gödel's and Turing's proofs, where the contradiction is resolved by excluding some true propositions from the system of provable propositions and by excluding a program from the system of halting-decidable programs, respectively.

7 As Floyd, 2012, p. 36 points out, Wittgenstein's interest in the diagonal argument is not restricted to the special case of the uncountability of the real numbers or

MATHEMATICS AS PHYSICS

What are we to make of such an incompleteness proof from the perspective of Wittgenstein's philosophy? In contrast to Berry's paradox, which we could easily brush aside as a paradoxical source of amusement that is ultimately without any practical use, the undecidability of Kolmogorov complexity seems to demonstrate an inherent limit of many rather ordinary concepts. The proof is relevant even in the context of natural languages if we replace the notion of programs with unambiguous English descriptions and search through all of these English descriptions lexicographically, going from the smallest description to the largest. The result then seems to preclude the possibility of ever being certain that a particular description (whose size exceeds a particular threshold) truly is the most compact description that we could give.

Such an undecidability proof can appear to demonstrate an almost physical impossibility, because it seems that if there were a general decision procedure for Kolmogorov complexity, it would lead to a situation where a less complex program could produce a more complex program, in other words less information would generate more information. This production of information *ex nihilo* would amount to a reduction of entropy, as "program-size complexity is like the idea of entropy in thermodynamics" (Chaitin, 2000, p. 148), with entropy already appearing as a central concept in classical information theory, "as defined in certain formulations of statistical mechanics" and with a direct reference to Boltzmann (Shannon, 1948, p. 393). As the second law of thermodynamics tells us, entropy in a closed system can only ever increase, a decrease is thus a *physical impossibility*. Have we perhaps made an analogous discovery in algorithmic information theory, so that the impossibility of calculating the Kolmogorov complexity for any possible program corresponds to the physical possibility in thermodynamics? Algorithmic information theory seems to suggest as much:

> Ideas from theoretical physics and theoretical computer science are definitely leaking across the traditional boundaries between these two fields. And this holds for AIT too, because its two central concepts are versions of *randomness* and of *entropy*, which are ideas that I took with me from physics and into mathematical logic. [Chaitin, 2004, p. 235]

---

a particular (diagonal) notation or arrangement. Wittgenstein's own variant of the diagonal argument (see Chapter 3) allows us to "imagine an enumeration in any way we like, and Wittgenstein does not restrict its presentation. He is articulating, in other words, a generalized *form* of diagonal argumentation. The argument is thus generally applicable, not only to decimal expansions, but to any purported listing or rule-governed expression of them; it does not rely on any particular notational device or preferred spatial arrangements of signs." While it shall not be suggested that Wittgenstein had arguments such as Chaitin's in mind when formulating his own variant of the diagonal argument, Wittgenstein's broad interest in the family of diagonal arguments can help to justify the application of his remarks in the context of algorithmic information theory.

What Chaitin's incompleteness result seems to prove is that, as far as entropy is concerned, "you can only get out as much as you put in" (Chaitin, 1993, p. 89), in algorithmic information theory as well as in physics. What you cannot do, neither in theory nor in practice, is to create something out of 'thin air'. Chaitin further elaborates this point in the slightly different context of an information-theoretic proof of Gödel's theorem, which is information-theoretically similar enough to apply to the current discussion. As Chaitin notes, the "viewpoint of thermodynamics and statistical mechanics" applies to information theory, it is thus impossible to derive more information than you put in as axioms, with the result that you can never derive a "twenty-pound theorem" from only "ten pounds of axioms":

> Gödel's original proof constructed a paradoxical assertion that is true but not provable within the usual formalizations of number theory. In contrast I would like to measure the power of a set of axioms and rules of inference. I would like to able to say that if one has ten pounds of axioms and a twenty-pound theorem, then that theorem cannot be derived from those axioms. And I will argue that this approach to Gödel's theorem does suggest a change in the daily habits of mathematicians, and that Gödel's theorem cannot be shrugged away.
>
> To be more specific, I will apply the viewpoint of thermodynamics and statistical mechanics to Gödel's theorem, and will use such concepts as probability, randomness, entropy, and information to study the incompleteness phenomenon and to attempt to evaluate how widespread it is. [Chaitin, 1982b, p. 113]

A view that interprets a mathematical proof as something akin to a physical discovery is of course what Wittgenstein would warn against, since we are misled by identical terms (such as "entropy" and "information") in completely different contexts, with the differences in use being obscured by this superficial similarity. It is crucial to remember that we *call* a program minimal or of a particular complexity as the result of a particular definition and that as a result any impossibilities arising in the system are logical impossibilities, rooted in the grammar of our concepts, not in a pseudo-empirical and "ultra-physical" (Ts-221a/b, 143.3, Ts-222, 11.3 / *RFM I* §8, see Section 0.1) world of mathematics. At the contradictory point in the incompleteness proof, where the program $\mathscr{D}$ seemingly would have produced a more complex program (in other words a program with more information) than itself, there is no magical surplus of information popping into existence, there are only two programs (the search program $\mathscr{D}$ and the program resulting from its search) for which the mechanical application of the rules of the system leads to Kolmogorov complexities that contradict our concept of what the relationship between these two programs must be (based on the size of these two programs in bits). The conclusion of Chaitin's incompleteness theorem, which rejects the assumption of a general decision procedure for Kolmogorov complexity, is of course a valid and even natural choice, but it is not the result of some ultra-physical mathematical reality that we are forced to accept.

In the case of physics and thermodynamics, the entropy in a system can be measured (at least approximately) and *then*, based on this measurement, we are able to make predictions about how much energy for useful work remains available in the system. These predictions have proven so accurate and useful that we are certain that a system with maximum entropy has no more energy available for useful work, but we would certainly revise our theory of thermodynamics if it ever turned out that in some peculiar situation entropy in a closed system could be observed to decrease, contrary to the second law of thermodynamics. In contrast, the concepts of entropy and information in (algorithmic) information theory are *defined* in terms of program-size complexity, not as the result of mathematical observations or experiments.

To say that we cannot calculate program-size complexity beyond a particular threshold, because otherwise a sequence of bits with less information would generate a sequence of bits with more information, suggests that generating more information from less information would violate the laws of algorithmic information theory in a similar way as an observation of a decrease of entropy in a closed system would violate the laws of physics. But we need to keep in mind that the minimal program size to generate a particular bit sequence is what we *call* the information contained in this bit sequence. A program that generates a sequence with more information than itself is impossible only because there is nothing that we would be willing to call a program with less information than what it generates, precisely because as soon as a program is shown to generate a sequence with 'more' information we immediately say that this larger sequence has only as much information as the smaller program that acts as a description of it. Sequences are *measured* purely by their smallest program, without any recourse to any external standard of measurement. To say that we have found a program that generates a bit sequence with more information than the program itself would be like saying that we have found a 1 metre stick that is longer than 1 metre. It is *logically* impossible, because the stick itself is the standard of measurement. In the case of physics the observation of a decrease of entropy in a closed system would be an astounding discovery, but in the case of algorithmic information theory all that a logical impossibility shows is what we call a correct measurement.

It is not surprising that if the rules of the game allow the construction of a program that generates a sequence which it, by definition, cannot generate, the resulting program will be paradoxical. The rules of the game lead to a program of the form 'Do What You Do Not Do' (cf. Section 3.1) and we can either exclude such a program from our game (by rejecting the assumption that a decidable algorithm for Kolmogorov complexity exists and thereby ensure that no program of the form 'Do What You Do Not Do' can be constructed) or accept that the resulting program must be paradoxical. It is the old choice between incompleteness and inconsistency already offered by Gödel's theorem and it is understandable that mathematics, with its ideal of consistency, chooses incompleteness over inconsistency. But the incompleteness result should not blind us to the fact that it is a choice, that the incompleteness proof leads us to make a conceptual decision at the crucial point of the argument, but that it is not an inevitable discovery, because we could also have chosen to play another (inconsistent) game instead.

The fact that incompleteness in the form of the undecidability of Kolmogorov complexity is ordinarily accepted as the inevitable conclusion of such an argument points to another interesting issue at play here. Why were we interested in a decision procedure for Kolmogorov complexity in the first place? What *practical* use did we originally have in mind when we set out to find such an algorithm? The ease with which we give up the notion of such a decision procedure suggests that we had no clear concept in the first place and that the incompleteness result offers us absolution from any attempt to clarify this concept. It would be easy to object that the practical use would have come in the form of a universal compression scheme that guaranteed for every sequence of data (no matter of what origin) a transmission of the information content in a minimal amount of bits. After all, nobody could conceivably dispute the usefulness of finding the most general description of any data.

But here we risk being blinded by a utopian vision that had no chance of being realisable, for *practical* reasons, not for theoretical reasons as the incompleteness result suggests. Surely if our aim was to transmit descriptions as efficiently as possible, the inconsistency in the case of paradoxical descriptions would have been a low price to pay *in practice*. After all, how often do we transmit paradoxical descriptions of the form "the number not nameable in under ten words"? Any practical machine has certain defects, which we mitigate through redundancy and safety measures, and we could conceivably have done something similar in the case of the decision procedure for Kolmogorov complexity.

Instead, the real reason why the universal compressor would have been doomed in any practical setting is much simpler and has practical reasons, which we can see if we consider the following program: "If the *unsolved mathematical problem X* can be proved and is true, print 0 as the output, otherwise print 1." In the past, the canonical example for such a problem would have been Fermat's last theorem, which by now has been proved, but it does not matter which problem is chosen as long as it is extraordinarily hard to solve. It is clear that the minimal description of the program will be either "print 0" or "print 1", if the provability of the problem X can be decided at all, and for the sake of the argument we assume that the problem indeed turns out to be provable. Our universal compressor would then have to solve an extraordinarily difficult mathematical problem and it is obvious that by feeding the compressor only programs of this sort we effectively turn it into a theorem prover of arbitrarily difficult theorems. At this point Gödel's theorem looms again, but we will sidestep it here and instead focus on decidable mathematical problems that simply take a very long time to solve, long past the eventual heat death of the universe, for example the calculation of arbitrarily large prime numbers or of incredibly long running computations like the Ackermann function. Would we really call the universal compressor a practical compressor if it routinely failed to calculate the minimal program size in our own lifetime or even the lifetime of any imaginable civilisation? In the end, there is no practical difference between an algorithm that theoretically will eventually come to a halt but practically runs forever and one which in many cases does not even theoretically halt.

The incompleteness result is thus not an end to any practical search for a compression scheme, because a search that relies on a *general* compressor in the sense of algorithmic information theory would have been impractical even in the absence of any incompleteness result. That a decision procedure for Kolmogorov complexity is nothing else than a universal theorem prover in different clothing (and that consequently Gödel's and Turing's results lurk behind every corner) demonstrates that the concept of a universal compression scheme is as vague *in its applications* as Hilbert's program and the search for a formalisation of all of mathematics were for mathematics. The idea that such an incompleteness result could have blocked any avenue for mathematical or logical progress suggests that mathematicians surprised by these results expected certain theoretical endeavours to do much more than they possibly could have achieved, namely to clarify a concept (with practical uses) where there actually was none.

It must be pointed out that Chaitin himself has a remarkably different view of the implications of these incompleteness results than many other mathematicians and clearly articulates an interpretation that shows parallels with many ideas held by Wittgenstein, even though the wording is often quite different. The most obvious parallel is that, in the eyes of Chaitin at least, incompleteness results such as Gödel's do not give cause for concern or even depression, because they are not a hindrance to any practical mathematical work, even though they certainly came as a shock to the foundational efforts in mathematics:

> In 1946 Hermann Weyl said that the doubt induced by such discoveries as Gödel's theorem had been "a constant drain on the enthusiasm and determination with which I pursued my research work." From the point of view of information theory, however, Gödel's theorem does not appear to give cause for depression. Instead it seems simply to suggest that in order to

> progress, mathematicians, like investigators in other sciences, must search for new axioms. [Chaitin, 1975, p. 24]

Of course such an extreme view as exemplified by Weyl is not common among mathematicians anymore and Chaitin says as much himself, while going one step further by advocating for a more relaxed attitude towards the inclusion of new axioms into mathematics:

> How have the incompleteness theorem of Gödel, the halting problem of Turing and my own work affected mathematics? The fact is that most mathematicians have shrugged off the results. Of course, they agree in principle that any finite set of axioms is incomplete, but in practice they dismiss the fact as not applying directly to their work. Unfortunately, however, it may sometimes apply. Although Gödel's original theorem seemed to apply only to unusual mathematical propositions that were not likely to be of interest in practice, algorithmic information theory has shown that incompleteness and randomness are natural and pervasive. This suggests to me that the possibility of searching for new axioms applying to the whole numbers should perhaps be taken more seriously.
>
> Indeed, the fact that many mathematical problems have remained unsolved for hundreds and even thousands of years tends to support my contention. Mathematicians steadfastly assume that the failure to solve these problems lies strictly within themselves, but could the fault not lie in the incompleteness of their axioms? For example, the question of whether there are any perfect odd numbers has defied an answer since the time of the ancient Greeks. [...] Could it be that the statement "There are no odd perfect numbers" is unprovable? If it is, perhaps mathematicians had better accept it as an axiom. [Chaitin, 1988, p. 37]

But when could mathematicians feel justified in accepting axioms? Are they to accept them simply because these axioms have proven to be useful, have stood the test of time and turned out to be too difficult to prove? Such a rather practical view is bound to clash with the platonistic conception held by most mathematicians and notably differs from Gödel's own interpretation of the incompleteness result, who viewed mathematics as *a priori* and remained a staunch platonist all his life. It is important to note that Gödel nevertheless came to accept axioms based on "their fruitfulness in mathematics and, one may add, possibly also in physics" (Chaitin, 2003, pp. 193–94, quoting Gödel). Chaitin goes even further than Gödel in this regard and is ready to accept a more nuanced view that is not strictly platonistic:

> I think that the work I've described, and in particular my own work on randomness, has not spared the whole numbers. I always believed, I think most mathematicians probably do, in a kind of Platonic universe. "Does a diophantine equation have an infinite number of solutions or a finite number?" This question has very little concrete computational meaning, but I certainly used to believe in my heart, that even if we will never find out, God knew, and either there were a finite number of solutions or an infinite number of solutions. It was black or white in the Platonic universe of mathematical reality. It was one way or the other.
>
> I think that my work makes things look gray, and that mathematicians are joining the company of their theoretical physics colleagues. I don't think that this is necessarily bad. We've seen that in classical and quantum physics randomness and unpredictability are fundamental. I believe

> that these concepts are also found at the very heart of pure mathematics. [Chaitin, 1989, p. 514]

In Chaitin's view, algorithmic information theory gives rise to a conception of mathematics that Gödel only alluded to, a "pseudo-empirical, or quasi-empirical position" (Chaitin, 2003, p. 195). This is a direct consequence of the result that a 'twenty-pound theorem' needs at least 'twenty pounds of axioms', which implies that not all theorems can be derived from a fixed set of immediately intuitive and simple laws of thought. Chaitin thus takes another step towards a position that accepts axioms as at least partially invented and judges them based on their usefulness to derive theorems with manifold applications in mathematics, a view which seems to echo Wittgenstein's own notion of mathematics as being primarily invented and not discovered.

Of course the similarities between the views of Chaitin and Wittgenstein should not be overstated. Chaitin distinguishes himself from many other mathematicians by an openness for a less dogmatic and non-platonistic view of mathematics, an openness with which Wittgenstein would have certainly agreed, but the idea that as a consequence of information-theoretic incompleteness results the axioms of mathematics should be treated 'quasi-empirically' like hypotheses in physics is at odds with Wittgenstein's strict separation between experiment and calculation and thus between empirical propositions and grammatical propositions. To say that axioms are grounded empirically, even quasi-empirically, would risk misunderstanding their role as rules in a language game that lie beyond doubt and uncertainty. Chaitin's approach seems to provide a new source of certainty and escape from the doubt raised by Gödel's results, whereas Wittgenstein wants to emphasise that mathematical propositions operate outside of the game of verification which empirical propositions are subjected to. Quite ironically, Chaitin less dogmatic view opens the door to another form of dogmatism, which understands mathematical theorems less as discoveries in the platonic world of mathematics and more as discoveries similar to physics. In so far as 'quasi-empiricism' forms a distinct philosophical position and understands at least parts of mathematics as *a posteriori* science, Wittgenstein would have found it problematic for the same reasons that theses and dogmatism in general are to be examined with suspicion in philosophy.

Chaitin tends to combine a conception of mathematics grounded purely by *a priori* deductive reasoning with a conception grounded by *a posteriori* inductive science (cf. Chaitin, 2003, p. 196), but this merely extends the assumption that mathematics is directly and explicitly grounded in abstract principles that we can articulate, whereas Wittgenstein sees mathematics as being implicitly grounded in our *acts* ("So handle ich eben.", Ts-227a, 153.2 /*PI* §217). Chaitin blurs the dividing line between mathematics and natural sciences, while Wittgenstein wants to do just the opposite and emphasises that mathematics has no need for the empirical certainty provided by physical laws, because mathematical certainty shows itself in our grammar. Despite these differences, Chaitin's work provides interesting philosophical 'raw material' for a Wittgensteinian investigation of a variety of interesting concepts.

INCONSISTENT COMPLEXITY

Philosophically, the proof of the undecidability of Kolmogorov complexity outlined above raises questions similar to those investigated by Wittgenstein in the context of Cantor's diagonal argument. The computationalist view at the heart of the incompleteness results in algorithmic information theory, which treats every possible description uniformly as a program with a single particular output (or no output in the case of programs that do not halt), proceeds analogously to the extensionalist view at the heart of Cantor's diagonal argument, which treats every number purely extensionally as an infinite decimal expansion. Wittgenstein's critique in *RFM II* ("what is the method of calculating, and what the result, here? You will say that they are *one*", Ms-117, 99 / *RFM II* §3, see Section 1.2) applies just as much in the context of algorithmic information theory: Why are we willing to say that a formalised version of Berry's paradox as a program is a valid description? In contrast to other programs, for which we can distinguish between the result (the program output in form of a binary sequence) and the method (the particular algorithm that computes the binary sequence), the paradoxical program constructed based on the assumption of a decision procedure for Kolmogorov complexity depends almost 'parasitically' on the particular computing model, in the same way that the diagonalised number in Cantor's diagonal argument depends on the base of the numbering system. In the same way that the extensionalist view of mathematics presents a uniform picture of numbers and risks obscuring the variety of numerical methods exhibited by a more intensionalist view, the computationalist view presents an equally uniform picture of descriptions in terms of programs.

