Faculty of Arts
University of Helsinki

GÖDEL ON INTUITIONISM AND CONSTRUCTIVE FOUNDATIONS OF MATHEMATICS

Maria Hämeen-Anttila

DOCTORAL DISSERTATION

To be presented for public discussion with the permission of the Faculty of Arts of the University of Helsinki, in Lecture Hall 107, Athena, on the 2nd of November, 2020 at 12 o'clock.

ISBN 978-951-51-5922-9 (nid.)
ISBN 978-951-51-5923-6 (PDF)
Unigrafia
Helsinki 2020

Abstract

Kurt Gödel's early views on intuitionism and constructive foundations of mathematics became publicly known in his three posthumously published lectures: "The present situation in the foundations of mathematics" (1933), "Lecture at Zilsel's" (1938), and "In what sense is intuitionistic logic constructive?" (1941). The aim of the study is to examine these works in light of Gödel's unpublished notes to construct a more detailed picture of his views. A wide selection of materials, including so far unpublished lecture notes as well as shorthand notes on mathematics and philosophy primarily from 1940–1941, was studied for this purpose.

The analysis shows three phases in the development of Gödel's foundational views. Gödel's earliest studies in intuitionistic logic focused on its classical interpretations, and these shaped his belief that the found interconnections between intuitionistic and classical logics revealed something suspicious about intuitionism. The second phase, comprised of the 1933 and 1938 lectures, is shown to be characterised by a firm belief in formalisation and the Hilbert Programme. There is a strong parallelism between Gödel's and Hilbert's critique of intuitionism, and Gödel also agreed with Hilbert on the need for constructive consistency proofs for mathematical systems. In the 1938 lecture, it is suggested that a system based on functionals of finite types could fulfil this purpose while remaining properly finitary.

The third phase begins with the lecture course in Princeton, where the functional system is finally introduced. However, it is now presented as an interpretation and a proof of constructivity of intuitionistic arithmetic, not as a finitistic consistency proof. In his notes written around the same time, Gödel started to reconsider the relationship between classical and intuitionistic systems, considering intuitionistic interpretations of classical logic as well as interpretations of intuitionism in its own terms. This gradual resignation of the formalistic viewpoint coincides with several failed formal endeavours and a slow turn towards philosophy.

Acknowledgements

During the writing of this dissertation, I have been blessed to receive advice, help, and support from a host of different people. Had the book even been completed without them, it would have looked, to say the least, very different. I offer my deepest gratitude to each and every one of them.

First and foremost, I wish to thank my supervisor and mentor Jan von Plato. It was his lecture course on history of arithmetic in spring 2016 that made me decide to become a historian. Some months later, he agreed to supervise my dissertation. I did not expect that this first meeting would lead me to learn, in addition to a good amount of logic, two new languages and Gabelsberger shorthand. Not only has his experience and knowledge been invaluable to me but he has also given me the space to grow and find my voice as a young researcher, having faith in me even when I did not have faith in myself.

I am most grateful to the external examiners of my dissertation, Göran Sundholm and Mark van Atten, for their valuable comments and suggestions. Moreover, their works on history of logic, Gödel, and intuitionism have greatly inspired me in my own research.

I wish to thank my teachers in logic and philosophy of mathematics. Gabriel Sandu taught me how to think and write philosophically; when I first decided to specialise in logic, he patiently guided me through the labyrinth of intermediate logic. Sara Negri's course on proof theory in 2015 inspired me to learn more about the field which I barely knew at the time. I am thankful for her support in my early (somewhat simple) works on proof theory and the encouragement to continue my studies in logic.

It has been a privilege to be able to work with my colleagues Annika Kanckos and Tim Lethen. From the beginning, I have admired Annika's skill in logic and mathematics; in spring 2019, we held a reading group in ordinal proof theory, which is probably one of the hardest things I have ever done (and a plenty of fun). I am thankful for her advice and support. Tim has shown incredible patience helping me with shorthand reading. I would be envious of his ability to read and interpret Gödel so beautifully if he did not happen to be such a warm and generous person.

I am thankful to the various people who have supported me and helped to improve my work on Gödel and intuitionism. I am especially thankful to Juliette Kennedy for some beautiful conversations as well as encouraging me to widen my philosophical horizon. I am grateful for Panu Raatikainen's insightful comments on several papers, which have helped me to refine and rethink my views. Mirja Hartimo's comments have also turned invaluable to my work on Gödel's philosophical views. Fernando Ferreira helped me with Gödel's functional interpretation, and I am thankful to him for answering all of my many questions. Matthias Baaz, William Howard, Karl Sigmund, and Friedrich Stadler have all been very kind to answer my questions concerning various details of Gödel's life and intellectual development. I greatly value the discussions with John and Cheryl Dawson, which not only gave me a wealth of information on Gödel but also have encouraged me to continue, if one may say so, in their footsteps, making Gödel's unpublished papers available to the wider academic community.

I am grateful to Edi Pavlović for logical help as well as heated political discussions and Sanna Mattila for (less heated) discussions on logic and metaphysics. I am also grateful to Andreas Fjellstad, Jaakko Hirvelä, Tuukka Tanninen, Pii Telakivi, and Tuomas Vesterinen for their support and practical advice.

In 2017, a bunch of post-graduate students from various philosophical backgrounds decided to put together what we came to call our PhD support group. Our meetings have not only inspired me to keep on working on my dissertation, but given me something much more precious: friends that have been there for me both in moments of self-doubt and moments of joy. I wholeheartedly thank Tuukka Brunila, Minna-Kerttu Kekki, Polaris Koi, Mattias Lehtinen, Joonas Martikainen, Sini Pentikäinen, Joni Puranen, and Pieta Päällysaho.

My family's support during my post-graduate studies means more to me than can be expressed in words. I am thankful for my parents' enthusiastic support of my work; they have given me two very different but equally inspiring role models in my academic career. My husband Joonas has shared the dissertation journey with me in a very concrete way, as he is currently finishing his own doctoral thesis. I could not wish for a better travel companion in academia or in life.

Kurt Gödel's unpublished papers are quoted courtesy of the Kurt Gödel Papers, the Shelby White and Leon Levy Archives Center, Institute for Advanced Study, Princeton, NJ, USA, on deposit at Princeton University.

This work has been completed under the research project GODELIANA funded by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 787758).

Contents

Chapter 1

Introduction

Kurt Gödel plays an essential role in the development of constructive mathematics. On the one hand, his incompleteness theorems showed the infeasibility of what is known as the Hilbert Programme, whose aim was to justify classical mathematics by proving its consistency with only elementary, intuitively evident methods. On the other hand, Gödel contributed to both the development of intuitionistic logic and the "extended Hilbert Programme," the aim of which is to extract as much constructive content as possible from a given classical system. Even in his earlier works, Gödel was sensitive to questions of constructivity, and he placed a great deal of importance on some of his works on the topic of constructive foundations.

This side of Gödel was not too well known during his lifetime. In fact, before the publication of the third volume of Gödel's *Collected Works* (Gödel, 1995), his only known works on intuitionism and its logic from the 1930s and 1940s were three short and very formal articles he published in 1932 and 1933. Gödel's only philosophical comments related to intuitionism and constructive foundations could be found from the article "Über eine bisher noch nicht benutzte Erweiterung des finiten Standpunktes," based on results of 1941 but published in 1958 in the journal *Dialectica*. Even the revised version of the article, where Gödel further refined his philosophical views on constructivity, originally meant to be published in 1970, had to wait until the second volume of the *Collected Works* (Gödel, 1990). This type of thing happened relatively often: Gödel's perfectionist tendencies and his obsessive focus on the problems he saw as important – such as the incompleteness theorems, and later the consistency of analysis and the Continuum Hypothesis – resulted in setting many smaller ideas aside.

The third volume of the *Collected Works* was published in 1995 and it contained a selection of Gödel's unpublished lectures and manuscripts. The three lectures Gödel gave on intuitionism in 1933, 1938, and 1941 shed more light on his views. In these lectures, he investigated the concept of constructiveness of a logic, and criticised Heyting's logic for having non-constructive (or less than ideally constructive) elements. In 1933 and 1938, he called for an extension of Hilbert's Programme to obtain consistency proofs for classical theories. In 1941, he devised an interpretation of intuitionistic arithmetic which not only meant to provide a relative proof of consistency for classical arithmetic but also to show that intuitionistic arithmetic is constructive.