The formalised version of the Berry paradox is a program that 'escapes' any attempt to assign it a fixed and correct Kolmogorov complexity via the assumed decision procedure. It is tempting to say that this construction produces a program that is different from all of the programs for which the Kolmogorov complexity is decidable, as Chaitin does, but such a conclusion conceals the conceptual decision at the heart of the proof, as Wittgenstein points out in the context of his remarks on Cantor:

> 'Ich will Dich eine Methode lehren wie Du in einer Entwicklung allen diesen Entwicklungen nach der Reihe *ausweichen* kannst.' So eine Meth-

> ode ist das Diagonalverfahren. – "Also erzeugt sie eine Reihe, die von allen diesen verschieden ist." Ist das richtig? – Ja; wenn Du nämlich diese Worte auf diesen, oben beschriebenen Fall anwenden willst. [Ms-117, 101.2 / *BGM II* §8]
>
> ---
>
> "I want to shew you a method by which you can serially *avoid* all these developments." The diagonal procedure is such a method. – "So it produces a series that is different from all of these." Is that right? – Yes; if, that is, you want to apply these words to the described case. [*RFM II* §8]

By choosing to apply these words in this case, we not only choose to call the formalised Berry paradox a program, but also choose to draw a line between the undecidability and decidability of Kolmogorov complexity based on whether or not the result of the algorithm can be consistent. The first decision, to call the paradoxical program a program, will not be discussed here, since any possible objection would at most lead to a different sort of incompleteness result, by excluding paradoxical descriptions from all those descriptions that are considered to be programs.[8] The more interesting decision is the one which draws the dividing line between the decidable and the undecidable: Why not conclude from the proof that the notion of Kolmogorov complexity is inconsistent and that, if we allow the construction of arbitrary programs, formalised versions of the Berry paradox can be shown to have a contradictory Kolmogorov complexity that can only be expressed as a pair of program lengths, namely the length of the formalised Berry paradox in combination with the length of the first program found as a result of the systematic search through the programs?

The issue at stake is the "therefore", exactly as in the case of Cantor's diagonal argument (Ms-117, 102.2 / §10). Chaitin's incompleteness result appears to demonstrate that the formalised Berry paradox will lead to a contradictory answer and that *therefore* there cannot be a decision procedure for Kolmogorov complexity. The "therefore" seems to *force* us along a chain of steps in the proof and gives the proof the appearance of a *discovery* of a mathematical fact without any alternative. But even purely *invented* rules can force us to do or say something, as Wittgenstein points out in Ms-124:

> "Die Regeln zwingen mich zu etwas", nun das kann man schon sagen, weil, was mir mit der Regel übereinzustimmen scheint ja nicht von meiner Willkür abhängt. [...]
>
> Was zwingt mich denn? – Der Ausdruck der Regel? – Ja; wenn ich einmal so erzogen bin. Aber kann ich sagen, er zwingt mich, ihm zu folgen? Ja; wenn man sich hier die Regel nicht als Linie denkt, der ich nachfahre, sondern als Zauberspruch der uns im Bann hält.

---

8 Furthermore, the importance of Chaitin's results lies precisely in the fact that the 'homespun' nature of Turing's computing model (or any other equivalent formalism) seems to make it impossible not to consider even a paradoxical program to be a proper program, contrary to the case of the informal Berry paradox, which we can more easily exclude from our legitimate language and call it "an illegitimate notion" (Russell, 1908, p. 155).

> [“schlichter Unsinn, & Beulen . . . ”] [Ms-124, 105.2, 107.3 / *BGM VII* §27]
>
> ---
>
> “The rules compel me to. . . ” – this can be said if only for the reason that it is not all a matter of my own will what seems to me to agree with the rule. [...]
>
> What is it that compels me? – the expression of the rule? – Yes, once I have been educated in this way. But can I say it compels me to follow it? Yes: if here one thinks of the rule, not as a line that I trace, but rather as a spell that holds us in thrall.
>
> ((“plain nonsense, and bumps. . . ”)) [*RFM VII* §27]

The alternative to an interpretation of the incompleteness result as a discovery forced upon us is *not* a naive form of ‘anything goes’, but rather the notion that even though we are forced by a particular form of life, the conceptual decision could end up being different in the context of a different form of life. The misleading aspect of the “therefore” is thus not the emphasis on the ‘force’ of the argument, but the inability to see *why* we are forced to follow the proof the way we do: not because some “ultra-physical” reality would dictate it, but rather because this way of following the proof the way we do is embedded in a whole form of life. This contrast between “it *must* follow” and “it *follows*” was already pointed out by Wittgenstein in a different context in Ms-117:

> “Aus ‘alle’, wenn es so gemeint ist, muß doch *das* folgen.” – Wenn es *wie* gemeint ist? Überlege es Dir, wie meinst Du es? Da schwebt Dir etwa noch ein Bild vor – und mehr hast Du nicht. – Nein, es *muß* nicht – aber es *folgt*: Wir *vollziehen* diesen Übergang.
>
> Und wir sagen: Wenn das nicht folgt, dann waren es eben nicht *alle*! – – und das zeigt nur, wie wir mit Worten in so einer Situation reagieren. – [Ms-117, 1.1 (p. 13–14), Ts-221a/b, 147.2, Ts-222, 15.1 / *BGM I* §12]
>
> ---
>
> “From ‘*all*’, if it is meant *like this*’, *this* must surely follow!” – If it is meant like *what*? Consider how you mean it. Here perhaps a further picture comes to your mind – and that is all you have got. – No, it is not true that it *must* – but it *does* follow: we *perform* this transition.
>
> And we say: If this does not follow, then it simply wouldn’t be *all* – and that only shews how we react with words in such a situation. [*RFM I* §12]

Here we see one of the reasons why Wittgenstein’s philosophy of mathematics has so often been dismissed: It is all too easy to misinterpret Wittgenstein’s remarks as a rather trivial form of constructivistic ‘anything goes’ that completely misunderstands why and how proofs force us to say that something follows. But Wittgenstein does not want to deny that we are forced to say that $2 + 2 = 4$ in a way that does not apply to $2 + 2 = 5$, he merely wants to emphasise that *it follows*, not that “therefore, it follows”. The proof was already perfectly fine without the additional “therefore” and the extra word merely feigns additional certainty when in fact the question of certainty does not even enter into the language game, because the certainty shows itself in our acts. Wittgenstein thus walks a very fine line, which at times narrows down to a seemingly innocuous word like “therefore”.

The remarks on Cantor that could be applied in the present context extend well into Ms-121 and are too numerous to be discussed at length (see Section 1.3 for more context). It shall only be pointed out that in both contexts the comparability of rules that 'stand on their own' with rules that depend on the entire 'system of systems' is called into question by Wittgenstein:

> Warum sollten wir nicht sagen: die Regel, die Diagonale zu verändern, sei mit den Regeln des Systems *unvergleichbar*?
> "tamper with the extension" [Ms-121, 41r.3]

> Why should we not say: the rule of changing the diagonal is *incomparable* with the rules of the system?
> "tamper with the extension"

The formalised Berry paradox might be considered a program, but it is distinguished from other programs by its lack of practical use. Its sole purpose is 'parasitic' and depends so much on the context of the incompleteness proof that the program is unthinkable outside of it, because it cannot 'survive' without the assumption of a decision procedure for Kolmogorov complexity that is discarded at the end of the proof. It is not "anchored" in any practice, but only "a piece of mathematical architecture which hangs in the air, [...] but not supported by anything and supporting nothing" (Ms-121, 41v.2 / §35).

But if there is no practice to anchor the contradictory program, that is to say no *extra-mathematical* use for it, the rejection of the initial assumption due to the resulting contradiction becomes philosophically questionable or at the very least open to a further investigation. A deductive principle such as *reductio ad absurdum* is used in practice because we have no use for contradictory premises in our ordinary language games, where the premises are ordinary observations about the world such as "It is raining". But an inherently paradoxical description such as Berry's paradox is obviously of a very different kind and can even be distinguished clearly from ordinary programs by virtue of being 'higher order' and depending on the whole system of programs. It is thus not immediately obvious why we have to reject the resulting contradiction *in this particular case*. This is not an invitation for trivialism, nor the advocacy for a wholesale dismissal of the law of non-contradiction, only an emphasis on the possibility of accepting the contradiction "in tiny doses":

> Wir sagen, der Widerspruch würde den Kalkül *vernichten*. Aber wenn er nun sozusagen in winzigen Dosen aufträte, gleichsam blitzweise, nicht als ein ständiges Rechenmittel, würde er da {das Spiel // den Kalkül} auch vernichten? [Ms-124, 65.5 / *BGM VII* §15]

> We say that the contradiction would '*destroy* the calculus. But suppose it only occurred in tiny doses in lightning flashes as it were, not as a constant instrument of calculation, would it nullify the calculus? [*RFM VII* §15]

In Ms-178d, a short fragment of only a few loose pages[9], Wittgenstein goes even further and questions the rejection of contradictions *if the function of the contradictory propositions is not yet determined*:

> 'Aus seinem Gegenteil läßt sich ein Widerspruch ableiten.' – Nun, vielleicht macht er hier nichts.
>
> Den Widerspruch zu vermeiden ist eine mathematische Methode. Sie führt zu brauchbaren {Gebilden // Sätzen} & brauchbar ist hier ähnlich unbestimmt wie eine Pointe haben.
> Ist aber die Funktion eines {Satzes // irgendwie satzähnlichen Gebildes} gänzlich unbestimmt, warum soll er nicht ein Widerspruch sein? Warum sollte sich ein Mathematiker prinzipiell vor {*jedem* // dem} Widerspruch bekreuzigen. (Man {könnte // möchte} sagen: hab keine Angst er beißt nicht!) [Ms-178d, 1.2–1.3]

> 'A contradiction can be derived from its opposite.' - Well, maybe it doesn't cause any trouble here.
>
> Avoiding the contradiction is a mathematical method. It leads to usable {constructions // propositions} & usable here is similarly undetermined as having a punch line.
> But if the function of a {proposition // proposition-like construction} is completely undetermined, why should it not be a contradiction? Why should a mathematician in principle cross himself before {*every* // the} contradiction. (One could say: don't be afraid, it doesn't bite!)

This freedom to accept the contradiction instead of excluding it is a consequence of the lack of a role that the contradictory proposition plays in our ordinary language. Its use is undetermined, as it is not yet used in our practice and "supports nothing" that could be toppled by our decision to accept or exclude the contradiction. The same cannot be said for a contradiction such as "It is raining and it is not raining", an acceptance of which would require us to question and adjust a whole form of life with all of its associated practices.[10] This is not to say that to conclude the undecidability from the incompleteness proof sketched above would be somehow wrong, only to suggest that it is easier than in most other situations to imagine a form of life that calls into question the application of the *reduction ad absurdum* and as a result considers such an incompleteness proof as a conceptual invention rather than as a discovery.

THE HALTING PROBABILITY Ω

Up until now, Kolmogorov complexity has been examined in the present discussion as a concept independent of Turing's halting problem. The approach of introducing Kolmogorov complexity on its own

9 See Section 1.4 and Section 3.4 for a discussion of the context of the remark in Ms-178d as well as an attempt to properly date the fragment in relation to Ms-121.

10 Ms-121, 70r.2 is relevant in this context and shows further parallels with Wittgenstein's remarks on Cantor: "Ich möchte das rechtfertigen indem ich sage: Es ist eben hier *alles* anders, ich bin nicht mehr [...] *gezwungen* dies so zu nennen." See Section 1.5 for a detailed discussion.

has certain benefits, as it allows for an investigation that targets aspects unique to algorithmic information theory without bringing in all the heavy machinery of Turing's own reflections on questions of algorithmic decidability (see Chapter 3). To quietly pass over the close connection between the halting problem and information-theoretic complexity would however paint a distorted picture of this field and ignore some of the most interesting philosophical interpretations of algorithmic information theory, which is why the following sections will examine this connection in more detail, starting with "Chaitin's mystical, magical number, Ω" (Cover and Thomas, 2006, p. 484).

Before looking at the number Ω itself, it helps to reframe the proof of the undecidability of Kolmogorov complexity in terms of the undecidability of the halting problem. The decidability of Kolmogorov complexity is a sufficient and necessary condition for the decidability of the halting problem, in other words either one can be reduced to the other, since an algorithm for one can be used to construct an algorithm for the other. This can be demonstrated by showing that a decision procedure for the halting problem implies a decision procedure for Kolmogorov complexity and vice versa:

1. Halting decidability implies Kolmogorov decidability: Let us assume that there is an algorithm for deciding the halting problem. Then the Kolmogorov complexity of a particular program P with length n can be decided by checking all programs with length up to and including length n, executing only those programs that halt. The length of the shortest program that produces the same output as P must be the Kolmogorov complexity of P.

2. Kolmogorov decidability implies halting decidability: Let us assume that there is an algorithm for deciding the Kolmogorov complexity of any particular program. Then we can decide whether a particular program P halts by constructing a program P' that first runs P and then runs another program X for which the Kolmogorov complexity is known to be different from that of a non-halting program and returns the result of the program X as the output of P'. By deciding the Kolmogorov complexity of P' we can then decide whether P halts, because if P does not halt then X will never run and thus the Kolmogorov complexity of P' will be equal to the Kolmogorov complexity of P, but if P halts then the output (and thus the Kolmogorov complexity) of P' will be equal to the output (and thus the Kolmogorov complexity) of X.[11]

11 The choice of X is static but depends on how we define the Kolmogorov complexity of a non-halting program: If it is defined as the length of the smallest program that does not halt, X is chosen to be the first halting program larger than the smallest non-halting program. If it is defined as the length of the particular non-halting program

There are many other ways to prove the second part of this equivalence (see Chaitin, Arslanov, and Calude, 1995).[12] Information-theoretic complexity and the halting problem are thus intimately linked, which is further emphasised by Chaitin's most famous concept, the number Ω. This number can be understood as a generalisation of the halting decision, which instead of considering whether a particular program halts considers the probability that an arbitrary program, chosen at random, will halt (in the context of a chosen computing model, for example a universal Turing machine or a particular Turing-complete programming language). As Chaitin explains:

> What exactly is the halting probability? I've written down an expression for it:
>
> $$\Omega = \sum_{p \text{ halts}} 2^{-|p|}$$
>
> Instead of looking at individual programs and asking whether they halt, you put all computer programs together in a bag. If you generate a computer program at random by tossing a coin for each bit of the program, what is the chance that the program will halt? You're thinking of programs as bit strings, and you generate each bit by an independent toss of a fair coin, so if a program is N bits long, then the probability that you get that particular program is 2^{-N}. Any program p that halts contributes $2^{-|p|}$, two to the minus its size in bits, the number of bits in it, to this halting probability. [Chaitin, 1993, p. 85]

A crucial property that must hold for the computing model is that all programs are "self-delimiting", which means that if a particular bit sequence (such as for example '011') is a valid program then no extension of this bit sequence can be a valid program (so that for the example of the valid program '011' neither '0111', '0110' nor '01101' can be valid programs). This is necessary to ensure that Ω never grows larger than 1, because if for example '00', '01', '10' and '11' are all valid programs then each of them already contributes 0.25 to the halting probability (for a maximum total of $4 \times 0.25 = 1$ if all 4 programs

itself so that different non-halting programs have different complexities, X is chosen to be the first program with a complexity smaller than its own length. If it is defined to be 0, X is chosen to be the first program with length greater than 0, etc.

What matters is that for any imaginable definition of the Kolmogorov complexity we can always pick at least one program X which will be distinguishable by its Kolmogorov complexity from all non-halting programs and the Kolmogorov complexity decision procedure can then be used to distinguish the result of P' based on whether or not the program proceeds past P to X.

12 The above sketch of the second part is an application of Rice's theorem that "all non-trivial properties of the recursively enumerable languages are undecidable" (Hopcroft, Motwani, and Ullman, 2001, pp. 387–90). The program P' is only a specific instance (for Kolmogorov complexity) of Rice's more general theorem, because for every algorithm that decides a non-trivial property it is possible to construct a program similar to P' which first runs a (possibly non-halting) program P and then a fixed program X for which the non-trivial property is known to be different from all non-halting programs. Any non-trivial decision procedure could thus be used to decide the halting problem for arbitrary P.

halt) and longer programs could push the number Ω beyond 1 if programs were allowed to be non-self-delimiting (so that extensions of valid programs could potentially also be valid programs).[13]

Not all models of computation are naturally self-delimiting, for example Turing machines that interpret two adjacent empty squares on the tape as an end marker would not satisfy the condition. But since all Turing-complete models of computation are equivalent via the Church-Turing thesis, a computing model with self-delimiting programs can be assumed as a precondition for Ω with no loss of generality. The exact numerical value of Ω will of course vary based on the particular computing model being used (it is thus more precise to write Ω with a subscript as Ω_C for the halting probability in the computing model C), similar to how the Kolmogorov complexity will vary based on the particular computing model that is being considered, but is a constant in the context of a particular computing model (and often written without subscript simply as Ω since all Turing-complete computing models are theoretically equivalent).