The extent to which Gödel devoted his attention to intuitionistic logic and its philosophy is surprising, considering that he had published nothing on the topic during the 1930s and 1940s. Given the consensus on Gödel's mathematical realism which he himself has confirmed, the interest in constructive views was not entirely expected. In his commentary to the 1933 lecture, Feferman calls Gödel's statement that the axioms of classical mathematics presuppose "a kind of Platonism, which cannot satisfy any critical mind" (Gödel, 1933b, 50) "most surprising" (Feferman, 1995, 39). According to Sieg and Parsons (1995, 62), "[a] surprising general conclusion from the three [lectures on intuitionism] is that Gödel in those years was intellectually much closer to the ideas and goals pursued in the Hilbert school than has been generally assumed (or than can be inferred from his own published accounts)." Davis (2005) goes so far as to conclude that this, as well as related remarks in Gödel's early letters to Paul Bernays, refutes Gödel's own statement in 1975 that he was a Platonist since 1925 (Gödel, 2003a, 444).

Compared to the *Dialectica* article of 1958 and its intricate ideas on the notions of proof and computability, Gödel's early thoughts on constructivity and intuitionistic logic have received relatively little attention in the literature. There are very good reasons for this. First of all, apart from the three lectures, there is precious little published material to work with. Gödel's focus in his early works is predominantly mathematical, and his philosophical views never take centre stage, in contrast to his works after 1943, which are mostly philosophical in nature. Another issue could be younger Gödel's style of writing, which is, as Kreisel puts it, "concise and cavalier, apparently scoffing [...] at the antics of the rhetoric" (Kreisel, 1987, 144). Gödel's mathematical reasoning is exact, but several philosophical remarks are simply stated and left unexplained; the “kind of Platonism” remark is one of these cases. The extent to which Gödel *had a philosophy* of mathematics in the 1930s and early 1940s is not entirely clear in his published writings.

In my dissertation, I will show that Gödel did think more deeply about issues of constructivity, as well as philosophy of mathematics and foundational programmes, in general, and that even if he did not have a philosophy, he did explore philosophies. There is still much to gain from looking at Gödel’s personal notes and yet unpublished drafts in his *Nachlass* to complement the interpretation of the early articles and lectures. In general, Gödel’s papers are still largely an unknown territory due to the fact that he used an obsolete form of shorthand known as the *Gabelsberger-Schrift*. Unlike in Edmund Husserl’s case, where all of his Gabelsberger notes were eventually transcribed, the work on Gödel’s notes has only just begun.

The main question I want to answer is: how did Gödel’s views on intuitionism and constructive foundations develop in the 1930s and early 1940s, and how should we interpret his interest and motivation to study these topics? One might argue that there was no deeper motivation than that he was asked to talk about foundations of mathematics, and that he had a temporary interest to solve some problems related to intuitionistic logic; perhaps we should not put too much weight on these works. However, looking at Gödel’s notes from 1941, one can see that most of his attention was actually focused on intuitionistic logic and mathematics as well as the functional interpretation. In spring 1941, Gödel gave a full lecture course on intuitionistic logic and his functional interpretation of Heyting Arithmetic. In his *Arbeitshefte*, mathematical workbooks, we find hundreds of pages devoted to intuitionistic logic and mathematics, all written between late 1940 and late 1942. Even in the philosophical notebooks, there is a good handful of remarks related to intuitionism. It certainly does not appear that he came up with his results for the sake of a formal exercise.

Nevertheless, Gödel did not publish anything at all on this topic. The amount of notes that did not contribute to any published work is astonishing. Just as curious is the fact that very soon after his investigations into intuitionism, the mathematical notes run dry. There are barely any mathematical notes from 1943, and in 1944, he published his first purely philosophical paper in the Russell volume of the *Library of Living Philosophers*. After this, he wrote on little else than philosophy for a long time.

The story whose details I want to fill in is Gödel's philosophical development in the years leading to this sudden halt, as seen through his 1933–1941 lectures on intuitionism. I will support my research with a careful study of Gödel's unpublished notes. In doing so, I hope to shed some more light on Gödel's views of intuitionism as well as his early thoughts on questions of constructivity in the foundations of mathematics.

1.1 Beyond the crisis: Hilbert's and Brouwer's legacies

In this section, I will give a sketch of the general historical context in which we are to interpret Gödel's ideas. Gödel enrolled in the University of Vienna in 1924 and finished his doctoral thesis in 1929. By this time, there was already a heated discussion on the foundations of mathematics. Dissatisfied with Russell's logicist foundations, which assumed too much, and annoyed with Brouwer's intuitionistic reconstruction of mathematics, David Hilbert developed a foundational programme to ground classical mathematics. His idea, to put it roughly, was that mathematics could be divided into two parts, the real and the ideal, and that the consistency of the whole structure – serving as a proof of its adequacy – is to be shown in the secure and more elementary realm of real mathematics. Natural numbers, to give an example, are real objects; the general concept of a set or a proof would be an ideal object. The elementary methods that Hilbert claimed to be secure enough were called finitary, and his methodical viewpoint finitism. Only finitary objects were given meaning, and the ideal part of mathematics was treated as purely formal without any intrinsic meaning to them. This is why Hilbert's approach is often called formalism.

The idea of finitism in the context of the Hilbert Programme was developed in the early 1920s. Luitzen Egbertus Jan Brouwer, the founder of intuitionism, introduced his constructive viewpoint already in his dissertation in 1907. Intuitionism is, in a sense, a more radical view to grounding mathematics: instead of merely proving the consistency of mathematics, Brouwer wanted to rebuild it by more constructive standards. Nevertheless, Brouwer's intuitionism was not generally seen as a contestant to Hilbert's formalistic approach to the foundations before Hermann Weyl, one of Hilbert's favourite students, converted to intuitionism. Weyl's article "Über die neue Grundlagenkrise der Mathematik" (Weyl, 1921) in the *Mathematische Zeitschrift* did, in effect, create a crisis.

Hilbert answered soon with a series of talks, given in Hamburg later in 1921, titled "Neubegründung der Mathematik" (Hilbert, 1922), in which he first presented the idea of his Programme and the quest for the consistency proof.

Soon afterwards, the logic of intuitionism became a topic of extensive discussion, and between 1925 and 1930, intuitionistic logic was formalised. The earliest complete formalisation of a subsystem of intuitionistic logic was given by Andrei Kolmogorov, whose work would not be known to the broader community of logicians until later, and a full formalisation would be carried out by Arend Heyting, Brouwer's student, in 1930. Over time, Brouwer's quasi-phenomenological, formalism-free idea of intuitionistic mathematics was replaced by Heyting's more conventional approach, which made some aspects of Brouwer's ideas more widely accessible. This is the constructive viewpoint that underlies a great part of modern proof theory.

Towards the end of the 1920s, the *Grundlagenstreit* had heated up to the point where Brouwer came to blows with Hilbert, who then used his authority to throw Brouwer out of the editorial board of the *Mathematische Annalen*. As a consequence, Brouwer soon became relatively isolated, setting aside his work on intuitionism for a long time. There is ironic justice in the fact that Gödel's incompleteness theorems, in a sense, proved the Hilbertian dream of a universal consistency proof impossible and validated Brouwer's view of the impossibility of completion of formal mathematics. However, before talking about Gödel's achievements, let us go briefly over the developments in the late 1920s on both Hilbert's and Brouwer's side.

1.1.1 Brouwer's intuitionism

Intuitionism is often characterised by the rejection of the Principle of Excluded Middle (PEM) that is expressed in the axiom $A \lor \neg A$. Another classical form of reasoning that is not accepted by the intuitionists is the Double Negation Elimination (DNE), $\neg\neg A \supset A$. In quantified logic, the intuitionists reject the right-to-left direction of the classical theorem $\exists x A \equiv \neg \forall x \neg A$.

A standard example of a theorem of classical mathematics that is not intuitionistically valid is the Bolzano-Weierstrass theorem, which states that an infinite set of points S on a bounded interval $[a, b]$ has an accumulation point. The standard proof proceeds by dividing $[a, b]$ at the middle and picking as the next interval $[a, \frac{1}{2}(a+b)]$ if it contains infinitely many points in S, or else, the second half $[\frac{1}{2}(a+b), b]$. The process is then repeated with the new interval.

With continued iteration of the process, we obtain the result. The construction, however, is not effective in the sense that no instructions are given for deciding which half of the interval contains infinitely many points in S (Dummett, 2000, 7). Therefore, we have no general way to figure out the accumulation point. The problematic part of the proof is an instance of PEM which tells us that one or the other half must contain infinitely many points, and that the truth of this matter is enough for us to justify the inference, whether we can actually verify it or not.

Already in 1908, Brouwer interpreted PEM as solvability of every mathematical problem, a conviction – Hilbert is referred to here – for which "[there] is not a shred of a proof" (Brouwer, 1908, 109).[1] He believed that there are "fleeing" properties (*vluchtende eigenschappen, fliehende Eigenschaften*), decidable properties of natural numbers for which one cannot indicate whether there is a number satisfying the property or whether no number satisfies the property. He then used this idea to create "Brouwerian counterexamples" involving fleeing properties. Brouwer was not the first to see that the existence of such properties lead to counterexamples to the definability of real numbers. In his solution to Richard's Paradox,[2] Émile Borel – whose work Brouwer knew well – noted that one can give a real number by a finite definition which is yet ambiguous in the sense that its decimal expansion could depend on the solution of an unsolvable problem (Borel, 1908, 446).

The first Brouwerian counterexample appears in Brouwer (1924b):

> Let d_v be the vth digit after the decimal point in the decimal expansion of π and let $m = k_n$, if it occurs for the nth time in the progression of the decimal expansion of π at d_m that the segment $d_m d_{m+1} \dots d_{m+9}$ creates the sequence 0123456789. Moreover, let $c_v = \left(-\frac{1}{2}\right)^{k_1}$, if $v \geq k_1$, or else $c_v = \left(-\frac{1}{2}\right)^{v}$. Then the infinite sequence $c_1, c_2, c_3, \dots$ of this decimal expansion defines a real number r for which neither $r = 0$ nor $r > 0$ nor $r < 0$ holds.[3]

[1] Page numbering in Brouwer (1908, 1924b,a) refers to Brouwer (1975).

[2] Richard's Paradox, presented by the French mathematician Jules Richard, states roughly this: The set of real numbers that can be defined in finitely many words must be countable. Let us, then, list them based on the length of their definitions: $r_1, r_2, \dots r_n, \dots$. Now define a new real by diagonalising on the list, starting with 0, where the first decimal place is different from the first decimal place of r_1, the second decimal place different from that of r_2, and so on. This real number is clearly finitely definable, yet it cannot occur anywhere on the list.

[3] Sei d_v die v-te Ziffer hinter dem Komma der Dezimalbruchentwicklung von π und $m = k_n$, wenn es sich in der fortschreitenden Dezimalbruchentwicklung von π bei d_m zum n-ten Male ereignet, dass der Teil $d_m d_{m+1} \dots d_{m+9}$ eine Sequenz 0123456789 bildet. Sei weiter $c_v =$

Here "neither ... nor ..." has to be interpreted in terms of unprovability, not in terms of provability of the negation, for otherwise the statement would imply the contradiction $r \neq 0$ & $\neg r \neq 0$.

Brouwer did not think of PEM or DNE in terms of axioms or logical rules in the formal sense of the word. In the dissertation, titled *Over de grondslagen der wiskunde* (Brouwer, 1907), two theses are explicitly stated: mathematics is independent of logic and logic is dependent of mathematics. It should be mentioned that despite not being fond of formalism, Brouwer knew well the logic that he talked about: in *Grondslagen*, he discusses both Hilbert's axiomatics and Russell's logicism. His views towards logic changed from mostly negative to somewhat positive between 1907 and 1923, but he never let go of this idea about the place of logic with respect to mathematics. In order to understand intuitionistic logic, then, we must begin from intuitionistic mathematics.

Mathematics, in Brouwer's sense, is not to be identified with the language in which mathematics is written down. In its most basic sense, mathematics is a thought-activity, whose only framework is time, and whose first source is what Brouwer calls the "Primordial Intuition" (*het oer-phenomeen*). In essence, the Primordial Intuition consists of the recognition of apartness of two moments in time which are connected yet qualitatively different (Brouwer, 1907, 81). The act of separation, and the stripping out of any other qualities from the two moments of time save their structure, constructs the discrete when iterated. The same intuition gives rise to the continuum when the empty "two-ity" – Brouwer's own English translation of *tweeheid* first occurs in Brouwer (1949) – is again re-joined into one (Brouwer, 1913, 85–86). Neither structure is defined in terms of the other, but rather, the same framework of intuition gives rise to both in the recognition of unity and apartness.

Van Stigt (1990, 166) describes the Brouwerian hierarchical construction of mathematics. From the elements of construction and the elementary operations used to create structures such as natural numbers, integers, countable ordinals, etc., one gets a series of more complex constructions. The Creating Subject (*het scheppende subject*) – a term used by Brouwer in late 1940s to characterise the concept of an idealised human mind where the mathematics in fact happens – is then free to build more complex structures from these primitive constructions by fitting them together in different ways, creating properties,

$\left(-\frac{1}{2}\right)^{k_1}$, wenn $v \geq k_1$, sonst $c_v = \left(-\frac{1}{2}\right)^v$, dann definiert die unendliche Reihe $c_1, c_2, c_3, \ldots$ eine reelle Zahl r, für welche weder $r = 0$, noch $r > 0$, noch $r < 0$ gilt. (Brouwer, 1924b, 270).

relations, and more complex objects.

Whereas Brouwer did write on logic and its fallacies, such as PEM, his relationship to the subject remained quite complicated. In general, his attitude towards logic was negative in theory and somewhat disinterested in practice. Brouwer's original suspicion towards logic originated from his pessimistic view on natural language. Especially in his early works, language is treated as not only inaccurate but harmful and almost immoral.[4] For him, all language was essentially private, as reference to any external world is entirely absent from his picture of the language, and thus communication never quite succeeds. Moreover, Brouwer saw language primarily as an instrument of power and control, both at the level of individuals and at the level of society, this aspect of his view being more strongly present in his earlier writings (see, e.g., Brouwer, 1905).

The symbolic language of logic inherits the weaknesses of natural language. Franchella (1995, 306) notes that Brouwer's dislike for symbolism is also reflected in his early topological works. From the early view of language as a tool for manipulation and control, she says, follows that logic is, in a sense, the worst language one could adopt because it is cold and impartial and therefore unsuitable for pursuing power. The somewhat later notes offer, however, a contrary viewpoint to the situation: in the light of Brouwer's "signific interlude"[5] in the 1920s, one can also interpret mathematical language as superior to natural language in that it has a "low degree of egocity" and thus can be more objective (Van Dalen, 1999, 9).

Indeed, Brouwer's later views are not as negative, and he seems to have accepted the need for communicating one's ideas to others - the other use of mathematical language for him was to aid the memory (Van Stigt, 1990, 205) – in a form more accessible to the broader community of mathematicians. In 1923, two papers were published. In "Über die Bedeutung des Satzes vom ausgeschlossenen Dritten in der Mathematik" (Brouwer, 1924b),[6] Brouwer criticises certain classical principles, implying that there is a more or less

[4] The most controversial comments on language in Brouwer's dissertation were omitted, as his advisor, D.J. Korteweg, did not see them as appropriate in a dissertation on mathematics. They can be found in Appendix 3 of Van Stigt (1990).

[5] The *Significhe Kring* was founded by a group of Dutch intellectuals including Brouwer and the poet and psychiatrist Frederik van Eeden in 1916. The aim of the signific movement was to reform Western languages, impoverished of moral and spiritual value and used as an instrument of power and control, in order to create a better society (see Van Stigt, 1990, 65–71).

[6] Originally published in Dutch under the title "Over de rol van het principium tertii exclusi in de wiskunde, in het bijzonder in de functietheorie."

correct way of formulating the logic of intuitionistic mathematics. The 1923 article "Intuitionistische Zerlegung mathematischer Grundbegriffe" (Brouwer, 1924a)[7] gives an idea of how classical logic should be purified to give way to a new intuitionistically acceptable logic. Brouwer rejects PEM as well as its weaker form, $\neg A \vee \neg\neg A$. He also gives two correct principles involving negation (or in his terminology, absurdity), namely $A \supset \neg\neg A$ and $\neg\neg\neg A \equiv \neg A$. From the latter principle it follows that of three or more negations, one can cross out two at a time until one ends up with one or two negations.

Following his usual style, Brouwer does not attempt to formulate a system of axioms, and he uses almost no formalism: e.g., the last-mentioned principle $\neg\neg\neg A \equiv \neg A$ is formulated as "the absurdity-of-absurdity-of-absurdity is equivalent with absurdity" (Brouwer, 1924a, 277). He was still not interested in the task of formalising intuitionistic logic himself, and in any case, in Brouwer's sense of intuitionism, this task could never be quite completely achieved. As Van Atten (2017) writes, any formalisation of intuitionistic logic is, to a degree, incomplete, as "logic is as open-ended as the mathematics it depends on."