One of the most interesting properties of Ω is that it acts as a "philosopher's stone" and that a knowledge of its first n bits would enable a computer to find all possible proofs for theorems that can be expressed in less than n bits:

> Ω is a "philosopher's stone". Knowing Ω to an accuracy of n bits will enable us to decide the truth of any provable or finitely refutable mathematical theorem that can be written in less than n bits. Actually, all that this means is that given n bits of Ω, there is an effective procedure to decide the truth of n-bit theorems; the procedure may take an arbitrarily long (but finite) time. Of course, without knowing Ω, it is not possible to check the truth or falsity of every theorem by an effective procedure (Gödel's incompleteness theorem). [Cover and Thomas, 2006, pp. 484–85]

Given a knowledge of the first n bits of Ω, we can start to run all programs with a length of less than n bits in an interleaved fashion, executing each program for a single step before switching to the next program, so that more and more programs come to a halt the longer this interleaved execution is allowed to run. The contributions to the halting probability of each halting program are summed (so that a halting 1-bit program contributes 2^{-1}, a halting 2-bit program 2^{-2}, etc.) until the resulting sum equals Ω with its first n bits. All programs that have not halted at this point will never halt, because the contribution to the halting probability after the first n bits of Ω can only have come from programs with a program size equal to or larger than n:

13 As a result of the condition that all programs must be self-delimiting, the halting probability can be understood as the result of repeatedly tossing a coin *until a valid program is generated*, so that if both '0' and '1' were valid programs only a single toin coss would ever be made to generate one of two valid programs, each contributing 0.5 to the halting probability. A thorough explanation of the need for self-delimiting programs can be found in Chaitin, 2008, p. 271.

> Thus, we will ultimately know whether or not any program of less than n bits will halt. This enables the computer to find any proof of the theorem or a counterexample to the theorem if the theorem can be stated in less than n bits. Knowledge of Ω turns previously unprovable theorems into provable theorems. Here Ω acts as an oracle. [Cover and Thomas, 2006, p. 485]

This "oracle" allows us to solve any mathematical question that we can state in less than n bits. To prove whether a particular property about numbers holds for all natural numbers, for example, it suffices to formalise this property in x bits and then write a program that checks all numbers one by one in a loop and halts if any number does not have this property, which adds an additional y bits for the loop-and-check routine. To know whether the property holds for all numbers is then equivalent to the knowledge of whether the program ever halts with a counterexample or runs forever, which can be answered if we know at least the $x + y + 1$ first bits of Ω. Many extremely difficult or still unsolved mathematical problems can be specified in this way, such as Fermat's last theorem or the Goldbach conjecture (cf. Cover and Thomas, 2006, p. 486).

This interpretation of Ω as the ultimate source of mathematical truth can be found in a variety of texts by different authors. It has been called "God's number" (Chown, 2007, p. 321), "a mysterious number Ω ('the secret number', 'the magic number', 'the number of wisdom', 'the number that can be known of but not known'), [...] that encodes very compactly any 'cornerstones' of undecidability" (Rozenberg and Salomaa, 2007, p. 178), a number with "magic bits" (Rozenberg and Salomaa, 2007, p. 208), "Chaitin's mystery number Ω" (Li and Vitányi, 2019, p. 16), "truly the number of Wisdom" (Li and Vitányi, 2019, p. 229), a number whose digits encode "the secret of the universe" (Chown, 2007, p. 330) and it has even been suggested "that knowledge of Omega could be used to characterise the level of development of human civilisation" (Chown, 2007, p. 331). The list of authors includes Chaitin himself, who, borrowing from Bennett, calls the number "a suitable subject for worship by mystical cultists":

> Ω is a suitable subject for worship by mystical cultists, for as Charles Bennett (Gardner, 1979) has argued persuasively, in a sense Ω contains all constructive mathematical truth, and expresses it as concisely and compactly as possible. Knowing the numerical value of Ω with N bits of precision, that is to say, knowing the first N bits of Ω's base-two expansion, is another N-bit axiom that permits one to deduce precisely which programs of size less than N halt and which ones do not. [Chaitin, 1982b, p. 122]

It is worth quoting the passage from Bennett in full, as it is by far the most exuberant and poetic description of Ω:

> Throughout history mystics and philosophers have sought a compact key to universal wisdom, a finite formula or text that would provide the answer to every question. The use of the Bible, the Koran and the I Ching for divination and the tradition of the secret books of Hermes Trismegistus and the medieval Jewish Cabala exemplify this belief or hope. Such sources of universal wisdom are traditionally protected from casual use by

> being difficult to find as well as difficult to understand and dangerous to use, tending to answer more questions and deeper ones than the searcher wishes to ask. The esoteric book is, like God, simple but undescribable. It is omniscient, and it transforms all who know it. The use of classical texts to foretell mundane events is considered superstition nowadays, yet in another sense science is in quest of its own Cabala, a concise set of natural laws that would explain all phenomena. In mathematics, where no set of axioms can hope to prove all true statements, the goal might be a concise axiomatization of all 'interesting' true statements.
>
> Ω is in many senses a Cabalistic number. It can be known of through human reason, but not known. To know it in detail one must accept its uncomputable sequence of digits on faith, like words of a sacred text. The number embodies an enormous amount of wisdom in a very small space inasmuch as its first few thousand digits, which could be written on a small piece of paper, contain the answers to more mathematical questions than could be written down in the entire universe – among them all interesting finitely refutable conjectures. [Gardner, 1979, pp. 33–34; a very similar version can be found in Bennett, 1979, pp. 9–10]

Of course this prosaic description of a mathematical concept should be taken with a grain of salt and not necessarily read as an endorsement of any mystical cultism on the part of Bennett himself, nevertheless it is certainly reflective of a more widespread fascination with Ω as the ultimate yet ungraspable source of mathematical knowledge and truth. Bennett readily admits that it can only be "known of through human reason, but now known", yet exactly this explicit mention of "human reason" appears to open the backdoor to a less rational and more mystical understanding of Ω, since it immediately raises a follow-up question: If it cannot be known through human reason, can it be known through other means? Does God know Ω, we might ask? Bennett does not explicitly raise this question, let alone answer it, but his phrasing certainly suggests that *there is* mathematical knowledge or truth that transcends human reason and that it makes sense to speak of it, even if it is not possible to ever articulate such a knowledge. This knowledge would allow us to answer all "'interesting' true statements", meaning all statements which can be formalised using a certain 'complexity budget', in other words in less than a particular number of bits (Bennett, 1979, p. 7). Bennett continues:

> The wisdom of Ω is useless precisely because it is universal: the only known way of extracting the solution to one halting problem, say the Fermat conjecture, from Ω is by embarking on a vast computation that would at the same time yield solutions to all other simply stated halting problems, a computation far too large to be actually carried out. Ironically, however, although Ω cannot be computed, it might be generated accidentally by a random process, such as a series of coin tosses or an avalanche that left its digits spelled out in the pattern of boulders on a mountainside. The first few digits of Ω are probably already recorded somewhere in the universe. No mortal discoverer of this treasure, however, could verify its authenticity or make practical use of it. [Gardner, 1979, p. 34]

Bennett thus reaffirms a view of Ω as existing independently of our ability to practically use or even theoretically compute it as a number.

Despite this limitation, it is 'out there', either already "recorded" or yet to be "generated accidentally by a random process" somewhere in the universe. But of course this knowledge will not help us find Ω and all except its first few digits must remain unknown and inaccessible to mathematics. The number Ω with all its wisdom is thus written somewhere in the book of nature, but we do not know where to look, not unlike the situation described in Jorge Luis Borges' *Library of Babel*:

> These examples made it possible for a librarian of genius to discover the fundamental law of the Library. This thinker observed that all the books, no matter how diverse they might be, are made up of the same elements: the space, the period, the comma, the twenty-two letters of the alphabet. He also alleged a fact which travelers have confirmed: *In the vast Library there are no two identical books.* From these two incontrovertible premises he deduced that the Library is total and that its shelves register all the possible combinations of the twenty-odd orthographical symbols (a number which, though extremely vast, is not infinite): in other words, all that it is given to express, in all languages. Everything: the minutely detailed history of the future, the archangels' autobiographies, the faithful catalogue of the Library, thousands and thousands of false catalogues, the demonstration of the fallacy of those catalogues, the demonstration of the fallacy of the true catalogue, the Gnostic gospel of Basilides, the commentary on that gospel, the commentary on the commentary on that gospel, the true story of your death, the translation of every book in all languages, the interpolations of every book in all books.
>
> When it was proclaimed that the Library contained all books, the first impression was one of extravagant happiness. All men felt themselves to be the masters of an intact and secret treasure. There was no personal or world problem whose eloquent solution did not exist in some hexagon. [...] At that time it was also hoped that a clarification of humanity's basic mysteries – the origin of the Library and of time – might be found. [Borges, 1964, pp. 64–65]

The solution to any problem, as long as it is expressible, exists *somewhere* in this vast library of books, exactly like how the solution to any "finitely refutable mathematical conjecture" (Bennett, 1979, p. 6) exists (as the binary digits of the real number Ω) *somewhere* among all imaginable decimal expansions of the real numbers. But what good is this book of nature, if there is no way to find the right page? Why not simply 'roll the dice' instead? In the *Library of Babel*, this is considered blasphemy:

> As was natural, this inordinate hope was followed by an excessive depression. The certitude that some shelf in some hexagon held precious books and that these precious books were inaccessible, seemed almost intolerable. A blasphemous sect suggested that the searches should cease and that all should juggle letters and symbols until they constructed, by an improbable gift of chance, these canonical books. The authorities were obliged to issue severe orders. The sect disappeared, but in my childhood I have seen old men who, for long periods of time, would hide in the latrines with some metal disks in a forbidden dice cup and feebly mimic the divine disorder. [Borges, 1964, p. 66]

Borges' short story illustrates the absurdity of treating such a behaviour as "blasphemy": A library contains useful information only because an ordinary library is a *selection* of books from the set of all possible books, a set which would of course consist mostly of what we would describe as indecipherable nonsense. If the 'library' contained all possible books, it would cease to be a selection and consequently cease to be useful as a library. To find a legible English book in such a library is not any easier than simply generating a book by repeated dice roll, in fact these two approaches are completely indistinguishable if only the resulting book and the likelihood of legibility are considered.

This is of course easily confirmed by the basic notions of information theory: A selection of a few books, no matter whether they are randomly generated nonsense or legible English text, usually contain an amount of information roughly proportional to the number of books. If Borges' short stories were to be stored in a compressed form as bits, it would certainly be possible to compress them rather well thanks to the fact that any natural language contains a lot of regularity and thus redundancy, but each added story would nevertheless increase the amount of information so that 20 stories contain roughly twice the amount of information as 10 stories. But if all possible combinations of letters are to be stored, this 'collection' of books can, in the spirit of algorithmic information theory, be described by a very short program that merely generates all possible combinations one by one.[14] The information contained in the whole *Library of Babel* is less than that of the single short story that describes it, even though this short story itself would certainly be found (more than once) somewhere on the shelves in the library.

The number Ω is, according to algorithmic information theory, as informative as any number can ever possibly be, because Ω is completely random and incompressible. "In fact, Ω is a totally informative message, a message which appears random because all redundancy has been squeezed out of it, a message which tells us only things we don't already know" (Bennett, 1979, p. 9). But how informative is it really, if we have no way to *search* for it, or rather more importantly, no way to know when we have *found* it? It might seem to exist independently of us and only wait for someone to discover it, with the unfortunate limitation that we will never be able to discover it, because it *transcends* the power of human reason. But exactly like a book in the *Library of Babel*, which contains a significant amount of information when considered on its own as a *selection* from the vast library but ceases to be informative against the backdrop of the library as a whole, the number Ω ceases to be informative against the backdrop of all the other real numbers, because we have no way to compute it and thus no way to know when we have found it. This

14 Of course all these information-theoretic results already follow from classical information theory, in fact this whole discussion is information-theoretically a trivial consequence of Shannon's definitions of information and entropy, see Shannon, 1948, pp. 392–396.

is the flip side of Ω: It is pure information, but only if it is separated from the noise of all the other real numbers, which we cannot do. Against the backdrop of all numbers, it ceases to be informative, because the information of 'all' real numbers taken together is 0.

SEARCHING FOR THE PHILOSOPHER'S STONE

If the number Ω is examined against in the context of all real numbers, information theory turns out to be an unlikely ally of Wittgenstein's critique of an overly extensionalist view in mathematics. Any uniform treatment of the real numbers purely in terms of their resulting decimal expansions, with little to no intensional regard for the methods that produced them, will see the world of numbers as a vast Borgesian library that already contains any number that could possibly be written. We might not 'know' most of these numbers (in Bennett's sense, of "knowing" in contrast to "knowing of") and will in fact be unable to ever know some of them, such as Ω, but they seem to exist in a way that could lead us to say that even if we will never know Ω, God certainly does, because Ω exists. But as information theory and the example of Borges' *Library of Babel* make clear, a library that contains anything at all is not a library filled to the brim with astounding insights just waiting for us to discover them, but rather a trivial collection that could with no loss of insight be replaced by a "dice cup" to "mimic the divine disorder".

The idea of Ω as a "philosopher's stone" is nonsensical from the perspective of the Wittgenstein of the early 1930s, because it appeals to notions such as *searching for* and *finding* real numbers even though these notions are not applicable in the case of Ω. For Wittgenstein, the possibility to search for an answer is a fundamental prerequisite for asking a mathematical question. If there is no way to search for an answer, the question is nonsense, at least mathematically:

> Wo man fragen kann, kann man auch suchen und wo man nicht suchen kann, kann man auch nicht fragen. Und natürlich auch nicht antworten.
>
> Meine Erklärung darf nicht das mathematische Problem aus der Welt schaffen. D.h. es ist nicht so, daß ein mathematischer Satz erst dann gewiß einen Sinn hat, wenn er (oder sein Gegenteil) bewiesen worden ist. (In diesem Falle hätte nämlich sein Gegenteil nie Sinn (Weyl)) andererseits könnte es sein, daß gewisse scheinbare Probleme den Charakter des Problems – der Frage nach Ja und Nein – verlieren. [Ts-209, 69.5–69.6 / *PB* §148]

> Where you can ask, you can also search, and where you can't search, you can't ask. And of course you can't answer either.
>
> My explanation mustn't wipe out the existence of mathematical problems. That is to say, it isn't as if it were only certain that a mathematical proposition made sense when it (or its opposite) had been proved. (This would mean that its opposite would never have a sense (Weyl).) On the other hand, it could be that certain apparent problems lose their character as problems – the question as to Yes or No. [*PR* §148 (only second remark, first remark my translation)]

It has to be kept in mind that the above quote, written in the context of Wittgenstein's rather blunt dismissal of undecidability in mathematics, originates in a series of documents that are at times considerably more dogmatic than the nuanced positions in the *RFM* from 1937 and later.[15] Nevertheless, Wittgenstein's remarks in the 1930s are relevant in the present context of algorithmic information theory and the number Ω, as they arise in the context of (at the time) undecided questions of mathematics that would be answered by Ω, such as Fermat's last theorem:

> Was uns, abgesehen vom angeblichen Beweis Fermat's, dazu treibt, uns mit der Formel $x^n + y^n = z^n \ldots$ (F) zu beschäftigen, ist die Tatsache, daß man nie auf Kardinalzahlen gestoßen ist, die der Gleichung genügen; aber das gibt dem allgemeinen Satz *keinerlei* Stütze (Wahrscheinlichkeit) und ist also kein guter Grund zur Beschäftigung mit dieser Formel. Wohl aber kann man sie einfach als Schreibweise einer bestimmten allgemeinen Form ansehen und sich fragen, ob sich die Syntax in *irgend* einer Weise mit dieser Form beschäftigt. [Ts-209, 70.4 / *PB* §149]

> What, apart from Fermat's alleged proof, drives us to concern ourselves with the formula $x^n + y^n = z^n \ldots$ (F), is the fact that we never happen upon cardinal numbers that satisfy the equation; but that doesn't give the *slightest* support (probability) to the general theorem and so doesn't give us any good reason for concerning ourselves with the formula. Rather, we may look on it simply as a notation for a particular general form and ask ourselves whether syntax is *in any way at all* concerned with this form. [*PR* §149]

As the above remark shows, the Wittgenstein of the early 1930s disagrees with Chaitin's "pseudo-empirical" approach that would accept propositions as axioms if they are only *probably true*, in other words if no counterexample has been found even after an extensive search. For Wittgenstein, probability cannot justify a mathematical proposition as a general proposition. Even though such an objection might at first appear at odds with Wittgenstein's later views, especially his remark that an empirical proposition can indeed be fixed in a way that lets it act as a grammatical proposition (Ms-127, 224.1; Ms-124, 199.5 / *RFM VII* §74, see also Section 3.4) and his remarks on the changing river bed in *On Certainty*, these later views only emphasise that propositions can *change* their status, so that an empirical

15 Wittgenstein's views on undecidability at mathematics in the early 1930s cannot be discussed here at length. It should only be pointed out that even his rather dogmatic dismissal during that time period, written before the fundamental undecidability results of the 1930s, does not necessarily need to be read as an objection to undecidability results of the sort by Gödel and Turing or the undecidability of Kolmogorov complexity illustrated above. In these cases, the search for an answer is well defined and the undecidability stems from the fact that none of the possible answers will ever be considered correct (because either answer leads to a contradiction). In contrast, the undecidable mathematical problems that Wittgenstein seems to have in mind in the 1930s are mathematical questions that only appear undecidable because we lack a clear concept of how to search for an answer. Consequently, Wittgenstein's objection applies only to these latter mathematical questions, which are not undecidable but simply nonsense in need of philosophical clarification and surveyability.

propositions is turned into a grammatical proposition or vice versa. This view is not in conflict with Wittgenstein's remark from 1930, because what this earlier remark denies is the view that there is no categorical difference between empirical and grammatical propositions and that grammatical propositions are merely generalised empirical propositions that are *justified* by probability or certainty. Grammatical propositions are grammatical not because they are close to 100% certain, but rather because as hinges they are exempt from the language game of certainty and the need to justify these propositions empirically. This underlines the fundamental disagreement between Wittgenstein and Chaitin: Both want to argue against a rigid and dogmatic form of platonism in mathematics, but in the process Chaitin treats even mathematical axioms as empirically and probabilistically justified and thereby considers empirical and grammatical propositions to be only gradually different, whereas Wittgenstein emphasises their categorical difference, even if some propositions might change their status as a result of a change in our form of life.

Wittgenstein follows the above quote with a further remark on the *search* for mathematical answers:

> Ich sagte: Wo man nicht suchen kann, da kann man auch nicht fragen, und d.h.: Wo es keine logische Methode des Findens gibt, da kann auch die Frage keinen Sinn haben.
>
> Nur wo eine Methode der Lösung ist, ist ein Problem (d.h. natürlich nicht "nur wo die Lösung gefunden ist, ist ein Problem").
>
> D.h. dort wo die Lösung nur von einer Art Offenbarung erwartet werden kann, ist auch kein Problem. Einer Offenbarung entspricht keine Frage.
>
> Das {ist // wäre} so, wie wenn man nach den Erfahrungen eines Sinnes fragen wollte, den man noch nicht hat. Uns einen neuen Sinn geben, das würde ich Offenbarung nennen. [Ts-209, 70.5–70.8 / *PB* §149]

> I said: Where you can't look for an answer, you can't ask either, and that means: Where there's no logical method for finding a solution, the question doesn't make sense either.
>
> Only where there's a method of solution is there a problem (of course that doesn't mean 'Only when the solution has been found is there a problem').
>
> That is, where we can only expect the solution from some sort of revelation, there isn't even a problem. A revelation doesn't correspond to any question.
>
> It would be like wanting to ask about experiences belonging to a sense organ we don't yet possess. Our being given a new sense, I would call revelation. [*PR* §149]

The number Ω is a revelation in the purest sense: We might be led to believe that we can search for the number because the extensionalist view of the real numbers conjures up the image of a systematic search space in terms of decimal expansions, better yet, we can even approximate Ω by executing all possible programs concurrently and adding up the contributions of those that halt. But what we lack is a *method to find* Ω, because we would have no way to know that a particular number really is Ω even if we stumbled upon it by chance. We might think that we know the sense of Ω (because we "know of" Ω, in Bennett's words), but Ω without its value (in terms of a decimal expansion) has no use and 'finding' (or rather inventing) its value would give it a new sense. We can neither search for it nor find it, because we lack a *systematic* way to search for it:

> Der Fermat'sche Satz hat also keinen *Sinn*, solange ich nach der Auflösung der Gleichung durch Kardinalzahlen nicht *suchen* kann.
>
> Und "suchen" muß immer heißen: Systematisch suchen. Es ist kein suchen, wenn ich im unendlichen Raum nach einem Goldring umherirre. [Ts-209, 71.11 / *PB* §150]

> Thus Fermat's proposition makes no *sense* until I can *search* for a solution to the equation in cardinal numbers.
>
> And 'search' must always mean: search systematically. Meandering about in infinite space on the look-out for a gold ring is no kind of search. [*PR* §150]

It is no accident that Wittgenstein's remarks on Fermat's last theorem are applicable in the case of Ω, because Ω merely shifts around the problem of nonsensical concepts and hides any undefined notions under the cloak of a seemingly well-defined mathematical description. Ω suggests that undecided problems (in Wittgenstein's lifetime) such as Fermat's last theorem already have a sense and *could* be decided if only we could *discover* which machines would halt, whereas Wittgenstein held the view that any solution to such a conjecture would require the *invention* of new mathematical methods and calculi. Ω blinds us with its use of the notion of halting programs: We know what we mean when we say that a program halts, even if we have no idea how to decide whether a program halts in complicated cases. But this is exactly what Wittgenstein warns against. The application of concepts that are familiar and well understood in simple cases cannot simply be extrapolated to the 'more complicated' cases, as if the more complicated cases were like the simpler cases, just gradually more intricate. Instead, programs for which we cannot decide whether they halt are systems of a different kind, for which we need to invent new methods because we do not yet fully understand them.

To put it slightly more dogmatically, whether or not a program halts is in many cases not *discovered*, it is *invented*. We are only led to believe that we could discover in all cases whether a program halts by analogy with the simple cases that actually do halt on a computer and which are simple enough that we can exclude errors in the computer as possible explanations for the behaviour of these programs. In these simple cases, we can use a practical computer as our standard of measurement and will see that this standard agrees with our own conceptual standard in all useful theoretical models of computation. But in the more complicated cases, where we have no idea how to prove theoretically whether a program halts, the computer cannot act as our practical standard of measurement, since for highly complicated computations which could potentially take years or longer to run we can never be sure that even if the program were to halt it was not caused by a fault in the practical implementation of the computer. In these cases, where the possibility of an empirical measurement on a practical computer falls away, the decision whether a program halts can only be decided through theoretical means, which might require the *invention* of new methods.

In contrast, Ω suggests that the sense of all mathematical propositions is already fixed and that God could decide Fermat's last theorem by knowing whether a particular program halts. The number appears to give most mathematical conjectures an *unambiguous* sense, even if this sense is inaccessible to our human reason. Wittgenstein discusses such a picture in Ms-116, in a remark which later found its way into the *PI*:

> Ein Bild wird heraufbeschworen, das *eindeutig* den Sinn zu bestimmen scheint. Die wirkliche Verwendung scheint etwas Verunreinigtes der gegenüber, die das Bild uns vorzeichnet. Es geht hier wieder, wie in der Mengenlehre: Die Ausdrucksweise scheint für einen Gott zugeschnitten zu sein, der weiß, was wir nicht wissen können; er sieht die ganzen unendlichen Reihen und sieht in das Bewußtsein des Menschen hinein. Für uns freilich sind diese Ausdrucksformen quasi ein Ornat, das wir wohl anlegen, mit dem wir aber nicht viel anfangen können, da uns die reale Macht fehlt, die dieser Kleidung Sinn und Zweck geben würde.
>
> In der wirklichen Verwendung der Ausdrücke machen wir gleichsam Umwege, gehen durch Nebengassen; während wir wohl die gerade breite Straße vor uns sehen, sie aber freilich nicht benützen können, weil sie permanent gesperrt ist. [Ms-116, 162.3; Ms-120, 15r.3–16r.1; Ts-227a, 238.2 / *PU* §426]

> A picture is conjured up which seems to fix the sense unambiguously. The actual use, compared with that traced out by the picture, seems like something muddied. Here again, what is going on is the same as in set theory: the form of expression seems to have been tailored for a god, who knows what we cannot know; he sees all of those infinite series, and he sees into the consciousness of human beings. For us, however, these forms of expression are like vestments, which we may put on, but cannot do much with, since we lack the effective power that would give them point and purpose.
>
> In the actual use of these expressions we, as it were, make detours, go by side roads. We see the straight highway before us, but of course cannot use it, because it is permanently closed. [*PI* §426]

For Wittgenstein, these side alleys are just as important as the main roads, as only the variety of side alleys shows how a concept is being used in our life. Set theory and especially Cantor's uniform treatment of the real numbers from an extensionalist viewpoint create the illusion that such a general view would have sense even in the absence of any specialised methods and uses, since an omniscient being could understand these expressions purely extensionally. But the variety of intensional methods is not secondary to any uniform treatment, they are not merely "detours". This critique of extensionalism as a viewpoint of God is developed further in Ms-124 and Ms-127:

> Ein Beweis der zeigt, daß die Figur "777" in der {Extension // Entwicklung} von π vorkommt aber nicht zeigt *wo*. Nun, so bewiesen wäre dieser 'Existenzsatz' für gewisse Zwecke *keine Regel*. Aber könnte er nicht z.B. als Mittel der Einteilung von Entwicklungsregeln dienen. Es wäre etwa auf analoge Art bewiesen daß "777" in π^2 nicht vorkomme, wohl aber in $\pi \times e$ etc. Die Frage wäre nun: Ist es vernünftig von dem betreffenden Beweis zu sagen: er beweise die Existenz von "777" in dieser Entwicklung. Dies kann einfach irreführend sein. Das ist eben der Fluch der Prosa, & besonders der Russellschen Prosa, in der Mathematik.
>
> Was schadet es, z.B., zu sagen, Gott kenne *alle* irrationalen Zahlen? Oder: sie seien schon alle da, wenn wir auch nur gewisse kennen? Warum sind diese Bilder nicht harmlos?
>
> Einmal verstecken sie gewisse Probleme. – [Ms-124, 138.3–139.2 / *BGM VII* §41, with the second remark also in Ms-127, 57.2–57.3]

> A proof that shews that the pattern '777' occurs in the expansion of π, but does not shew *where*. Well, proved in this way this 'existential proposition' would, for certain purposes, not be *a rule*. But might it not serve e.g. as a means of classifying expansion rules? It would perhaps be proved in an analogous way that '777' does not occur in π^2 but it does occur in $\pi \times e$ etc. The question would simply be: is it reasonable to say of the proof concerned: it proves the existence of '777' in this expansion? This can be simply misleading. It is in fact the curse of prose, and particularly of Russell's prose, in mathematics.
>
> What harm is done e.g. by saying that God knows *all* irrational numbers? Or: that they are already all there, even though we only know certain of them? Why are these pictures not harmless?
>
> For one thing, they hide certain problems. – [*RFM VII* §41]

In Ms-124, Wittgenstein does not immediately explain what these problems might be, but he returns to the idea nearly a hundred pages later in two remarks that were published directly following the above remarks in *RFM VII*:

> Angenommen die Menschen berechnen die Entwicklung von π immer weiter & weiter. Der allwissende Gott weiß also, ob sie bis zur Zeit des Weltuntergangs zu einer Figur 777 gekommen sein werden. Aber kann seine *Allwissenheit* entscheiden, ob die Menschen nach dem Weltuntergang zu jener Figur gekommen *wären*? Sie kann es nicht. Ich will sagen: Auch Gott {kann // könnte} Mathematisches nur durch Mathematik entscheiden. Auch für ihn kann die bloße Regel des Entwickelns nichts entscheiden, was sie für uns nicht entscheidet.
>
> Man könnte das so sagen: Ist uns die Regel der Entwicklung gegeben, so kann uns nun eine *Rechnung* lehren, daß an der fünften Stelle die Ziffer "2" steht. Hätte Gott dies, ohne diese Rechnung, bloß aus der Entwicklungsregel wissen können? Ich will sagen: Nein. [Ms-124, 175.2–175.3 / *BGM VII* §41]

> Suppose that people go on and on calculating the expansion of π. So God, who knows everything, knows whether they will have reached '777' by the end of the world. But can this *omniscience* decide whether they *would* have reached it after the end of the world? It cannot. I want to say: Even God can determine something mathematical only by mathematics. Even for him the mere rule of expansion cannot decide anything that it does not decide for us. [*RFM VII* §41]

It is clear that the example given by Wittgenstein is one of the conjectures that would be decidable with the knowledge of Ω, in fact Wittgenstein's example of the figure '777' in π can be understood as a paradigmatic case for any mathematical proposition decidable by knowing whether the calculation ever comes to an end, in other words by a solution to the halting problem as provided by Ω. From the standpoint of algorithmic information theory, the question raised by Wittgenstein about the existence of the figure '777' in π is well defined, even if we might not be able to answer it for practical reasons.[16] An omniscient being would certainly be able to answer the question, since omniscience would include knowledge of Ω and thus knowledge of whether the search routine for the figure in π would ever come to a halt, even after the end of the world.

But such a view is exactly what Wittgenstein wants to deny: No omniscient God can decide such a question, because no answer could be considered an answer to *our* mathematical question employing *our* mathematical concepts. After all, what would it mean to say that a program does not halt before the end of the world but would halt some time *after*? There is no way for us to verify this statement, let alone use it in a more meaningful way than the opposite statement. To say that a program halts only after the end of the world is just as nonsensical as to say that we will not be able to prove that a program will halt until the end of the world and that it will in fact never halt.

This situation is not comparable to a case of a trivially looping program for which we can prove that it will never halt: To say that such a trivial loop will never halt is not a prediction about the program running on an actual computer, as we would not call our prediction incorrect if after a billion years the program did turn out to halt due to a malfunction of the computer. It is instead a rule of grammar, which leads us to exclude from our language game the *possibility* of *calling* anything the end of the program. If the program did turn out to halt after a billion years we would chalk it up to programming or hardware errors, because our theoretical model of computing is how we measure the practical execution on a computer in this case, not the other way around.

We might be tempted to think that a divine revelation could give us the *certainty* that if we run a particular program it will not suddenly halt after a few billion years, but this is a *practical* certainty that extends at most until the end of the world and thus depends only on

16 Wittgenstein's example assumes that a figure such as '777' will not be found until the end of the world. Of course computers of today and all the more so improved computers in the future will be able to find simple figures such as '777' in π by calculating it. This should not detract from Wittgenstein's point, as he clearly assumes that the figure has not been found until the end of the world and that any figure that has not appeared in π up until that moment would do the job. The specific figure '777' should thus be understood as a placeholder for an arbitrarily long figure that cannot or has not been successfully calculated by even the fastest computers.

the knowledge of whether the program will come to a halt *before* the end of the world. Any answer that transcends these boundaries by answering the halting question after the end of the world must be unintelligible for us, because it employs concepts that have no analogue in our life.

This is why not even God can decide a mathematical conjecture with the knowledge of Ω: If the halting decision needed to decide the conjecture lies beyond the end of the world, it transcends our limits of the world and our concepts. An answer would then not be an answer to our question, but to an entirely different question only meaningful to omniscient beings, while our question remains merely nonsense. If the halting question for the particular program is decided within a practical timeframe and thus does not transcend the limits of our world (for example if the program turns out to halt after a few minutes), then no knowledge of Ω is necessary to decide the mathematical conjecture, at least not knowledge of more than finitely many bits of Ω, which can always (if only rather impractically) be approximated. But then Ω loses its role as a mathematical oracle or "God's number" and is instead seen in a more 'homespun' way, as a compact but entirely impractical representation of the halting probability for programs up to a certain length. From the perspective of Wittgenstein, Ω is thus philosophically irrelevant: It never had any practical relevance to begin with and any theoretical aspirations to transcend the limits of mathematics turn out to be philosophical nonsense.

The philosophical confusion at play here is a result of a misunderstanding of *surveyability*. Extensionalism in mathematics in general and applications of Cantor's diagonal method in particular suggest that whatever is surveyable in the finite case remains surveyable in the infinite case, thanks to the "u.s.w. ad inf." and the application of mathematical induction. This leads to the picture of a God that can survey the vast landscape of the real numbers, all of them already existing, waiting for us to discover those that can be discovered. But in the case of Ω as well as in many other cases, there is no transcendental vantage point that could survey the system of all systems from above, because any generalised explanation would have to invent new concepts and fail to describe the existing concepts. If the aim is to clarify the existing concepts, all that can be done is to *describe* what is already there, by calling to mind the variety of uses with its interconnections in a *surveyable representation*.

B

SURVEYABLY REPRESENTING THE "RFM"

The "works" of Wittgenstein that were published after his death by the three literary executors G.E.M. Anscombe, Rush Rhees and Georg Henrik von Wright have often been the target of criticism, giving rise to the question of what can even be considered a "work" by Wittgenstein. The *Remarks on the Foundations of Mathematics* ("*RFM*"), first published in 1956 and then heavily revised and expanded in 1974, are certainly one of the most problematic publications in this regard, as the literary executors themselves pointed out:

> The *Remarks on the Foundations of Mathematics* occupy a nearly unique, and not altogether happy, position among the posthumous publications. In addition to the relatively finished Part I, corresponding to typescripts 222, 223, and 224 of the catalogue and constituting the second half of the prewar version of the *Investigations*, the *Remarks* contain *selections* from several manuscripts (117, 121, 122, 124, 125, 126, and 127). In the revised edition of 1974 (English translation 1978) the selections from those manuscripts were somewhat enlarged and a further manuscript (164) which was not known to the editors at the time of the first edition was added, practically without omissions. A publication of the manuscripts *in toto*, however, seemed to us excluded even at the time of preparing the new edition. [Von Wright, 1993, p. 502]

The following pages will present visualisations of the similarities and differences between the parts II–VII of the *RFM* (which never advanced to the typescript phase) and their underlying *Nachlass* documents. It is explicitly not the aim to answer whether the editorial decisions by the literary executors were justified or which alternative ways of publication might have been feasible in the historical context shortly after Wittgenstein's death. Instead, the following pages attempt to introduce and describe a "surveyable representation" of these editorial decisions, which allows the reader to make up their own mind about the editorial interventions and, more importantly, gain a synoptic understanding of the differences between the published work and the corresponding documents in the *Nachlass*. The term "surveyable representation", borrowed from *PI §122*, is to be interpreted here only as a motivating idea, ignoring the considerable philosophical context that is attached to it in Wittgenstein's philosophy. The following visualisations are meant only as a philological tool, it should be clear that they can only illuminate the broad strokes of the editorial work in the *RFM*, even when they are paired with a short commentary. They are thus not intended as a replacement for existing philological commentaries and critiques, but rather as a visual supplement and jumping-off point for deeper philological work.

A CLASSIFICATION OF EDITORIAL INTERVENTIONS

Broadly speaking, the editorial interventions that are surveyed in the following visualisations can be classified into four groups: reordered remarks, excluded remarks, inserted remarks and edits related to missing contextual indications (e.g. a separating line or substantial chronological break that is not reflected in the published work).

R/R+/R++	Reorderings of 1, up to 5, or more than 5 remarks
E/E+/E++	Exclusions of 1, up to 5, or more than 5 remarks
I/I+/I++	Insertions of 1, up to 5, or more than 5 remarks
C	Loss of context due to a chronological jump

Table 6: Classification of Editorial Interventions

Reorderings are usually rather benign and are often hinted at in the source document by Wittgenstein himself, either as a continuation or reformulation of a previous remark (with the characteristic three dashes "– – –") or through the explicit use of arrows.