However, fragments can be formalised, and indeed they were. Brouwer's "Intuitionistic splitting" caught the eye of many mathematicians and logicians, and several papers were published on the topic in the 1920s. The nature of intuitionistic quantifiers, as well as PEM, was not at first entirely clear, and several attempts were made to formalise intuitionistic logic before Arend Heyting published his full formalisation of first-order logic and arithmetic. Incomplete or not, the work of Heyting (as well as of people such as Kolmogorov, Glivenko, Gödel, and Gentzen) is probably the reason intuitionism is still known today, for Brouwer's phenomenological picture of it, as beautiful as it is, was in the end too obscure to draw the interest of the wider community of logicians and mathematicians. Intuitionistic logic, however, survived and is still a fruitful field of research especially for proof theorists and computer scientists.

1.1.2 Towards intuitionistic logic

One of the first responses to Brouwer's 1923 article was Rolin Wavre's article in the *Revue de métaphysique et de morale*, which was, judging from the number

[7] Originally published in Dutch under the title "Intuitionistische splitsing van mathematische grondbegrippen."

of references to it in the following years, quite widely read. In "Y a-t-il une crise des mathématiques?" (Wavre, 1924), Wavre not only placed intuitionism in the context of the "foundational crisis" and against Hilbert's formalist programme but also discussed the intuitionistic notion of existence and the rejection of PEM as the main principles behind intuitionism which separate it from classical mathematics. Two years later he published a second article in the same journal titled "Logique formelle et logique empirique" (Wavre, 1926a), where he attempted to formalise intuitionistic ("empiristic") logic without the problematic principle of PEM. Whereas Wavre's article more or less lists principles already recognised in Brouwer (1924b), it was an important opening of the discussion on the nature of intuitionistic logic (Mancosu, 2010, 91).

First concrete steps towards formalising intuitionistic logic were taken already in 1925, when Andrei Kolmogorov published a paper "On the Principle of Excluded Middle." In his article, Kolmogorov presents an axiomatisation of what is known as minimal logic, i.e., intuitionistic logic without the principle of Ex Falso Quodlibet (EFQ) that states that from a contradiction, anything can be derived. Kolmogorov's formalisation does not, however, contain the connectives $\vee, \&$. Kolmogorov saw the need for a rule, as opposed to an axiom, for the introduction of a universal quantifier:

$$\cfrac{\vdots \\ A(y)}{\forall x A(x)}\ \forall I$$

although he does not give a similar rule for existence elimination

$$\cfrac{\exists x A(x) \qquad \begin{array}{c}[A(y)]_1 \\ \vdots \\ C\end{array}}{C}\ \exists E, 1$$

Nevertheless, it turns out that this rule is nevertheless derivable in Kolmogorov's system.[8]

Kolmogorov's central result, anticipating Gödel's and Gentzen's later works, is that classical propositional logic can be interpreted in the corresponding in-

[8] The conclusion of $\exists E$, given $\exists x A(x)$ and a derivation of C from $A(y)$, is derived using Kolmogorov's Axiom III for quantifiers (the inferences $\supset I, E$ are here represented as rules, but given as axioms in Kolmogorov's calculus):

tuitionistic logic, which is to say that intuitionistic propositional logic is consistent if its classical counterpart is. The original paper (Kolmogorov, 1925) was written in Russian, which, unfortunately, made it inaccessible to most. Oddly enough, it seems that Kolmogorov never mentioned the paper in his correspondence with Heyting nor to Bernays when he visited Göttingen in 1930–1931 (von Plato, 2017a, 174). One can find a mention of the article in Glivenko's letter of 13th October 1928 to Heyting (Troelstra, 1990b, 12) – Glivenko writes that Kolmogorov rejects the principle of EFQ[9] along with PEM – as well as an undated letter from 1933 or later[10] from Kolmogorov to Heyting (Troelstra, 1990b, 16); however, Hesseling (2003, 241) notes that there is no reference to Kolmogorov in the published articles on intuitionistic logic before 1933.

Here also Paul Bernays' name should be mentioned. In 1930, he wrote to Heyting that he started thinking about axiomatising intuitionistic logic after Brouwer's lectures in Göttingen in 1924, coming to the conclusion that the classical theorem $\neg\neg A \supset A$ should be left out (Troelstra, 1990b, 8–9). However, Bernays published nothing on the topic; in general, his published remarks on intuitionism in the 1920s are few and far between.

There was still plenty of confusion around the interpretation of PEM as well as the intuitionistic quantifiers. Marcel Barzin and Alfred Errera (1927) suggested that the rejection of PEM should be understood positively as the existence of a third truth value, "a tierce," where "p is tierce" is denoted by p'. However, they argue, it is not inconsistent to adopt the principle of the "Excluded Fourth," i.e., $p \vee p' \vee \neg p$. From this, they attempt to construct a refutation of the Brouwerian view. Using the above principle and their interpretation of intuitionistic logic, they derive the absurd statement $p \supset p'$, which makes the distinction between the three truth values meaningless (see Thiel, 1988, 69–70).

$$\dfrac{\exists x A(x) \qquad \dfrac{[\text{III.}]\ \forall x(A(x) \supset C) \supset \exists x A(x) \supset C \qquad \dfrac{\dfrac{\dfrac{\begin{array}{c}[A(y)]_1 \\ \vdots\end{array}}{C}}{A(y) \supset C}\supset I,1}{\forall x(A(x) \supset C)}\forall I}{\exists A(x) \supset C}\supset E}{C}\supset E$$

I am grateful to Jan von Plato for pointing this out to me.

[9] Here formulated as $\neg p \supset (p \supset q)$. Both Glivenko and Heyting accepted EFQ.

[10] Kolmogorov mentions here Gödel's negative translation between intuitionistic and classical arithmetic, published in 1933.

Barzin and Errera's suggestion was met with criticism from many sides. Perhaps the most interesting to us is Valerii Glivenko's short paper (Glivenko, 1928), where he gives an incomplete formalisation of intuitionistic propositional logic by which he obtains a direct syntactic proof against the Barzin-Errera conjecture. Glivenko's next paper (1929) presents a translation theorem for the full intuitionistic propositional calculus, showing that intuitionistic propositional logic proves $\neg\neg A$ if and only if classical logic proves A. Glivenko's result grew from the correspondence with Heyting after the publication of the 1928 article (see Troelstra, 1990b). Interestingly, Kolmogorov's name is nowhere to be found in Glivenko's paper, although it is clear from the Glivenko-Heyting correspondence that he knew of the 1925 paper.

The first complete formalisation of intuitionistic logic was developed by Heyting, who was one of Brouwer's students at the University of Amsterdam where Brouwer had become professor in 1912. Heyting wrote his dissertation on axiomatised intuitionistic projective geometry, a subject that was suggested by Brouwer. In 1928, the prize question for the annual contest of Dutch Mathematical Society concerned the logic of intuitionism: the task was to "1. formulate such a system and indicate the deviations of the formalism following from that system and Brouwer's theories, 2. to investigate if by means of a (formal) exchange of the principium tertii non exclusi and the principium contradictionis[11] an associated dual system can be derived."[12] The prize-winning essay was Heyting's. Brouwer was impressed with the work and suggested that it should be published in the prestigious *Mathematische Annalen*. However, very soon after he got into a quarrel with Hilbert, he refused to publish in the *Annalen* ever again and wrote to Heyting suggesting that he do the same (Van Dalen, 2005, 635). In the end, Heyting's work was published in three parts as "Die formalen Regeln der intuitionistischen Logik" and "Die formalen Regeln der intuitionistischen Mathematik I/II" (Heyting, 1930a,b,c) in the *Sitzungsberichte der preussischen Akademie von Wissenschaften*.[13]

[11] Law of Non-Contradiction, formalised as $\neg(A\&\neg A)$.

[12] Gerrit Mannoury to Brouwer, 26th January 1927, quoted and translated in Van Dalen (2005, 546–547).

[13] Heyting's original essay has apparently been lost; it is not known if it contained more than a draft of the first part of his three articles (Heyting, 1930a) on intuitionistic propositional logic. Jan von Plato drew my attention to the fact that in the published article, Heyting refers to Hilbert and Ackermann (1928) as one source for his formalisation of quantified intuitionistic logic – however, the book had not been published before Heyting submitted his prize essay in early 1928. *Principia Mathematica* could not have been Heyting's only source, for the system did not contain all the logical connectives and quantifiers.

Heyting's mode of presentation is axiomatic, with the standard rule of Modus Ponens as well as conjunction introduction as propositional rules, and universal introduction and existence elimination as quantifier rules. There are some weaknesses, such as inadequate substitution rules for predicate logic, lack of rules of multiplication for arithmetic, and no existence axioms for sets and functions in the formal system of analysis (Troelstra, 2011, 162). I will not list the axioms of Heyting's system here, as its details are not relevant for our purposes; for a thorough presentation, see Moschovakis (2009). Gödel's early works do refer to Heyting's 1930 axiomatisation, but the choice of particular axioms is mostly irrelevant.

Heyting's *Formalen Regeln* are, indeed, formal in nature. The first paper does begin with a slightly apologetic remark about the relationship between intuitionistic mathematics and its logic:

> Intuitionistic mathematics is a thought-activity, and any language, also the formalistic one, is only a tool for communication. In principle, it is impossible to set up a system of formulas that would be equivalent to intuitionistic mathematics, for the possibilities of thought do not allow for a reduction to a finite number of previously established rules.[14]

This is in line with Brouwer's own thought: although he supported Heyting in his efforts, Heyting himself recalls that Brouwer "always maintained that formalization is unproductive, a sterile exercise" (Van Stigt, 1990, 290). It is telling that the first thesis[15] in Heyting's dissertation (Heyting, 1925) reads, rather gloomily,

> The consequent formalism is irrefutable but worthless.[16]

Despite the inherent fallibility of the endeavour, Heyting maintains that the precision of symbolic language, as opposed to natural language, is important for the sake of communication; but he adds that mathematics will remain prior

[14] Die intuitionistische Mathematik ist eine Denktätigkeit, und jede Sprache, auch die formalistische, ist für sie nur Hilfsmittel zur Mitteilung. Es ist prinzipiell unmöglich, ein System von Formeln aufzustellen, das mit der intuitionistischen Mathematik gleichwertig wäre, denn die Möglichkeiten des Denkens lassen sich nicht auf eine endliche Zahl von im voraus aufstellbaren Regeln zurückführen. (Heyting, 1930a, 42)

[15] It is a convention in the Netherlands to include a number of "Stellingen" that the candidate must be prepared to defend in their doctoral defense.

[16] Het consequente formalisme is onweerlegbaar, maar waardeloos.

to logic, and the role of logic is only descriptive in nature. Nevertheless, Heyting does not elaborate much on the interpretation of intuitionistic connectives or the justification for his rules. Reflecting on his early works, he later regretted this mode of presentation for it "diverted the attention from the underlying ideas to the formal system itself" (Heyting, 1978, 15).

1.1.3 The proof interpretation

The natural way to interpret classical connectives is in terms of their truth conditions: this is reflected by the standard practice in logic textbooks that introduce the propositional connectives by way of truth tables. However, this approach does not work for intuitionistic logic, not even, as Glivenko had shown in 1928, if one were to use three-valued tables. (Indeed, Gödel would later show that no finitely valued semantics could interpret intuitionistic logic.) The central notion in intuitionistic mathematics is that of a construction that, in the case of logic, translates to the notion of proof.

The "proof interpretation"[17] or the BHK interpretation (for Brouwer-Heyting-Kolmogorov) can be formulated as follows.

(&) A proof of $A \& B$ consists of a proof of A and a proof of B.

($\vee$) A proof of $A \vee B$ consists of a proof of either A or B.

($\supset$) A proof of $A \supset B$ consists of a construction that transforms any proof of A into a proof of B.

In intuitionistic logic, $\neg A$ is defined as $A \supset \bot$. Given that we have

($\bot$) No proof of $\bot$ (absurdity) can be constructed.

it follows that $\neg A$ can be asserted only if any proof of A can be transformed into a proof of $\bot$, i.e., a contradiction can be derived from A.

The conditions of the quantifiers are

($\exists$) A proof of $\exists x A(x)$ consists of a proof of $A(b)$ for some b in the range of variable x.

[17] The term "proof explanation" is often used instead, underlining that the aim is to explain the meaning of the logical connectives rather than define them in formal terms. However, as "proof interpretation" or "BHK interpretation" is used more often, I will stick to this term.

($\forall$) A proof of $\forall x A(x)$ consists of a construction that, given any b in the range of the variable x, produces a proof of $A(b)$.

Gerhard Gentzen was the first to give a formal presentation of the BHK conditions in his rules of natural deduction, which can already be found in his 1932 notes under the title "Five different forms of natural calculi" (von Plato, 2017b, 113–115).

The above conditions express assertibility conditions for statements. I may assert $A \vee B$ if I can construct a proof of either A or B. What $A \vee B$ *means*, in this framework, is the set of *conditions* for a construction that picks either A or B and constructs it. To give another example, the meaning of $\exists x A(x)$ consists of the set of conditions for a construction that picks an object b of the domain and produces the construction $A(b)$. However, the BHK conditions of provability are not very informative in themselves, as "proof" and "construction" are left unspecified.

It is far from certain that Brouwer, Heyting, and Kolmogorov ever agreed on a common interpretation. Van Dalen (2004, 249) cites Heyting's remark to him that when he came up with the axioms and rules of intuitionistic logic in 1928, he had already thought of an interpretation. It is conceivable, Van Dalen continues, that Kolmogorov had also thought of one. Indeed, looking at Kolmogorov's paper, we get the impression that he was well read in both mathematics and philosophy, as in addition to frequent references to Brouwer, also Hilbert is referred to, and so are Leibniz and Aristotle. Moreover, the 1925 article does discuss the intuitionistic interpretation of Hilbert's axioms to motivate leaving out those that are not valid from the Brouwerian viewpoint. As mentioned, in the 1925 article, Kolmogorov does not include the rule EFQ, whereas Heyting includes it in his 1930 axiomatisation.

Kolmogorov's name, though, is not added after Brouwer and Heyting because of his early work of 1925, but because of a later article (Kolmogorov, 1932) which discusses the "problem interpretation" of intuitionistic logic. To give an example, where A and B are seen as problems, the problem $A \& B$ is to solve both problems A and B; to solve $A \supset B$ is to solve the problem B given a solution of A. In the 1932 article, Kolmogorov also accepts EFQ on grounds of a proof of unsolvability of the antecedent being enough for a proof of the whole implication. In his presentation of intuitionistic logic in his book of 1934, Heyting (1934, 14) mentions the problem interpretation, which he claims is essentially the same as his proof interpretation.

Why EFQ should be valid is not entirely intuitive, for one cannot ever present a solution to the contradictory problem, nor can one ever present a proof of a contradiction. Ingebrigt Johansson, who is known as the inventor of minimal logic (see Johansson, 1937) – which is intuitionistic logic without EFQ – also discussed the use of the principle, which he could not see as valid, with Heyting in 1935. Several letters were exchanged on the topic between the two, but Johansson maintained his belief that EFQ is not in line with the intuitionistic viewpoint. In a letter draft of September of 1935, Heyting writes that the interpretation of $a \supset b$ as reducing the solution of b to the solution of a, or showing the impossibility of a solution of a, is "effective" (*doelmatig*) (Van Atten, 2009, 132).

The interpretation of connectives in terms of constructibility conditions is already mentioned in Brouwer's dissertation of 1907. Of course, Brouwer avoids the use of formal language and gives no systematic presentation of the connectives, but he devotes plenty of thought to the study of the notions of existence, negation, and implication in their non-formal sense. Van Dalen (2008, 17) remarks that the idea of "proof = construction" can be found already in the dissertation. He describes Brouwer's thought as follows (ibid., 18):

> Mathematical statements describe constructs of the subject, made up of the basic entities of mathematics. In this description also the relations of the various building blocks are incorporated, that is to say the relations are themselves represented as constructions (or construction instructions) in a suitable mathematical way. A successful construction of the required structure then is the proof.

There is, however, no systematic discussion of the interpretation of (what corresponds to the intuitive concept of) logical connectives, and even in the later works, Brouwer does not attempt to give a conclusive explanation for the connectives. As mentioned, what exactly passes as a "construction" is never precisely defined, not out of sloppiness, but because the linguistic definition of such a notion is bound to remain open-ended.