Remarks present in the source document but excluded form the published version in *RFM* are by far the most frequent case of editorial interventions in *RFM*. Sometimes these exclusions are substantiated by indications in the source, but more often such editorial decisions were made either because the remark was already published in another work or because the remark was deemed irrelevant or inessential for publication.

Insertions of remarks happen much more rarely in the published works than the exclusion of remarks. Sometimes remarks need to be inserted because Wittgenstein continued writing in another document, at other times remarks are inserted because they are deemed relevant by the literary executors.

For reordered, excluded or inserted remarks, it is helpful to distinguish between edits involving single remarks, less than 5 remarks or many 5 or more remarks. An edit involving more than a single remark will be indicated by a '+' (e.g. 'E+'), one involving 5 or more by '++' (e.g. 'I++')

Wittgenstein's writing process was not always as straightforward or linear as the published works suggest. He often started working on another topic in the same document, separating groups of remarks with a horizontal line, or continued writing in the same document months, sometimes years later. This segmentation of documents is usually not indicated in the published version, which can lead to a loss of context compared to the source document, though the editors can hardly be faulted for not including every indication. A loss of context caused by a chronological jump in the source document that is not reflected in the published work will be marked by the indication 'C' in the following visualisations.

A SYNOPSIS OF THE RFM

Based on the visualisations on the following pages, the editorial decisions in the *RFM* can be briefly summarised as follows:

1. Part I is based on a typescript and exhibits only minor editorial interventions, it will therefore not be discussed here.
2. Part II may be the smallest of the published parts, but is compiled quite heterogeneously from two larger documents; a clearly delineated part of Ms-117 and the whole of Ms-121, with the frequent omission of sometimes large passages in the latter case.
3. Part III is heavily edited from start to finish and contains only around half of the remarks of its source documents Ms-117 and Ms-122, but these edits are rather uniform throughout the whole part and usually restricted to the exclusion of at most a handful of remarks.
4. Part IV is sourced primarily from Ms-125 and can be described as editorially inconspicuous for most of its first 50 sections, with the last 10 sections being the result of reorderings and a number of insertions from Ms-126, Ms-127 and Ms-121.
5. Part V is taken from Ms-126 (first two thirds) and Ms-127 (last third), with several reorderings and frequent omissions in the case of both documents and a jump of about 100 pages between remarks in Ms-127.
6. Part VI is editorially unproblematic and corresponds *in toto* and with only minor corrections to Ms-164, while omitting the last part of that document on private language.
7. Part VII is selected (with numerous remarks being excluded) exclusively from Ms-124, which shares similar or identical remarks with earlier draft stages (Ms-161 and Ms-163), contemporaneous documents (Ms-127) and later documents (Ms-129) as well as typescripts (Ts-227a/b and thus the *Philosophical Investigations*).

HOW TO READ THE VISUALISATIONS

The visualisations on the following pages present the editorial decisions described above for the parts II–VII of the *RFM*, with one visualisation per part. Each image shows three columns of remarks, ordered from top to bottom as they appear in the published work of the source documents. The remarks in the different columns are connected if they correspond to each other, either because a remark in the published work is identical (with minimal textual changes) to a remark in the source document or because a remark in a source document also appears in other related documents. The column on the left displays the remarks that are included in the published work, with diamond markers indicating the section boundaries in the work (and section numbers to the left). The column in the middle shows the remarks of one or several source documents (with page numbers to the right) that form the basis of the published work. The column on the right displays the remarks of secondary documents (with page numbers to the right) with textually similar remarks in the source document of the published work.

By displaying the connections from the published work (on the left) to the source documents (in the middle), the visualisation shows where the published work originates and which parts of the source document were reordered / excluded / inserted. By displaying the connection from the source document (in the middle) to the related documents (on the right), the visualisation can often hint at why a particular segment was or was not published (which will be elaborated in the commentary).

To the left of the vertical document lines (in the middle and right columns) appears a kind of bar code pattern of short horizontal lines with small or larger circle markers directly to the left. These lines and markers indicate that the remark is published in the work currently being visualised (indicated by a slightly larger circle marker) or is published in other works (either other parts of the *RFM* or other works, both indicated by a small circle marker).

Longer horizontal lines, both solid and densely dotted, which cross the document lines (in the middle and right columns) indicate a break between adjacent remarks in the document. Solid lines indicate a separating line by Wittgenstein himself, densely dotted lines indicate a chronological break of more than 30 days between remarks.

METADATA SOURCES

The following visualisations were all generated based on the metadata contained in the XML transcriptions of the Wittgenstein Archives Bergen. Metadata concerning published works is quite complete, but probably not entirely free of errors. The existing *Source Catalogue of the Published Texts* (Biggs and Pichler, 1993) shows that such a mapping between published works and *Nachlass* documents is far from trivial. In fact, back in 1993, many of the sections of the *RFM* (especially in the parts III, IV and V) could not be easily associated with their document sources, as evidenced by a number of "?" entries in the source catalogue. Nowadays, all of these sections are included in the XML metadata, but it is not always obvious where to look for them. This is one of the areas where visualisations can offer a starting point for researchers not intimately familiar with all the *Nachlass* documents.

The connections between primary (in the middle column) and secondary documents (in the right column) were automatically calculated based on the textual overlap between remarks in terms of shared word bi-grams and tri-grams. Here the caveat applies even more: These similarity connections have been checked by hand in many cases, but are only approximate and thus not free of errors. It should be pointed out that the similarity between remarks is not binary, but rather specified as a confidence value between 0 and 1 (which is reflected in the opacity of the visualised connections). It is not guaranteed that every single connection is correct, but taken as a whole, the method of visualisation should result in a "surveyable representation" that correctly displays the editorial decision in a synoptic fashion.

PART II

RFM II, which deals mainly with Cantor's diagonal argument, is by far the shortest of all the parts. The underlying documents, Ms-117 and Ms-121, are much longer, however, and contain a wealth of remarks not published in any part of the *RFM*.

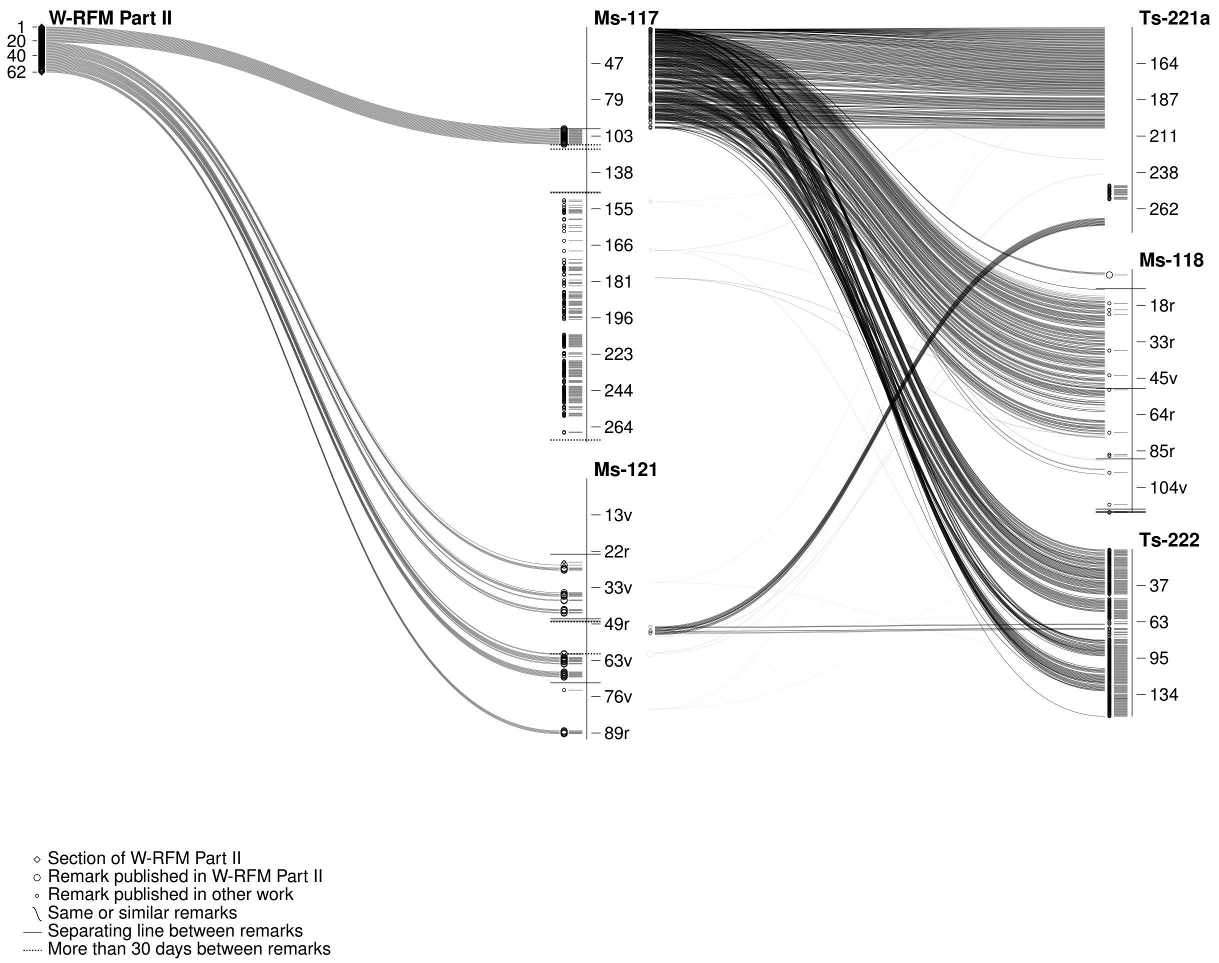
§§1–22: Ms-117

In the case of Ms-117, the reason is straightforward: There are 5 breaks in Ms-117, corresponding (with the exception of the last break at the end of Ms-117) to the different parts described in *The Wittgenstein Papers* (Von Wright, 1993, p. 495). The first 22 sections of *RFM II* correspond directly to the clearly separated first part in Ms-117. Many of the remarks in the fifth part of Ms-117 were published in another work (*RFM III*) and will be discussed on the next pages.

§§22-62: Ms-121

Ms-121 is heavily edited, however. The unpublished remarks up to 23v.2 mostly revolve around non-mathematical topics and are separated by a line from what follows. The rest deserves a closer look:

E++	23v.3–25r.2 cover topics such as the inner and the outer, expressions of pain and other non-mathematical issues.
E+	26r.3–27r.3 can be read as preliminary mathematical remarks leading up to the investigation of Cantor's argument, with one remark in between (26v.3) published in *Culture and Value*.
E+	27v.3–28r.2 are unpublished remarks between §23 and §24 clearly belonging to the remarks on Cantor.
E++	29v.1–35v.2 begin as a continuation of the remarks on Cantor after §27 and then gradually evolve into remarks on provability and Russell's logic.
E+	37r.3–38r.2 fill the gap between §32 and §33 with additional examples.
E++	39r.2–41r.4 are 11 unpublished remarks clearly belonging to Wittgenstein's remarks on Cantor.
E+, *E++*	45r.1–60r.1 have all been excluded from the published version. The first two of these remarks clearly continue the remarks on Cantor, most of the others revolve around other topics, however, with the last remark of the stretch being clearly separated from the next published remark by a break of more than three months.
E+	60v.3–61v.1 are remarks on Cantor between §41 and §42 that were dropped from the published work.
E	64r.4 is a single remark on Cantor that was dropped (it is rather short and forms only a supplementary example, however).
E++	65v.2–67v.2 (10 remarks) are remarks on Cantor and fill the gap between §50 and §51, but were not published.
E++	69v.3–72r.1 are all related to the remarks on Cantor, but all 7 remarks were dropped. After the last remark, Wittgenstein draws a separating line and then begins writing about contradictions in logic.
E++	72r.3–85r.2 revolve around issues of contradictions and provability and were not published. There is no clear separation between 85r.2, the last remark on provability, and 86r.1, the first remark explicitly referencing Cantor again, however.
E++	86r.1–87v.3 (7 remarks) clearly belong to the remarks on Cantor and lead up to the published §58, but were left out in *RFM II*.
E+	89v.2–93v.2 (8 remarks) after the last remark published in *RFM II*, §62, were all not published. The first half is still about Cantor, the second half revolves around Russell's logic, but the transition is relatively seamless.



W-RFM Part II
1
20
40
62
Ms-117
47
79
103
138
155
166
181
196
223
244
264
Ms-121
13v
22r
33v
49r
63v
76v
89r
Ts-221a
164
187
211
238
262
Ms-118
18r
33r
45v
64r
85r
104v
Ts-222
37
63
95
134
◇ Section of W-RFM Part II
○ Remark published in W-RFM Part II
∘ Remark published in other work
\ Same or similar remarks
— Separating line between remarks
······ More than 30 days between remarks

part iii

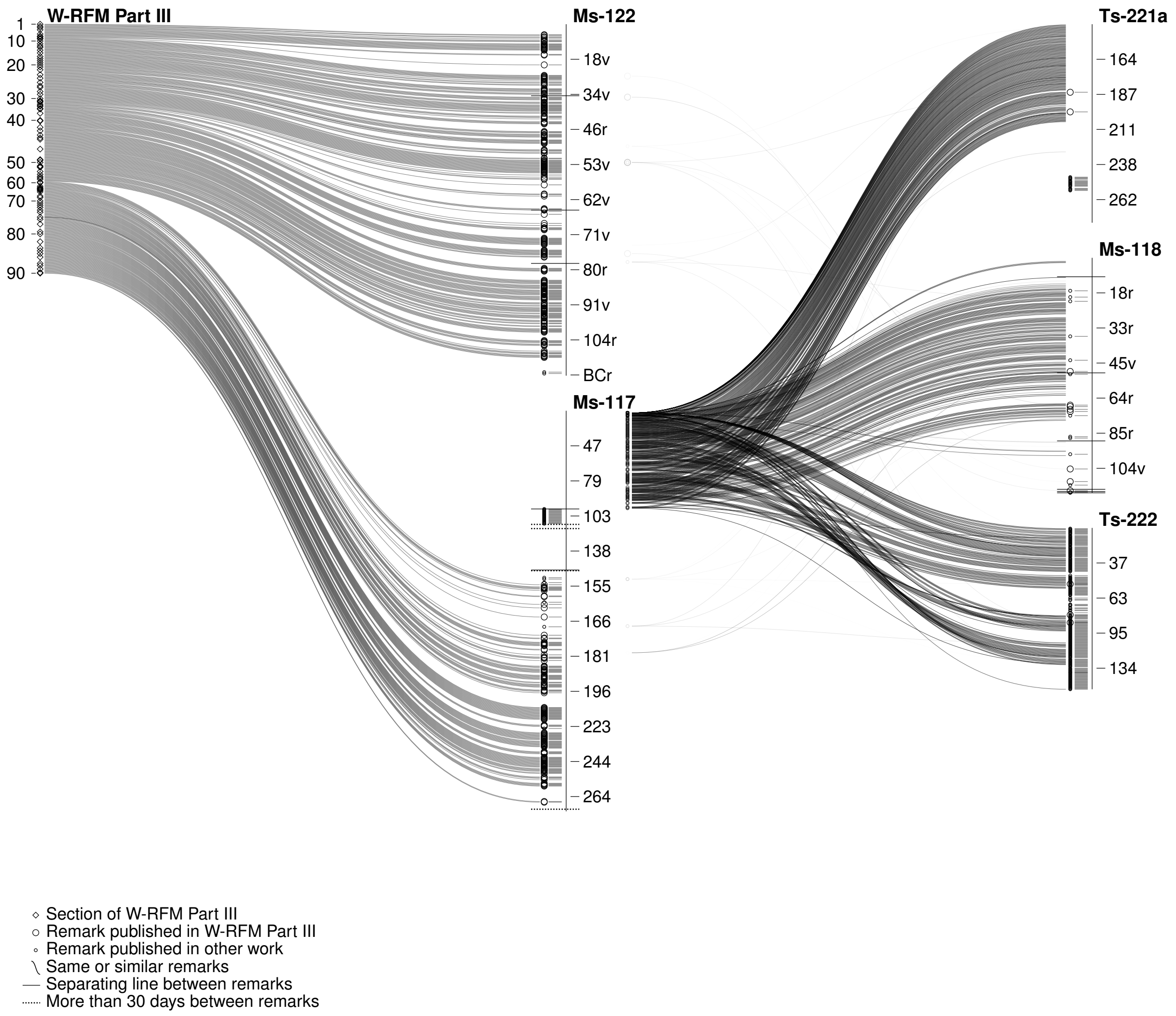
RFM III is the lengthiest part of the *RFM* and sourced exclusively from two documents, Ms-122 and Ms-117. As the visualisation on the following page shows, the number of excluded remarks is considerable, only about half of the remarks in the relevant parts of Ms-122 and Ms-117 made it into the published work. It would be infeasible to discuss all of these exclusions here, the following description will thus be limited to the most salient points and only highlight exclusions of 5 or more related remarks (with the count always excluding remarks in *Culture and Value*). A very thorough discussion of the unpublished passages can be found in the commentary of part III in Mühlhölzer, 2010.

§§1–58: Ms-122 Ms-122 is characterised by exclusions of thematically connected remarks, in addition to the exclusion a passage of vaguely related remarks in the beginning (1r.1–4v.4) and remarks at the end of the document that were published in *Culture and Value*.