Hypothetical judgement or implication as well as negation were key points of Brouwer and also of Gödel's later critique of intuitionism. The intuitionistic interpretation of an implication $A \supset B$ is explained already in Brouwer's dissertation. Given a structure A one obtains, by a series of transformations using tautologies,[18] a set of conditions for constructing the structure B (Brouwer, 1907, 126–127). How are we to understand this "given a structure A"? It does not seem plausible that one has to *construct* a proof of A, for that would amount to giving an intuitionistic proof of $A \wedge B$. Noting this, Van Atten (2009, 128) suggests the following interpretation: given the conditions for constructing A, one has to obtain, by a chain of intuitionistically valid transformations, the conditions for the construction of B. Therefore one can take A as an assumption and attempt to prove B from the consequences of the assumption. In formal terms, the two interpretations are illustrated by the difference between admissibility and derivability, i.e., between the cases where A is assumed to derive B, and where A is assumed to *have been proven*, to derive the proof of B:

$$\frac{\vdash A}{\vdash B} \qquad\qquad \frac{\begin{matrix}[A]\\ \vdots \\ B\end{matrix}}{A \supset B}$$

Clearly, the first version cannot be taken to represent the conditions for introducing an implication: from $\vdash A \vee B$, $\vdash \neg B$ one can derive A, but $((A \vee B)\&\neg B) \supset A$ is not derivable.

The proof of a negation is a special case of that of an implication. To prove a negation $\neg A$ means that in an attempt to derive something from the assumption A, one hits an inconsistency or incompatibility with some other acceptable construction. Brouwer often says simply that the construction does not go further (*gaat niet verder*), that one gets stuck or that the construction cannot fit in with what has been previously constructed (Van Stigt, 1990, 242–243). The word "impossibility," as in impossibility of fitting a construction another construction, is used often, as well. This suggests a reading of "does not go further" as "*cannot* go further," i.e., that one can actually show that the construction does not fit in or is in contradiction with something known to be the case. However, as Van Stigt (ibid., 245) notes, Brouwer later uses the phrase "impossible for now and always" (*onmogelijk voor nu en altijd*) instead of absurdity, where the latter could also refer to not being able to actually do something.

Although Heyting does not discuss the notion of intuitionistic validity in the "Formalen Regeln," where he expresses the validity of a formula by the

[18] "Tautology" here should not be understood in the sense of a formula which is always *true* but rather as a form of inference that is seen as intuitionistically acceptable.

word "correctness," he does mention the idea of provability both in his "Sur la logique intuitionniste" (Heyting, 1930d) that was his response to Barzin and Errera, and in the Königsberg meeting of September 1930 (Heyting, 1931). In these papers, he explains that the assertion of a proposition p, from the Brouwerian point of view, means that it is provable (Heyting, 1930d, 959). The meaning of p is the expectation (*Erwartung*) that p can be constructed; Heyting (1931, 113) writes that the phenomenological concept of intention could be more suitable here. Then a proof of p fulfils this intention. A negation $\neg p$, then, means the expectation that from p one can construct a contradiction (ibid.); a disjunction $p \vee q$ means an intention that is fulfilled when at least one of the intentions of p and q are fulfilled (ibid., 114). The explanation for implication would have to wait until Heyting (1934), where $a \supset b$ is defined, as one would expect, as "the intention of a construction that transforms any proof of a into a proof of b" (Heyting, 1934, 14). Heyting describes the meaning of the quantifiers through Kolmogorov's problem interpretation (Kolmogorov, 1932): $\forall x A(x)$ is the problem of giving a general procedure for solving $A(x)$ for any x, and $\exists x A(x)$ a problem of giving a particular x together with a solution for $A(x)$.

Heyting discusses the relation of classical and intuitionistic logic in 1930 and 1931. He distinguishes between two meanings of a proposition p, namely, "[it is the case that] p", and "p is provable," the latter denoted as $+p$. By these interpretations, $\neg p$ and $\neg + p$ mean different things: the first says, "it is not the case that p," whereas the second, "it is not the case that p is provable," that is, "an (assumed) construction of p leads to a contradiction." The latter is the Brouwerian explanation of negation. Indeed, when an assertion of a proposition is understood as its construction, the difference between asserting p and asserting $+p$ disappears (Heyting, 1931, 114–115). The idea of representing intuitionistic logic as the logic of provability is an anticipation of Gödel's 1932 translation, although Heyting did not give any formalisation of such a logic.

It should be mentioned that Gödel knew at least Heyting's Königsberg lecture very well; he was there at the conference, and he also reviewed the published paper for the *Zentralblatt für Mathematik*. He also knew of the 1934 book because it had started as a joint project between him and Heyting. Heyting had been asked by Otto Neugebauer to write a book on foundations of mathematics, and he agreed on the condition that someone else write the part on logicism. Gödel was then suggested as a co-author (Hesseling, 2003, 279). Gödel agreed to the task, and there is correspondence between the two on the book, but despite Heyting finishing his part and sending it for review to Gödel, the latter was unable to finish his part in time, and the project was never finished. Heyting's typewritten drafts, which are virtually identical to the 1934 publication, can be found in Gödel's *Nachlass* (040019.5).

1.1.4 Meanwhile, in Göttingen

As Gödel's lectures on intuitionism not only talk about intuitionistic logic, but also about constructivity more generally, and because he was both a critic and a contributor to the "Extended Hilbert Programme," something should be said about the original Programme and its methodical standpoint.

In brief, the aim of the Programme, first introduced by Hilbert in 1921–1922, is to justify classical mathematics by means of providing a proof of consistency by only the simplest and most reliable of methods. Hilbert divides mathematics into two parts, the real and the ideal. The real part of mathematics, often referred to as *finite* mathematics, is to be taken as meaningful due to its being "intuitively evident" from the human perspective. The ideal part, Kant's *Ding an sich*, transcends this framework of human intuition. Unlike Kant, who was an obvious influence for Hilbert's conception of intuition, and also unlike Brouwer, Hilbert also accepted the ideal part of mathematics, as long as it could be proven consistent in its finite fragment.

Besides Kant, another influence was Leopold Kronecker, whose mathematical works inspired Hilbert in his early studies on algebra and geometry. In an obituary of Hilbert, Hermann Weyl describes Hilbert's relationship to Kronecker colourfully (Weyl, 1944, 613):

> When one inquires into the dominant influences acting upon Hilbert in his formative years one is puzzled by the peculiarly ambivalent character of his relationship to Kronecker: dependent on him, he rebels against him. Kronecker's work is undoubtedly of paramount importance for Hilbert in his algebraic period. But the old gentleman in Berlin, so it seemed to Hilbert, used his power and authority to stretch mathematics upon the Procrustean bed of arbitrary philosophical principles and to suppress such developments as did not conform: Kronecker insisted that existence theorems should be proved by explicit construction, in terms of integers, while Hilbert was an early champion of Georg Cantor's general

set-theoretic ideas.

This ambivalence explains Hilbert's strong reaction to Brouwer's intuitionism, which, although much more tolerant than Kronecker's finitism, was meant to replace classical mathematics: as Weyl puts it, "Hilbert's slashing blows are aimed at Kronecker's ghost whom he sees rising from his grave" (ibid.). However, the ambivalence is not seen only in Hilbert's earlier algebraic works and his subsequent turn to Cantor. The whole Hilbert Programme is, in a sense, bipolar. On the one hand, classical mathematics or "Cantor's Paradise," as Hilbert once called it, must be preserved (Hilbert, 1926, 170). On the other hand, however, classical mathematics cannot be taken at face value, but it must be justified by some *more* reliable means, means which Hilbert, still in 1931, characterised as essentially the same as Kronecker's (Hilbert, 2013, 976). In a sense, Hilbert attempts to contain two conflicting ideals in one theory: the Platonistic approach of Cantor, and the ultra-constructive viewpoint of Kronecker.

Hilbert's Programme was first presented in his lectures on proof theory given in 1921–1922 (see Hilbert, 2013, chapter 3). Here Hilbert first talks about finitary mathematics as opposed to transfinite mathematics, i.e., the ideal part of mathematics that transcends the finite part. Soon after Hilbert's Hamburg lecture of 1921 (Hilbert, 1922), Paul Bernays, Hilbert's closest collaborator whose role in the philosophical part of the Hilbert Programme was significant, gave a talk in Jena that was essentially a summary and a clarification of Hilbert's most important points (Bernays, 1922).

Bernays' talk describes the aim of the Programme as well as its philosophical foundation with beautiful clarity. The aim is to prove the consistency of mathematical systems by grounding its transfinite part "in such a way that only *primitive intuitive cognitions come into play*" (Bernays, 1922, 216).[19] Frege's attempt to ground mathematics in logic was found to result in contradictions, and Brouwer's constructivisation sacrificed the most useful tools of the working mathematician. Hilbert's approach, on the other hand, attempts to synthesise the two views in a way that both retains the reliability of method and allows for the full use of classical mathematics. From Frege's logicism arises the *formalistic* viewpoint of Hilbert: for the purposes of the consistency proof, ideal mathematics is to be seen only as pure formalism with no intrinsic meaning to the symbols that express the axioms and rules of the formalised theory

[19] Page numbering refers to the English translation in Mancosu (1998).

(ibid., 219–220). Only the finite part is seen as contentful, and it derives its content essentially from its intuitive evidence. The outcome is, says Bernays, "that the problems and difficulties that present themselves in the grounding of mathematics are transferred from the epistemologico-philosophical domain into the domain of what is properly mathematical" (ibid., 221–222).

How exactly one defines "evident" or "intuitive" is not a question that is directly answered by Hilbert and Bernays in the 1920s. An illustrative description is the frequently quoted passage from "Über das Unendliche" (Hilbert, 1926, 171):

> [As] a condition for the use of logical inferences and the performance of logical operations, something must already be given to our faculty of representation, certain extralogical concrete objects that are intuitively present as immediate experience prior to all thought. If logical inference is to be reliable, it must be possible to survey these objects completely in all their parts, and the fact that they occur, that they differ from one another, and that they follow each other, or are concatenated, is immediately given intuitively, together with the objects, as something that can neither be reduced to anything else nor requires reduction.[20]

Therefore we obtain the natural numbers, perceived as a potential infinity, i.e., a process that can be iterated arbitrarily many times as opposed to a completed totality, as well as the standard arithmetic operations of addition and multiplication. As for the exact scope of finite mathematics, Hilbert and Bernays never gave an exact definition: the examples of finitary operations and proofs are scattered around in different works. Whereas finitism is now equated with Primitive Recursive Arithmetic (PRA), this was never explicitly stated by Hilbert and Bernays. As Richard Zach (2003) notes, there is a division between the delineation of the philosophical viewpoint and the actual practices of the Programme, e.g., the methods used in Wilhelm Ackermann's consistency proof of 1924 and Hilbert's 1925 attempt of a proof of the Continuum Hypothesis, which clearly transcend PRA.

The actual aim of the Hilbert Programme in the 1920s was to obtain consistency proofs for the systems of arithmetic and hopefully analysis by the use of the safest and most reliable mathematical methods. The collection of finitary

[20] English translation is from Van Heijenoort (1967, 376).

methods is not necessarily equal to any formal system, although it seems to correspond roughly to PRA. The justification for methods of finitism was that they were intuitively evident. The finitistic objects are the natural numbers, considered as a potential infinity, and elementary operations on the natural numbers: in Hilbert and Bernays' writings, addition, multiplication, exponentiation, and quantifier-free induction, to give a few examples, are interpreted finitarily. There is, however, no complete presentation of *the* formal theory that corresponds to finitism, and some proofs that were considered finitary could not be carried through in PRA. Therefore the question about the precise scope of Hilbert's finitary mathematics, at least as it was practised, is left open.

1.1.5 Intuitionism and finitism

As mentioned above, Hilbert's attitude towards Brouwer, especially after one of his brightest students, Hermann Weyl, had loudly converted to Brouwer's intuitionism, was remarkably aggressive. The Hilbert Programme was put forward partly as a response to Weyl's "neue Grundlagenkrise" (Weyl, 1921), and was seen as an alternative to intuitionism. Brouwer's strong sense of justice made him react intensely to Hilbert's rage, eventually causing the latter to dismiss Brouwer from the editorial board of *Mathematische Annalen* in 1928. Brouwer took this as an immense insult and became increasingly bitter and isolated. As a result, he published very little on intuitionism in the 1930s and early 1940s.

Even considering Brouwer's austere personality, Hilbert's strong reaction seems somewhat out of place, for although Brouwer did not see all of classical mathematics as meaningful, his methodical approach was not that different from Hilbert's finitism, and it certainly was not more limited in scope. However, Hilbert and Bernays did not give a comprehensive answer to the question of which methods are to be considered finitary, and as late as in 1930, the difference between finitism and intuitionism, measured in terms of their strength, was still not precisely understood. For example, Bernays (1930, 351) writes that there is no meaningful difference as to the methodology of the Hilbert Programme and that of the intuitionists, although their approaches to the foundations of mathematics are entirely different. Tait (2006a, 83–84) notes that the statements about the relationship between finitism and intuitionism would perhaps be better described as ambiguity rather than misidentification: there were, he says, "two conceptions of arithmetic reasoning at the time, one more restrictive than the other, and both called intuitionistic by von Neumann and Herbrand" – and indeed, even Gödel seems to have been slightly ambiguous in his usage of the terms at the time of the discovery of the incompleteness theorems (see **section 2.1.1**).

It does not help that there were several conceptions of "intuitionism" floating around in the 1920s. There was plenty of discussion around how to interpret Brouwer's intuitionism in formal terms, the logic of intuitionism was not even fully formalised before 1928, and people like Hermann Weyl put forth a conception of intuitionism that was not entirely in line with what Brouwer himself wrote. To give an example, Weyl had a stricter interpretation of the quantifiers than Brouwer. The implication $\neg\forall x P(x) \supset \exists x \neg P(x)$ is a consequence of PEM; this direction of the quantifier dualities is intuitionistically unacceptable. Weyl counted the following as an instance of PEM: either all numbers have a property P or else there is a number which has the property $\neg P$, i.e., $\forall x P(x) \vee \exists x \neg P(x)$ (Weyl, 1926, 42). It follows from this that $\neg\forall x P(x)$ implies $\exists x \neg P(x)$. Therefore, Weyl concludes, one cannot negate universal statements without having the corresponding instance of a negative property. Weyl stated that quantified statements are not proper judgements but rather "judgement-instructions" (*Urteilsanweisungen*) or "judgement-abstracts" (*Urteilsabstrakte*). Indeed, Brouwer did notice that Weyl's conception of quantifiers was stricter than his (Van Dalen, 1995, 160). Hilbert, who did not read Brouwer,[21] got the impression that this was also Brouwer's view, going as far as to say that Brouwer's view is essentially the same as Kronecker's ultra-finitistic standpoint (Hilbert, 1922, 159). This may have further complicated the issue of differentiating between finitism and intuitionism.

Another issue is with the definition of finitist intuition and Brouwer's conception of intuition that both can be said to have roots in the Kantian conception, although neither is perfectly in line with Kant's arithmetic intuition, which was very restrictive (see Tait, 2010). Some phenomena that Brouwer does base on intuition (understood as a construction by the Creative Subject from the elements given in the intuition), such as choice sequences, do go beyond what Hilbert and Bernays saw acceptable. However, this does not necessarily mean that they disagreed on what intuition *is* or how we perceive its objects. Detlefsen (1998) suggests that the most significant difference lies

[21] Van Dalen (2005, 637) recalls that Paul Bernays told him that Hilbert never read Brouwer; this is obvious from his criticism of intuitionism, which could apply to Weyl's conception but not to Brouwer's.

in their respective concepts of the act of intuiting. For Brouwer, the Creative Subject has subjective certainty over their creation (Detlefsen, 1998, 319); however, Hilbert's idealised subject is inherently intersubjective, having to be able to communicate their thoughts to others (ibid., 322). To give an example, the Hilbertian view requires that in order for a proof to be valid, one should be able to convince another mathematician of its validity, whereas for Brouwer, it is enough to *see* that the proof is valid. The difference in the mode of intuiting explains why intuitionistic methods are less restrictive than finitary ones.

When looking at Gödel's contributions to Hilbert's Programme, one must also take into account that the finitism of the 1920s is not necessarily the same as finitism after Gödel's own incompleteness results. As the theorems showed the impossibility of proving the consistency of even arithmetic in PRA, the post-incompleteness Programme had to either search for alternatives or admit defeat. Obviously, Hilbert and Bernays were not ready to give up, and therefore called for an extension of the finitary framework in a way that would allow proving the consistency of arithmetic and hopefully even analysis. They did not, however, give an explicit characterisation of extended finitism, and there was no general idea of how to construct a consistency proof that would be finitary in this more relaxed sense.

In 1934, Bernays gave a lecture titled "Quelques points essentielles de la métamathématique" (Bernays, 1935a), where he suggested that the principle of transfinite induction up to ε_0, which is not provable in Peano Arithmetic but that is *intuitionistically* valid, could be used in a consistency proof. In 1936, Gerhard Gentzen proved the consistency of arithmetic using transfinite induction to ε_0 over primitive recursive predicates; both Gentzen and Bernays argued explicitly for the finitary nature of Gentzen's proof (Hämeen-Anttila, 2019). This brought finitistic methods even closer to intuitionistic ones.