E++	1r.1–4v.4 On contradictions in logic and the use of the word "heterologic".
E++	15r.2–16r.3 (5 unpublished remarks.)
E++	17r.2–21v.2 (13 unpublished remarks.)
E++	22r.2–26v.1 (14 unpublished remarks.)
E++	45v.2–46v.1 (10 unpublished remarks.)
E++	49r.2–49v.5 (7 unpublished remarks.)
E++	51v.2–52r.3 (6 unpublished remarks.)
E++	58v.2–59r.4 (7 unpublished remarks.)
E++	59v.2–61r.2 (12 unpublished remarks.)
E++	62r.2–64v.1 (17 unpublished remarks.)
E++	66r.3–68r.3 (12 unpublished remarks.)
E++	70v.3–72r.2 (11 unpublished remarks.)
E++	74r.3–75r.2 (7 unpublished remarks.)
E++	77r.3–79r.4 (15 unpublished remarks.)
E++	80r.5–82v.2 (12 unpublished remarks.)
E++	102v.1–104r.4 (11 unpublished remarks.)
E++	107r.2–109r.2 (7 unpublished remarks.)
E++	112v.4–118r.3 (20 unpublished remarks.)

§§59–90: Ms-117 Ms-117, 148.2 is explicitly marked by Wittgenstein as a continuation of Ms-122, so the following notes apply only the remarks after 148.2.

E++	Ms-117, 148.3–153.5 (17 unpublished remarks.)
E++	Ms-117, 156.4–158.2 (7 unpublished remarks.)
E++	Ms-117, 159.2–160.6 (9 unpublished remarks.)
E++	Ms-117, 161.5–164.2 (12 unpublished remarks.)
E++	Ms-117, 165.1–171.4 (23 unpublished remarks.)
E++	Ms-117, 178.5–180.2 (6 unpublished remarks.)
E++	Ms-117, 182.5–184a.1 (7 unpublished remarks.)
E++	Ms-117, 197.2–203.4 (19 unpublished remarks.)
E++	Ms-117, 209.2–221.2 (7 unpublished remarks.)
E++	Ms-117, 223.3–227.1 (7 unpublished remarks.)
E++	Ms-117, 260a.3–266.2 (21 unpublished remarks.)
E++	Ms-117, 267.4–271.2 (6 unpublished remarks.)


W-RFM Part III
1
10
20
30
40
50
60
70
80
90
Ms-122
18v
34v
46r
53v
62v
71v
80r
91v
104r
BCr
Ms-117
47
79
103
138
155
166
181
196
223
244
264
Ts-221a
164
187
211
238
262
Ms-118
18r
33r
45v
64r
85r
104v
Ts-222
37
63
95
134
Section of W-RFM Part III
Remark published in W-RFM Part III
Remark published in other work
Same or similar remarks
Separating line between remarks
More than 30 days between remarks

PART IV

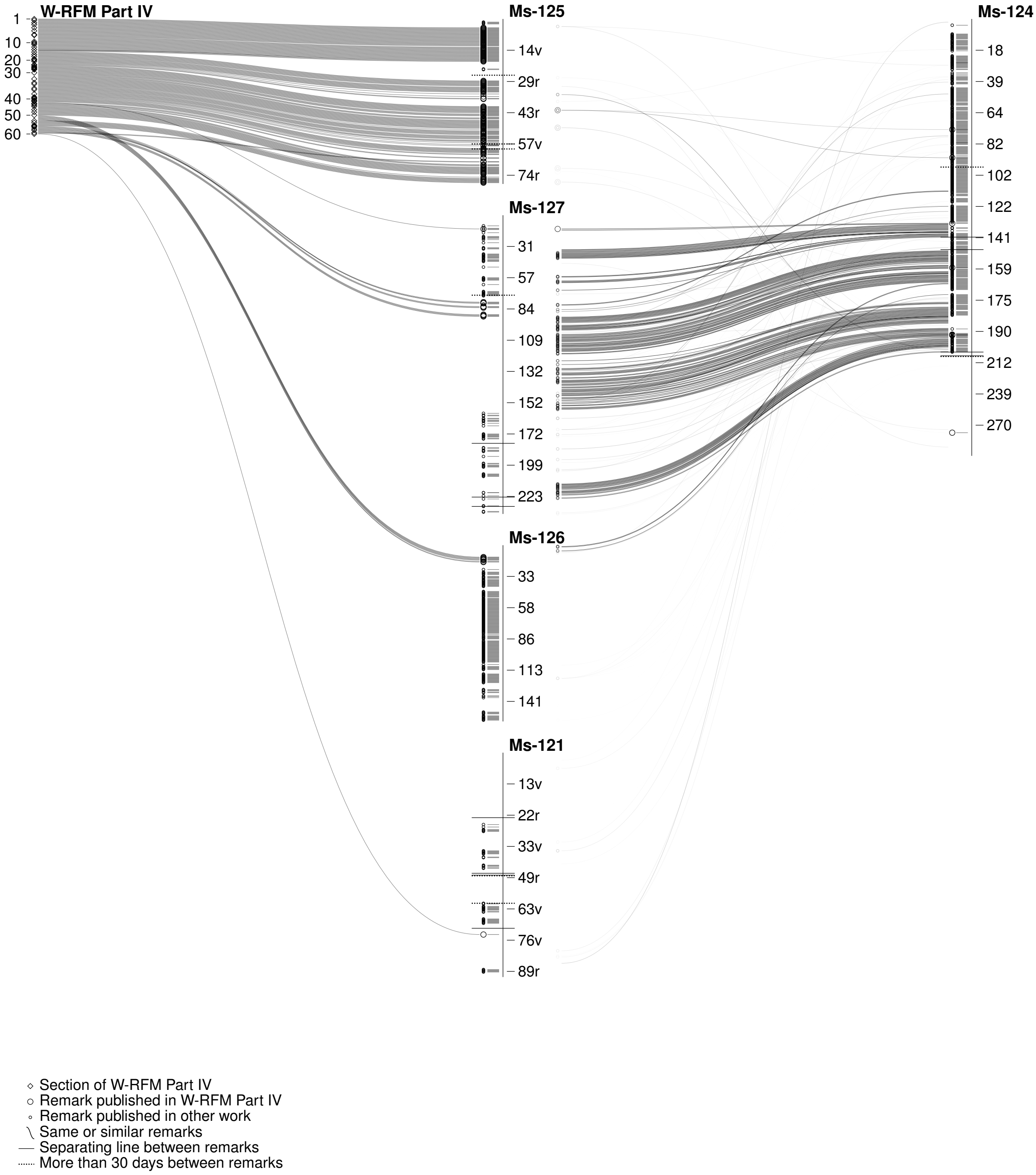
§§1–49: Ms-125, (Ms-127)

RFM IV is a rather peculiar part, as the extent of the editorial interventions changes quite substantially over the course of the work. Most of the first 49 sections are taken relatively straightforwardly from Ms-125, although some remarks are excluded and one remark (in §43) is inserted from Ms-127. Most exclusions before §50, however, are explained by the large number of personal diary entries in Ms-125, which were understandably omitted.

§§50–60: Ms-126, Ms-125, Ms-127, (Ms-121)

The last 11 sections, however, show more idiosyncratic editorial decisions and originate not only in Ms-125, but also Ms-127, Ms-126 and (in the case of the last remark) Ms-121. The insertions in §50 from Ms-126 concern the idea of a "word in reverse", but related remarks that follow in Ms-125 have not been published. The insertions in §§55–60 are concerned with the topic of contradictions in logic. Only the two remarks of §59 originate in Ms-125.

E+	Ms-125, 1r.1–5v.3 were deemed unrelated to the discussion in part IV, 4 of these remarks were published in *Culture and Value* instead. Apart from the first two remarks (personal diary entries written in code), there is no indication for a separation from the following remarks in the document.
E	Ms-125, 13v is a coded personal diary entry.
E++	Ms-125, 20r.3–28r.3 are all dropped from part IV, two remarks are published in *Culture and Value*, however. Ms-125, 26r.3 is a coded personal diary entry ("1.4.42"). The last dated remark before it is Ms-125, 3r.2 ("4.1"). Most of the remarks are related to the topic of *RFM IV*.
E	Ms-125, 31v.3 is another coded personal diary entry.
E++	Ms-125, 34v.1–39v.2 are, with the exception of 4 remarks, all excluded from *RFM IV*, but topically connected to the rest of Ms-125.
E	Ms-125, 39v.3–58r.3 contain only a few unpublished remarks: Ms-125, 44v.3, Ms-125, 46r.2, Ms-125, 50v.2 (a coded diary entry), Ms-125, 51r.2 (a personal entry, not coded, but in brackets), Ms-125, 53v.4 (coded diary entry), Ms-125, 54r.2, Ms-125, 55r.2, Ms-125, 56v.2 (coded diary entry), Ms-125, 57v.1 (coded diary entry), Ms12558v.2–59r.2 (coded diary entries). The dates can be puzzling: Ms-125, 26r.3: "1.4.42", Ms-125, 36v.2 / Ms-125, 36v.3: "9.2", Ms-125, 57v.1: "26.4", the next remark "18.5".
C	There is a chronological jump of more than three months between the published remarks Ms-125, 59v.2 (27.5.42) and Ms-125, 60v.2 (15.9.42).
I	§43 (third remark) is taken from Ms-127, 13.2 (written about half a year later), which explicitly mentions the distribution of prime numbers as "synthetic *a priori*". There is nothing in Ms-125 indicating such an insertion, the remark in Ms-127 is separated from the surrounding remarks by a long vertical line, however.
E	Ms-125, 63v.2 is a short remark on hidden contradictions.
E+	Ms-125, 64r.1–64v.2 introduce the picture of natural calculating machines, further developed in Ms-125, 65v.3–66v.2, but all of these remarks are dropped in *RFM IV*.
R+	Ms-125, 67r.2–68r.1 are two remarks on Russell's contradiction that were moved to the end of *RFM IV* (as §59).
E+	Ms-125, 68r.2–68v.2, Ms-125, 70r.2 are unpublished but related to *RFM IV*.
I++	Ms-126, 12.2–17.2 are inserted as §50 (with the first remark Ms-125, 73v.2).
E+	The insertion of remarks from Ms-126 is all the more puzzling given that the directly related remarks Ms-125, 73v.3–75r.1 are excluded in *RFM IV*.
E	Ms-125, 75v.1 is another coded diary entry.
E	Ms-125, 79r.2 (the last remark in Ms-125) is not published.
I++	§§55–58 are taken from Ms-12780.3–83.4 and discuss contradictions in logic. Inserted before *RFM IV* §59 from Ms-125.
I	§60, also on contradictions in logic, is taken from Ms-121, 74v.2.


W-RFM Part IV
1
10
20
30
40
50
60
Ms-125
14v
29r
43r
57v
74r
Ms-127
31
57
84
109
132
152
172
199
223
Ms-126
33
58
86
113
141
Ms-121
13v
22r
33v
49r
63v
76v
89r
Ms-124
18
39
64
82
102
122
141
159
175
190
212
239
270
Section of W-RFM Part IV
Remark published in W-RFM Part IV
Remark published in other work
Same or similar remarks
Separating line between remarks
More than 30 days between remarks

part v

§§1–34: Ms-126, (Ms-127)

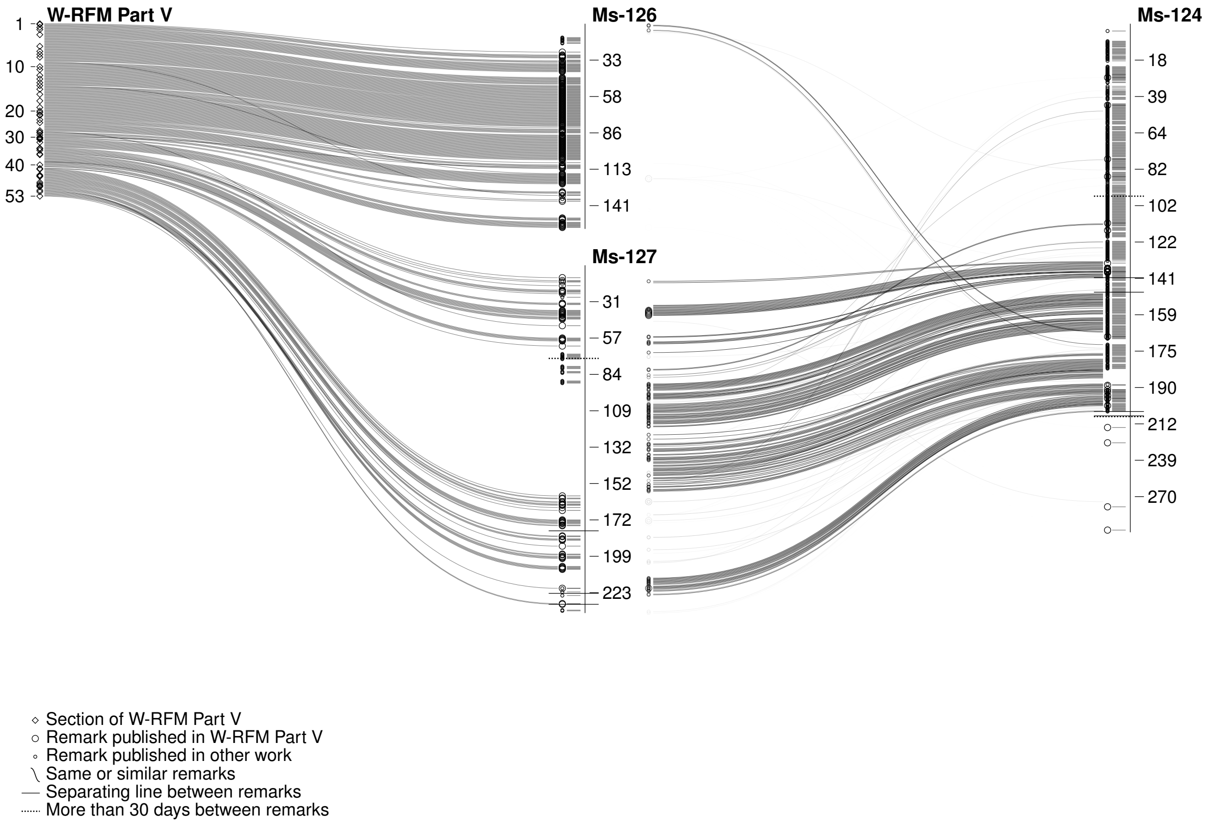
RFM V is compiled from two documents, Ms-126 and Ms-127, with the first part (§§1–34) being a relatively straightforward selection of remarks from Ms-126 on extensional view of mathematics, albeit with several reorderings, an insertion from Ms-127 (§27) and the omission of a few passages. (The following lists highlight only exclusions of 5 or more remarks.)

E++	1.1–28.2 are remarks on mathematics that were only partly published (in *RFM IV*).
E+	21.2–22.2, 86.2, 105.4, 132.4 are coded personal diary entries.
E++	43.1–45.2 are unpublished remarks about reflexive equality.
R	110.3 is a remark about Dedekind that was moved from between §24/§25 to §30.
E++	113.2–116.1 fit into the reflections of §25, but were not published.
R	116.2 discusses "abzählbar" versus "numerierbar" and was moved from §25 to §15.
E++	126.2–131.2 discuss the variety of proofs in Hardy, with the last remark on Gödel.
R	133.3 on the "expansion of an irrational number" is moved from §29/§31 to §9.
E++	138.3–147.2 discuss the continuity of curves, but were not published.

§§35–53: Ms-127

In Ms-127, Wittgenstein continues his discussion of extensions and the Dedekind cut, before embarking on a more general investigation of concept formation through proofs. A large part of the latter discussion is not published in part V, but appears in a similar form in Ms-124. Neither the chronological break (of nearly a year) nor the exclusion of a large portion of Ms-127 are evident in *RFM V*.

E++	1.1–9.2 discuss the Dedekind cut, but were not published.
R	13.3 appears between remarks of §36, but was moved to §38.
E++	17.3–20.2 continue with the discussion of the Dedekind cut, but were not published.
E++	28.2–31.3 discuss Russellian contradictions.
E++	33.2–36.1 consider a definition as a "determination of a concept".
E++	43.3–57 are all unpublished, with the exception of 47.4 (§27). The last 7 remarks revisit Cantor's diagonal argument.
R	47.4 was inserted between remarks from Ms-126 and published as §27.
E++	61.2–64.3 discuss concept determination / formation.
E++	65.2–70.2 are unpublished remarks on the continuum of the real numbers.
E++, C	70.2–73.2 were all published in *Culture and Value*, the last of these remarks is written nearly a year after the others.
E++	82.4–90.2 discuss contradictions in logic and were partly published in *RFM IV*.
E++	90.3–159.2 discuss rule following and related issues (with a lot of overlap in Ms-124), then gradually evolve into remarks on conceptual paths.
E	117.4 and 120.1 are coded personal remarks.
E++	167.2–172.2 continue the remarks on "conceptual paths", many are marked with the curved "S" section marked, indicating Wittgenstein's dissatisfaction.
E++	176.4–184.2 continue the remarks on concept formation through proofs.
E	181.2 is the draft of a letter to Yorick Smithies.
E++	189.2–193.2 discuss the cogency of axioms.
E++	195.2–198.2 continue the discussion of the cogency of proofs.
R++	198.3–200.2 appear before the remark published as §50, but were published as §52.
E++	200.3–202.4 discuss the relation of concept and proof.
E++	207.2–237.2 are practically all unpublished, except for a few remarks in *Culture and Value* and three remarks in *RFM V*.
R+	229.4–230.2 appear after the remark published as §53, but were published as §51.


W-RFM Part V
1
10
20
30
40
53
Ms-126
33
58
86
113
141
Ms-127
31
57
84
109
132
152
172
199
223
Ms-124
18
39
64
82
102
122
141
159
175
190
212
239
270
Section of W-RFM Part V
Remark published in W-RFM Part V
Remark published in other work
Same or similar remarks
Separating line between remarks
More than 30 days between remarks

part vi

§§1-49: Ms-164 According to the editors, *RFM VI* is the only part that corresponds to a *Nachlass* document *in toto*, namely Ms-164. The visualisation confirms this statement, as there are only a few very slight differences between the published work and Ms-164, which can be listed in detail:

R 19.1 are reordered in §7. The latter remark is a restatement and continuation (marked with Wittgenstein's three dashes "– – –") of Ms-164, 17.4 and it makes sense that the editors merged these remarks by moving Ms-164, 19.1 slightly.

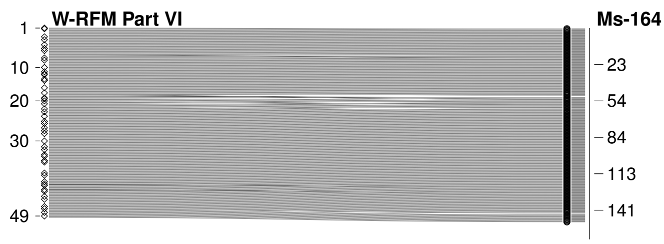
R, E 52.3–53.2 (§17) are another case of such a reordering, where additionally one remark was left out in the published version. Ms-164, 52.3 and Ms-164, 53.2 are a continuation (again marked using "– – –") of Ms-164, 50.4 and are thus reordered appropriately in the published work. Ms-164, 51.3, however, is simply left out. It is not immediately clear why: The remark is crossed out, but so are the remarks before and after it (which are published) and there is no section mark.

R 55.3 is reordered slightly, following Wittgenstein's own indications (with arrows).

E 59.3 is an unfinished remark that Wittgenstein crossed out with multiple strokes, which is why the remark was understandably dropped from the published work.

E 144.2 is not published, but is marked as separate by Wittgenstein through the use of vertical bars at the beginning and end of the remark and put inside quotation marks: | "Ich habe jetzt eingesehen: schlechte Augen sind ebenso gut als gute Augen." |

E++ 152.1–171.3 (24 remarks in total) are concerned with the notion of a private language (Page 153 ends with an explicit "Private Sprache" at the bottom of the page.) and were thus not published in *RFM VI*. The different parts are not clearly separated, however, in fact Wittgenstein moves rather seamlessly from "Übereinstimmung der Meinungen" (in logic, Ms-164, 151.2, published) over "Kriterien dafür [...] daß Einer eine Meinung hat" (Ms-164, 152.1, not published) to "ist eine Meinung haben ein Bewußtseinszustand" (Ms-164, 154.1, directly after the "Private Sprache" remark, not published). This makes sense, given the intimate connection between his remarks on rule following all throughout the published part of Ms-164, and the private language argument presented in the unpublished passage.



◇ Section of W-RFM Part VI
○ Remark published in W-RFM Part VI
∘ Remark published in other work
\ Same or similar remarks
— Separating line between remarks
······ More than 30 days between remarks

part vii

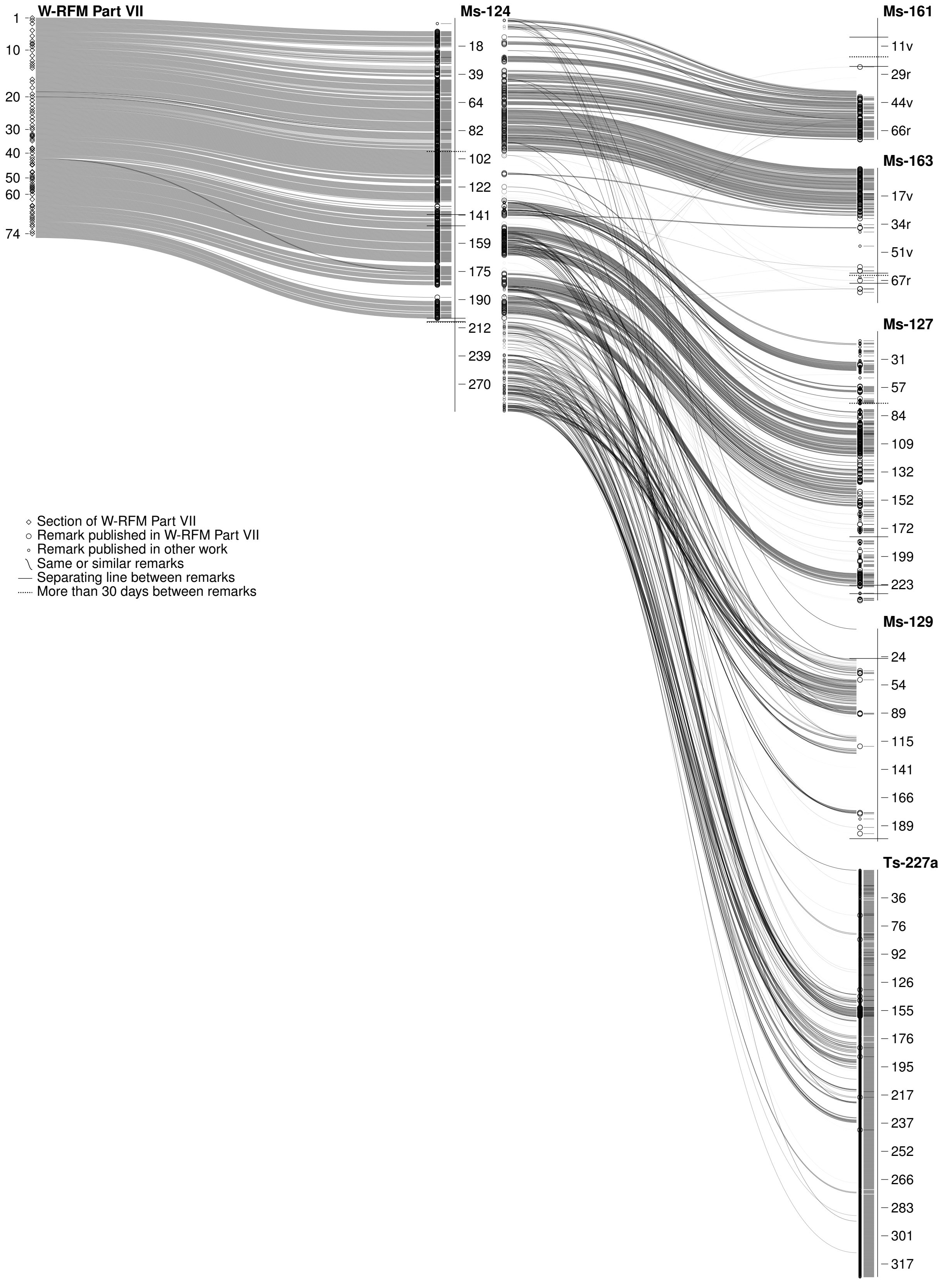
§§1–74: Ms-124

RFM VII contains remarks selected entirely from Ms-124 and shares identical or similar remarks with more documents than any other manuscript in *RFM*. Ms-124 was not written continuously, but with a break of almost three years occurring on page 96. The first part, written in 1941, contains a large portion of remarks that originate in the pocket notebooks Ms-161 and Ms-163, while the later parts, written in 1944, overlap not only with Ms-127 (published primarily as *RFM V*) and Ms-129, but also represent the origin of a number of remarks that later found their way into the *Philosophical Investigations* (Ts-227a). Due to the large number of excluded remarks, the following list will mention only exclusions of 5 or more remarks.

E++	1.1–7.2 discuss the difference between "Geneigt sein zu sagen" and "vouloir dire", expressions of pain, the inner and outer.
E++	19.2–23.2 are remarks about expressions of pain and behaviourism.
E++	31.2–32.5 discuss intuition and axioms.
E++	41.2–44.2 continue the preceding remarks and introduce the idea of following a rule "only once in life".
R, R, R	81.3, 82.4, 84.2 were all reordered and moved to the end of §18 before remark 81.2.
E++	120.4–121.5 discuss the role of empirical and mathematical propositions.
E++	133.3–134.6 are remarks on definitions as concept determinations.
E++	169.5–170.6 discuss rule following and the description of language games.
R+	175.2–175.3 are reordered to appear after 139.2 (§41). There is an explicit arrow by Wittgenstein connecting the following remark, 176.1, with 175.1.
E++	184.1–188.6, 189.3–191.2 are remarks on infinite series, "u.s.w." and rule following.

The rest of Ms-124 is clearly separated from the preceding remarks and focuses mostly on non-mathematical topics, although the remarks on rule following can certainly be read as a continuation of some earlier lines of thought in Ms-124. The remarks on private language are also not without precedence in the mathematical part of Ms-124, which contains several remarks that appear in the *PI*.

E++	200.3–202.3 are separated from the preceding and following remarks by a horizontal line and discuss rule following.
E++	205.1–292.2 are written nearly three months later than the preceding remarks and continue the discussion of rule following, private language and other topics, with numerous remarks corresponding to remarks in the *PI* between Ts-227a, 143.2 (§198) and Ts-227a, 236.4 (§421).


W-RFM Part VII
Ms-124
Ms-161
Ms-163
Ms-127
Ms-129
Ts-227a
1
10
20
30
40
50
60
74
18
39
64
82
102
122
141
159
175
190
212
239
270
11v
29r
44v
66r
17v
34r
51v
67r
31
57
84
109
132
152
172
199
223
24
54
89
115
141
166
189
36
76
92
126
155
176
195
217
237
252
266
283
301
317
Section of W-RFM Part VII
Remark published in W-RFM Part VII
Remark published in other work
Same or similar remarks
Separating line between remarks
More than 30 days between remarks

WRITINGS AND LECTURES BY WITTGENSTEIN

BGM *Bemerkungen über die Grundlagen der Mathematik. Werkausgabe Band 6.* Ed. by G. E. M. Anscombe, Rush Rhees, G. H. von Wright. Frankfurt a. M.: Suhrkamp, 2013.

BPP "Bemerkungen über die Philosophie der Psychologie" In: *Bemerkungen über die Philosophie der Psychologie. Werkausgabe Band 7. Bemerkungen über die Philosophie der Psychologie. Letzte Schriften über die Philosophie der Psychologie.* Ed. by G. E. M. Anscombe, G. H. von Wright. Frankfurt a. M.: Suhrkamp, 2014.

GB "Bemerkungen über Frazers 'Golden Bough'" / "Remarks on Frazer's 'Golden Bough'". In: *Philosophical Occasions. 1912–1951.* Ed. by James C. Klagge, Alfred Nordmann, pp. 115–155. Indianapolis & Cambridge: Hackett Publishing Company, 1993.

LFM *Wittgenstein's Lectures on the Foundations of Mathematics. Cambridge, 1939. From the Notes of R. G. Bosanquet, Norman Malcolm, Rush Rhees, and Yorick Smythies.* Ed. by Cora Diamond. Chicago & London: University of Chicago Press, 1976.

OC *On Certainty.* Ed. by G. E. M. Anscombe, G. H. von Wright. Oxford: Basil Blackwell, 1969.

PB *Philosophische Bemerkungen. Werkausgabe Band 2.* Ed. by Rush Rhees. Frankfurt a. M.: Suhrkamp, 2015.

PI "Philosophical Investigations". In: *Philosophische Untersuchungen. Philosophical Investigations.* Ed. by P. M. S. Hacker, Joachim Schulte, pp. 1–181. Chichester, West Sussex: Wiley-Blackwell, 2009.

PPF "Philosophie der Psychologie. Ein Fragment" / "Philosophy of Psychology. A Fragment". In: *Philosophical Investigations.* Ed. by P. M. S. Hacker, Joachim Schulte, pp. 182–243. Chichester, West Sussex: Wiley-Blackwell, 2009.

PR *Philosophical Remarks.* Ed. by Rush Rhees. Oxford: Basil Blackwell, 1975.

PU "Philosophische Untersuchungen". In: *Philosophische Untersuchungen. Philosophical Investigations.* Ed. by P. M. S. Hacker, Joachim Schulte, pp. 1–181. Chichester, West Sussex: Wiley-Blackwell, 2009.

RFM *Remarks on the Foundation of Mathematics.* Ed. by G. H. von Wright, R. Rhees, G. E. M. Anscombe. Oxford: Basil Blackwell, 1978.

RPP *Remarks on the Philosophy of Psychology.* Ed. by G. E. M. Anscombe, G. H. von Wright. Oxford: Basil Blackwell, 1980.

TLP *Tractatus Logico-Philosophicus.* Ed. by C. K. Ogden, translated by C. K. Ogden und F. P. Ramsey. International Library of Psychology, Philosophy and Scientific Method. London: Routledge and Kegan Paul, 1955.

WCL *Wittgenstein's Whewell's Court Lectures. Cambridge, 1938–1941. From the Notes by Yorick Smythies.* Ed. by Volker A. Munz, Bernhard Ritter. Chichester, West Sussex: Wiley-Blackwell, 2017.

Z "Zettel". In: *Über Gewißheit. Werkausgabe Band 8. Bemerkungen über die Farben. Über Gewißheit. Zettel. Vermischte Bemerkungen.* Ed. by G. E. M. Anscombe, G. H. von Wright. Frankfurt a. M.: Suhrkamp, 2015.

Baker, G. P. and P. M. S. Hacker (2005a). *Wittgenstein: Understanding and Meaning. Volume 1 of An Analytical Commentary on the Philosophical Investigations. Part II: Exegesis §§1–184*. Wiley-Blackwell.

— (2005b). *Wittgenstein: Understanding and Meaning. Volume 1 of An Analytical Commentary on the Philosophical Investigations. Part I: Essays*. Wiley-Blackwell.

Bays, Timothy (2004). "On Floyd and Putnam on Wittgenstein on Gödel." In: *The Journal of philosophy* 101.4, pp. 197–210.

— (2006). "Floyd, Putnam, Bays, Steiner, Wittgenstein, Gödel, Etc." In: *The Journal of Philosophy* 103, pp. 101–110.

Bennett, Charles H. (1979). "On Random and Hard-to-Describe Numbers." In: *Randomness and Complexity: From Leibniz to Chaitin*. Ed. by Cristian S. Calude. World Scientific, pp. 3–12.

Bernays, Paul (1959). "Comments on Ludwig Wittgenstein's *Remarks on the Foundations of Mathematics*." In: *Ratio* 2.1, pp. 1–22.

Berto, Francesco (2009). *There's Something About Gödel: The Complete Guide to the Incompleteness Theorem*. John Wiley and Sons.

Biggs, Michael and Alois Pichler (1993). "Wittgenstein: Two Source Catalogues and a Bibliography. Catalogues of the Published Texts and of the Published Diagrams, each Related to its Sources." In: *Working Papers from the Wittgenstein Archives at the University of Bergen* 7, p. 66.

Borges, Jorge Luis (1964). *Labyrinths: Selected Stories & Other Writings*. Ed. by Donald A. Yates and James E. Irby. New Directions paperbook. New Directions.

Brusotti, Marco (2014). *Wittgenstein, Frazer und die "ethnologische Betrachtungsweise"*. Vol. 2. Über Wittgenstein. De Gruyter.

Calude, Cristian S., ed. (2007). *Randomness and Complexity: From Leibniz to Chaitin*. World Scientific.

Chaitin, Gregory J. (1966). "On the Length of Programs for Computing Finite Binary Sequences." In: *Journal of the ACM (JACM)* 13.4, pp. 547–569.

— (1970). "On the Difficulty of Computations." In: *Information, Randomness & Incompleteness. Papers on Algorithmic Information Theory*. 2nd ed. Vol. 8. World Scientific, pp. 41–56.

— (1974). "Information Theoretic Computational Complexity." In: *Information, Randomness & Incompleteness. Papers on Algorithmic Information Theory*. 2nd ed. Vol. 8. World Scientific, pp. 57–73.

— (1975). "Randomness and Mathematical Proof." In: *Information, Randomness & Incompleteness. Papers on Algorithmic Information Theory*. 2nd ed. Vol. 8. World Scientific, pp. 11–28.

Chaitin, Gregory J. (1982a). "Algorithmic Information Theory." In: *Information, Randomness & Incompleteness. Papers on Algorithmic Information Theory*. 2nd ed. Vol. 8. World Scientific, pp. 75–82.

— (1982b). "Gödel's Theorem and Information." In: *Information, Randomness & Incompleteness. Papers on Algorithmic Information Theory*. 2nd ed. Vol. 8. World Scientific, pp. 111–130.

— (1988). "Randomness in Arithmetic." In: *Information, Randomness & Incompleteness. Papers on Algorithmic Information Theory*. 2nd ed. Vol. 8. World Scientific, pp. 29–39.

— (1989). "Undecidability and Randomness in Pure Mathematics." In: *Information, Randomness & Incompleteness. Papers on Algorithmic Information Theory*. 2nd ed. Vol. 8. World Scientific, pp. 503–515.

— (1990). *Information, Randomness & Incompleteness. Papers on Algorithmic Information Theory*. 2nd ed. Vol. 8. World Scientific.

— (1993). "Randomness in arithmetic and the decline & fall of reductionism in pure mathematics." In: *Thinking about Gödel and Turing: Essays on Complexity, 1970-2007*. World Scientific Publishing Company, pp. 75–98.

— (1995). "The Berry paradox." In: *Complexity* 1 (1), pp. 26–30.

— (2000). "A century of controversy over the foundations of mathematics." In: *Thinking about Gödel and Turing: Essays on Complexity, 1970-2007*. World Scientific Publishing Company, pp. 129–151.

— (2002). "Paradoxes of randomness." In: *Thinking about Gödel and Turing: Essays on Complexity, 1970-2007*. World Scientific Publishing Company, pp. 169–188.

— (2003). "Two philosophical applications of algorithmic information theory." In: *Thinking about Gödel and Turing: Essays on Complexity, 1970-2007*. World Scientific Publishing Company, pp. 189–200.

— (2004). "Leibniz, information, math & physics." In: *Thinking about Gödel and Turing: Essays on Complexity, 1970-2007*. World Scientific Publishing Company, pp. 227–239.

— (2007). *Thinking about Gödel and Turing: Essays on Complexity, 1970-2007*. World Scientific Publishing Company.

— (2008). "Irreducible Complexity in Pure Mathematics." In: *Wittgenstein and the Philosophy of Information. Proceedings of the 30th International Ludwig Wittgenstein-Symposium*. Ed. by Alois Pichler and Herbert Hrachovec. De Gruyter.

Chaitin, Gregory J., Asat Arslanov, and Cristian Calude (1995). *Program-size complexity computes the halting problem*. Tech. rep. Department of Computer Science, The University of Auckland, New Zealand.

Chown, Marcus (2007). "God's Number: Where Can We Find the Secret of the Universe? In a Single Number!" In: *Randomness and Complexity: From Leibniz to Chaitin*. Ed. by Cristian S. Calude. World Scientific, pp. 321–342.

Cover, Thomas M. and Joy A. Thomas (2006). *Elements of Information Theory*. 2nd ed. Wiley Series in Telecommunications and Signal Processing. Wiley-Interscience.

Davis, Martin (1985). *Computability and Unsolvability*. New edition. Mcgraw-Hill Series in Information Processing and Computers. Dover Publications.

Dummett, Michael (1959). "Wittgenstein's Philosophy of Mathematics." In: *The Philosophical Review* 68.3, pp. 324–348.

— (1970). "Wang's Paradox." In: *Truth and Other Enigmas*. Harvard University Press, pp. 248–268.

Floyd, Juliet (1995). "On saying what you really want to say: Wittgenstein, Gödel, and the trisection of the angle." In: *From Dedekind to Gödel*. Springer, pp. 373–425.

— (2001). "Prose versus Proof: Wittgenstein on Gödel, Tarski and Truth." In: *Philosophia Mathematica* 9.3, pp. 280–307.

— (2012). "Wittgenstein's Diagonal Argument: A Variation on Cantor and Turing." In: *Epistemology versus Ontology*. Ed. by P. Dybjer, Sten Lindström, Erik Palmgren, and G. Sundholm. Dordrecht: Springer Netherlands, pp. 25–44.

— (2016). "Chains of life: Turing, Lebensform, and the Emergence of Wittgenstein's Later Style." In: *Nordic Wittgenstein Review* 5.2, pp. 7–89.

— (2017). "Turing on "Common Sense": Cambridge Resonances." In: *Philosophical Explorations of the Legacy of Alan Turing*. Springer, pp. 103–149.

— (2018). "Lebensformen: Living Logic." In: *Language, Form(s) of Life, and Logic: Investigations After Wittgenstein*. Ed. by Christian Martin. Berlin: De Gruyter, pp. 59–92.

— (2019). "Wittgenstein and Turing." In: *Philosophy of Logic and Mathematics. Proceedings of the 41st International Ludwig Wittgenstein-Symposium*. Ed. by Gabriele M. Mras, Paul Weingartner, and Bernhard Ritter. De Gruyter, pp. 263–296.

— (2020). "The Generality of Cantor's Diagonal Procedure." In: *Wittgenstein's Annotations to Hardy's Course of Pure Mathematics: An Investigation of Wittgenstein's Non-Extensionalist Understanding of the Real Numbers*. Vol. 7. Nordic Wittgenstein Studies. Cham: Springer International Publishing, pp. 192–258.

Floyd, Juliet and Felix Mühlhölzer (2020). *Wittgenstein's Annotations to Hardy's Course of Pure Mathematics: An Investigation of Wittgenstein's Non-Extensionalist Understanding of the Real Numbers*. Vol. 7. Nordic Wittgenstein Studies. Cham: Springer International Publishing.

Floyd, Juliet and Hilary Putnam (2000). "A Note on Wittgenstein's "Notorious Paragraph" about the Gödel Theorem." In: *The Journal of Philosophy* 97.11, pp. 624–632.

Floyd, Juliet and Hilary Putnam (2006). "Bays, Steiner, and Wittgenstein's" Notorious" Paragraph about the Gödel Theorem." In: *The Journal of philosophy*, pp. 101–110.

— (2008). "Wittgensteins ‚berüchtigter' Paragraph über das Gödel-Theorem: Neuere Diskussionen." In: *Prosa oder Beweis? Wittgenstein's 'berüchtigte' Bemerkungen zu Gödel. Texte und Dokumente*. Ed. by Esther Ramharter. Berlin: Parerga, pp. 75–97.

Franzén, Torkel (2005). *Gödel's Theorem: An Incomplete Guide to Its Use and Abuse*. Wellesley, Massachusetts: A K Peters.

Frascolla, Pasquale (2006). *Wittgenstein's Philosophy of Mathematics*. Routledge.

Gardner, Martin (1979). "MATHEMATICAL GAMES. The random number omega bids fair to hold the mysteries of the universe." In: *Scientific American* 241.5, pp. 20–35.

Gefwert, Christoffer (1998). *Wittgenstein on Mathematics, Minds and Mental Machines*. Burlington: Ashgate Publishing.

Gödel, Kurt (1986). *Collected Works*. Ed. by Solomon Feferman, John W. Dawson, Jr., Stephen C. Kleene, Gregory H. Moore, Robert M. Solovay, and Jean van Heijenoort. Oxford : New York: Clarendon Press ; Oxford University Press.

Hopcroft, John E., Rajeev Motwani, and Jeffrey D. Ullman (2001). *Introduction to automata theory, languages, and computation*. 2nd ed. Addison-Wesley.

Kienzler, Wolfgang and Sebastian Sunday Grève (2016). "Wittgenstein on Gödelian 'Incompleteness', Proofs and Mathematical Practice: Reading Remarks on the Foundations of Mathematics, Part I, Appendix III, Carefully." In: *Wittgenstein and the Creativity of Language*. Ed. by Sebastian Sunday Grève and Jakub Mácha. Springer, pp. 76–116.

Knuth, Donald E (1976). "Mathematics and Computer Science: Coping with Finiteness: Advances in our ability to compute are bringing us substantially closer to ultimate limitations." In: *Science* 194.4271, pp. 1235–1242.

Kolmogorov, Andrei N. (1968a). "Logical Basis for Information Theory and Probability Theory." In: *IEEE Transactions on Information Theory* 14.5, pp. 662–664.

— (1968b). "Three Approaches to the Quantitative Definition of Information." In: *International Journal of Computer Mathematics* 2.1-4, pp. 157–168.

Kreisel, Georg (1958). *Wittgenstein's Remarks on the Foundations of Mathematics*. In: *The British Journal for the Philosophy of Science* vol. 9.34, pp. 135–158.

Lajevardi, Kaave (2021). "What he could have said (but did not say) about Gödel's second theorem: A note on Floyd-Putnam's Wittgenstein." In: *Wittgenstein-Studien* 12.1, pp. 121–129.

Li, Ming and Paul Vitányi (2019). *An introduction to Kolmogorov Complexity and its Applications*. 4th ed. Springer.
Lucas, J. R. (1961). "Minds, Machines and Gödel." In: *Philosophy* 36.137, pp. 112–127.
Majetschak, Stefan (2010). "Lebensformen und Lebensmuster. Zur Deutung eines sogenannten Grundbegriffs der Spätphilosophie Ludwig Wittgensteins." In: *Language and World. Part One: Essays on the Philosophy of Wittgenstein, Publications of the Austrian Ludwig Wittgenstein Society. New Series*. Ed. by Volker Munz, Klaus Puhl, and Joseph Wang. Vol. 14. Paris; Lancaster; New Brunswick: Heusenstamm: ontos verlag, pp. 265–290.
— (2012). "Die anthropologische Betrachtungsweise. Zum Einfluss von James George Frazers *The Golden Bough* auf die Entwicklung der Spätphilosophie Ludwig Wittgensteins." In: *Wittgenstein-Studien* 3, pp. 217–232.
— (2016). "Survey and Surveyability. Remarks on two central notions in Wittgenstein's later philosophy." In: *Wittgenstein-Studien* 7, pp. 65–80.
— (2020). "'Eine irreführende Parallele'. Wittgenstein über Begriffsverwirrung in der Psychologie und die Semantik psychologischer Begriffe." In: *Wittgenstein-Studien* 11, pp. 169–182.
Marion, Mathieu (1998). *Wittgenstein, Finitism, and the Foundations of Mathematics*. Oxford: Oxford University Press.
Matthíasson, Ásgeir Berg (2021). "Contradictions and falling bridges: what was Wittgenstein's reply to Turing?" In: *British Journal for the History of Philosophy* 29.3, pp. 537–559.
Monk, Ray (1991). *Ludwig Wittgenstein: The Duty of Genius*. New York: Penguin Books.
Moyal-Sharrock, Danièle (2015). "Wittgenstein on Forms of Life, Patterns of Life, and Ways of Living." In: *Nordic Wittgenstein Review* Special Issue 2015, Wittgenstein and Forms of Life, pp. 21–42.
Mühlhölzer, Felix (2010). *Braucht die Mathematik eine Grundlegung? Ein Kommentar des Teils III von Wittgenstein's Bemerkungen über die Grundlagen der Mathematik*. Frankfurt/Main: Vittorio Klostermann.
— (2020). "Wittgenstein on Cantor's Diagonal Method." In: *Wittgenstein's Annotations to Hardy's Course of Pure Mathematics: An Investigation of Wittgenstein's Non-Extensionalist Understanding of the Real Numbers*. Vol. 7. Nordic Wittgenstein Studies. Cham: Springer International Publishing, pp. 125–191.
Penrose, Roger (1991). *The Emperor's New Mind: Concerning Computers, Minds, and the Laws of Physics*. Penguin Books.
— (1994). *Shadows of the Mind: A Search for the Missing Science of Consciousness*. Oxford; New York: Oxford University Press.
Persichetti, Alessio (2021). "The later Wittgenstein's guide to contradictions." In: *Synthese* 198.4, pp. 3783–3799.

Post, Emil (1947). "On Computable Numbers, with an Application to the Entscheidungsproblem. A Critique." In: *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life, plus The Secrets of Enigma*. Oxford; New York: Clarendon Press; Oxford University Press, pp. 97–101.

Priest, Graham (1995). *Beyond the Limits of Thought*. Cambridge; New York: Cambridge University Press.

— (2004). "Wittgenstein's Remarks on Gödel's Theorem." In: *Wittgenstein's Lasting Significance*. Ed. by Max Kölbel and Bernhard Weiss. London; New York: Routledge, pp. 207–227.

— (2006a). *Doubt Truth to be a Liar*. Oxford: Oxford University Press.

— (2006b). *In Contradiction: a Study of the Transconsistent*. Oxford; New York: Clarendon Press; Oxford University Press.

Putnam, Hilary (2007). "Wittgenstein and the Real Numbers." In: *Wittgenstein and the Moral Life: Essays in Honor of Cora Diamond*. Ed. by Alice Crary, pp. 235–250.

Ramharter, Esther (2008). "Gödel und Wittgenstein — ein gleiches Paar?" In: *Prosa oder Beweis? Wittgenstein's 'berüchtigte' Bemerkungen zu Gödel. Texte und Dokumente*. Ed. by Esther Ramharter. Berlin: Parerga, pp. 7–19.

— (2010). "Are all contradictions equal? Wittgenstein on confusion in mathematics." In: *Philosophy of Mathematics: Sociological Aspects and Mathematical Practice*. Ed. by Th. Müller B. Löwe. London: College Publications, pp. 293–306.

— (2014). "Wittgenstein über vollständige Induktion." In: *Wittgenstein-Studien* 5, pp. 179–203.

— (2018). "In der Luft verankert? – Wittgenstein und Cantor zur Kardinalzahlarithmetik." In: *Wittgenstein und die Philosophie der Mathematik*. Ed. by Joachim Bromand and Bastian Reichardt. Paderborn: mentis, pp. 129–142.

— (2019). "Der Status mathematischer und religiöser Sätze bei Wittgenstein." In: *Philosophy of Logic and Mathematics. Proceedings of the 41st International Ludwig Wittgenstein-Symposium*. Ed. by Gabriele M. Mras, Paul Weingartner, and Bernhard Ritter. De Gruyter, pp. 485–493.

Redecker, Christine (2006). *Wittgensteins Philosophie der Mathematik: Eine Neubewertung im Ausgang von der Kritik an Cantors Beweis der Überabzählbarkeit der reellen Zahlen*. Vol. 9. De Gruyter.

Rodych, Victor (1999). "Wittgenstein's inversion of Gödel's Theorem." In: *Erkenntnis (1975-)* 51.2/3, pp. 173–206.

— (2000). "Wittgenstein's Critique of Set Theory." In: *The Southern Journal of Philosophy* 38.2, pp. 281–319.

— (2002). "Wittgenstein on Gödel: the newly published remarks." In: *Erkenntnis (1975-)* 56.3, pp. 379–397.

— (2003). "Misunderstanding Gödel: New Arguments about Wittgenstein and New Remarks by Wittgenstein." In: *Dialectica* 57.3, pp. 279–313.

— (2006). "Who is Wittgenstein's Worst Enemy? Steiner on Wittgenstein on Gödel." In: *Logique et Analyse* 49.193, pp. 55–84.

Rozenberg, Grzegorz and Arto Salomaa (2007). "The Secret Number. An Exposition of Chaitin's Theory." In: *Randomness and Complexity: From Leibniz to Chaitin*. Ed. by Cristian S. Calude. World Scientific, pp. 175–215.

Russell, Bertrand (1908). "Mathematical logic as based on the theory of types." In: *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*. Source Books in History of Sciences. Harvard University Press.

Schulte, Joachim (1990). "Chor und Gesetz. Zur ‚morphologischen Methode' bei Goethe und Wittgenstein." In: *Wittgenstein im Kontext*. Frankfurt/Main: Suhrkamp, pp. 11–42.

— (2021). "Philosophische Superlative und die Maschine als Symbol." In: *Wittgenstein-Studien* 12.1, pp. 1–36.

Shanker, Stuart (1987). *Wittgenstein and the Turning-Point in the Philosophy of Mathematics*. Albany, New York: State University of New York Press.

— (1988). "Wittgenstein's Remarks on the Significance of Gödel's Theorem." In: *Gödel's Theorem in Focus*. Routledge, pp. 155–256.

— (1998). *Wittgenstein's Remarks on the Foundations of AI*. London; New York: Routledge.

Shannon, Claude E. (1948). "A Mathematical Theory of Communication." In: *The Bell System Technical Journal* 27.3, pp. 379–423.

Sipser, Michael (2012). *Introduction to the Theory of Computation*. 3rd ed. Thomson South-Western.

Solomonoff, Ray J. (1964a). "A Formal Theory of Inductive Inference. Part II." In: *Information and Control* 7.2, pp. 224–254.

— (1964b). "A formal theory of inductive inference. Part I." In: *Information and control* 7.1, pp. 1–22.

Steiner, Mark (2001). "Wittgenstein as his Own Worst Enemy: The Case of Gödel's Theorem." In: *Philosophia Mathematica* 9.3, pp. 257–279.

Stern, David G (2004). *Wittgenstein's Philosophical Investigations: An Introduction*. Cambridge University Press.

Trächtler, Jasmin (2020). "Wittgenstein on 'Imaginability' as a Criterion for Logical Possiblity in 'The Big Typescript'." In:

— (2021). *Wittgensteins Grammatik des Fremdseelischen*. Bergen: Skipnes Kommunikasjon / University of Bergen.

Turing, Alan (1936). "On Computable Numbers, with an Application to the Entscheidungsproblem." In: *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and*

Artificial Life, plus The Secrets of Enigma. Oxford; New York: Clarendon Press; Oxford University Press, pp. 58–90.

Turing, Alan (1938). "Systems of Logic Based on Ordinals." In: *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life, plus The Secrets of Enigma*. Oxford; New York: Clarendon Press; Oxford University Press, pp. 146–204.

— (1940). "Letters on Logic to Max Newman." In: *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life, plus The Secrets of Enigma*. Oxford; New York: Clarendon Press; Oxford University Press, pp. 211–216.

— (1947). "Lecture on the Automatic Computing Engine." In: *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life, plus The Secrets of Enigma*. Oxford; New York: Clarendon Press; Oxford University Press, pp. 378–394.

— (1948). "Intelligent Machinery." In: *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life, plus The Secrets of Enigma*. Oxford; New York: Clarendon Press; Oxford University Press, pp. 410–432.

— (1950). "Computing Machinery and Intelligence." In: *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life, plus The Secrets of Enigma*. Oxford; New York: Clarendon Press; Oxford University Press, pp. 441–464.

— (1951). "Intelligent Machinery, A Heretical Theory." In: *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life, plus The Secrets of Enigma*. Oxford; New York: Clarendon Press; Oxford University Press, pp. 472–475.

— (1953). "Chess." In: *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life, plus The Secrets of Enigma*. Oxford; New York: Clarendon Press; Oxford University Press, pp. 569–575.

Turing, Alan and B. Jack Copeland (2004). *The Essential Turing: Seminal Writings in Computing, Logic, Philosophy, Artificial Intelligence, and Artificial Life, plus The Secrets of Enigma*. Oxford; New York: Clarendon Press; Oxford University Press.

Von Wright, Georg Henrik (1993). "The Wittgenstein Papers." In: *Philosophical Occasions: 1912-1951*. Ed. by James Klagge and Alfred Nordmann, pp. 480–506.

Watson, Alister (1938). "Mathematics and its Foundations." In: *Mind* XLVII.188, pp. 440–451.

Wheeler, Samuel J. (2021). "Defending Wittgenstein's Remarks on Cantor from Putnam." In: *Philosophical Investigations*, Online Version of Record before inclusion in an issue, First published 12 November 2021.

Wrigley, Michael (1980). "Wittgenstein on Inconsistency." In: *Philosophy* 55.214, pp. 471–484.

Yanofsky, Noson S. (2003). "A Universal Approach to Self-Referential Paradoxes, Incompleteness and Fixed Points." In: *Bulletin of Symbolic Logic* 9.3, pp. 362–386.

DECLARATION

I herewith give assurance that I completed this dissertation independently without prohibited assistance of third parties or aids other than those identified in this dissertation. All passages that are drawn from published or unpublished writings, either word-for-word or in paraphrase, have been clearly identified as such. Third parties were not involved in the drafting of the content of this dissertation; most specifically I did not employ the assistance of a dissertation advisor. No part of this thesis has been used in another doctoral or tenure process.

Berlin, Februar 2022

Frederic Kettelhoit