Not all were convinced, however: arguably, Gentzen's transfinite induction went well beyond what had originally been called finitistic. Gödel was one of the critical voices; in the Zilsel lecture (Gödel, 1938), he argued that Gentzen's construction of the transfinite ordinals, although highly intuitive, contains impredicative elements. He did, however, offer a positive solution (Gödel, 1941) for a more constructive consistency proof that used what is called a functional interpretation of Heyting Arithmetic. Now I will turn to Gödel's own contributions to intuitionistic logic from the early formal works of 1932–1933 to the more philosophical lectures of 1933, 1938, and 1941.

1.2 Gödel's contributions to constructive foundations

Gödel's best-known results on intuitionistic logic were all discovered between 1931 and 1941. The four key results are

1. Intuitionistic logic cannot be represented in a finite-valued semantics: between intuitionistic and classical logic, there is an infinite number of intermediate logics, each slightly stronger than the previous one.

2. Negative translation between classical and intuitionistic arithmetic, which proves the relative consistency of Peano Arithmetic with respect to Heyting Arithmetic.

3. Interpretation of intuitionistic propositional logic in a modal logic that corresponds to the Lewis system **S4**.

4. Interpretation of Heyting Arithmetic through finite-type functionals.

The first three results were all published between 1932 and 1933. The fourth result was not published until 1958, although Gödel first came up with the idea in the late 1930s.

As for Gödel's philosophical contributions in the 1930s and early 1940s, he talked on the topic of constructivity and intuitionism on three occasions in 1933, 1938, and 1941, and held one lecture course on the same topic in the Spring Semester of 1941. The functional interpretation was first mentioned in 1938 when Gödel presented it as an extension of Hilbert's finitism more constructive than Heyting's proof interpretation. In 1941, the functional system was introduced as an interpretation of Heyting Arithmetic and laid out in detail.

In this section, I will give a brief introduction to Gödel's known works on intuitionistic logic and constructive foundations. In the following chapters, I will give a more detailed account of the historical details as well as the contents of the three philosophical lectures and the Princeton course.

1.2.1 The logical works

As mentioned, there are three papers, all published in a very short timespan. The first work, "Zum intuitionistischen Aussagenkalkül" (Gödel, 1932), is barely two pages long and starts with a note "In Beantwortung einer von Hahn aufgeworfenen Frage," in reply to a question posed by Professor Hans Hahn. Here Gödel proves two results:

1. Heyting's propositional calculus **H** has no realisation with finitely many truth values.
2. Between **H** and the classical calculus **A**, there are infinitely many intermediate logics.

The first theorem is a generalisation of Glivenko's 1928 result that the principle of "Excluded Fourth" does not hold for intuitionistic logic. Already in 1927, Oskar Becker had suggested that there is no excluded nth (for n greater than two), either (Becker, 1927, 777), although he gave no proof. The second theorem, of which the first is an immediate consequence, is also interesting in its own right, and has led to the study of intermediate logics which are stronger than the standard intuitionistic calculus but weaker than classical logic. Gödel's result shows that there is, in fact, an infinite number of intermediate systems.

Gödel's second publication on intuitionistic logic (Gödel, 1933c) was a paper titled "Zur intuitionistischen Arithmetik und Zahlentheorie," presented in Karl Menger's Colloquium on 28th June 1932. The result was what is now called a negative translation between Heyting Arithmetic and its classical counterpart, Peano Arithmetic. The main idea of such a translation is to give a mapping $'$ from a formula A of one system to another formula A' in the other system – here Peano and Heyting arithmetic, respectively – so that theoremhood is preserved under $'$. Gödel refers to Glivenko (1929), who had given a translation for propositional logic, where the translation is simply $A' := \neg\neg A$. Also Kolmogorov (1925) anticipated Gödel's result, suggesting a translation which inserts a double negation before every subformula. Gerhard Gentzen came upon the same idea independently in 1932; by February of the next year, he had submitted his paper to the *Mathematische Annalen* (von Plato, 2017b, 19). Gödel learnt of Gentzen's translation from Heyting; in a letter of 16th May 1933, he writes to Heyting that his work should be known in Göttingen "since at least June 1932," as Oswald Veblen, who was present at Menger's Colloquium, arrived there soon after (ibid., 20). Gentzen withdrew his paper, and his version of the translation was not published until almost four decades later in his *Collected Works* (Szabo, 1969).

Gödel's 1932 translation is given as follows:

$(A)' = A$ where A is atomic

$(\neg A)' = \neg A'$

$(A \& B)' = A' \& B'$

$(A \vee B)' = \neg(\neg A' \& \neg B')$

$(A \supset B)' = \neg(A' \& \neg B')$

$(\forall x A)' = \forall x A'$

$(\exists x A)' = \neg \forall x \neg A'$

It is then shown that whenever A is provable in Peano Arithmetic, then A' is a theorem of Heyting arithmetic, and, moreover, that correctness is preserved under application of rules of inference. A direct consequence of Gödel's theorem is that Peano Arithmetic is consistent if Heyting Arithmetic is, for if $\vdash_{PA} \bot$, then by the translation theorem, also $\vdash_{H} \bot$. Conversely, if Heyting Arithmetic does not prove a contradiction, then Peano Arithmetic will not prove one either.

There is also philosophical value to Gödel's negative translation and the resulting relative consistency proof. The incompleteness theorems had already confirmed that the methods used so far by the finitists, all more or less falling within the scope of PRA, could not prove the consistency of Peano Arithmetic. What the negative translation showed was that intuitionistic methods did surpass the finitary ones, at least those used in the 1920s. In this rather formal sense, intuitionism is stronger than (original) finitism.

The third small article, "Eine Interpretation des intuitionistischen Aussagenkalküls" (Gödel, 1933a), is a translation of intuitionistic propositional logic into modal logic. As mentioned, Heyting (1931) denoted "p is provable" by $+p$, although he only introduced the idea informally. Gödel's provability logic essentially corresponds to the modal system **S4**. The intuitionistic connectives are expressed through the modal operation **B** for "beweisbar."[22] The translation goes as follows:[23]

[22] In a letter of 12th January 1931, to Gödel, von Neumann uses the notation B for Gödel's *Bew* predicate and notes that B$a \supset$ BBa – the axiom characterising **S4** – must hold (Gödel, 2003b, 342). In his later lecture notes on incompleteness, Gödel uses B instead of *Bew*.

[23] Gödel's original paper translates $p \wedge q$ simply as $p \wedge q$, but he notes that $\mathbf{B}p \wedge \mathbf{B}q$ works just as well.

$\neg p = \neg \mathbf{B}p$

$p \wedge q = \mathbf{B}p \wedge \mathbf{B}q$

$p \vee q = \mathbf{B}p \vee \mathbf{B}q$

$p \supset q = \mathbf{B}p \supset \mathbf{B}q$

Gödel notes that the sense of "provable" of the **B**-operator cannot always be interpreted as "provable in a formal system" because this would contradict the second incompleteness theorem. As $\mathbf{B}(\mathbf{B}p \supset p)$ is a theorem, by substituting $\perp$ for p we get $\mathbf{B}\neg\mathbf{B}\perp$, which claims that it can be proven that there is no proof of contradiction.

1.2.2 The lectures

The three lectures of 1933, 1938, and 1941 are, in contrast to the early works on intuitionistic logic, partly philosophical in nature, and they discuss the idea of constructivity of a logic as well as intuitionism as a foundation for mathematics. These lectures provide the background against which I will examine the development of Gödel's views on intuitionism. Here I will give only a short overview of the three lectures.

On 30th December 1933, Gödel gave a lecture in Cambridge, MA, titled "The present situation in the foundations of mathematics" (Gödel, 1933b). He first considers the status of foundations after the incompleteness theorem and then discusses the question of justification for certain sets of axioms and rules of inference. The meaning of the classical axioms that allow for non-constructive proofs of existential statements and impredicativity presupposes "a kind of Platonism, which cannot satisfy any critical mind [...]" (Gödel, 1933b, 50). One needs a more secure foundation for a proof of consistency of mathematics. Gödel notes that "what remains of mathematics if we discard these methods [...] is the so-called intuitionistic mathematics" (ibid.). "Intuitionistic or constructive" mathematics forms a hierarchy of more or less strictly constructive systems. Here the strictest form of a constructive system satisfies the three properties of

1. the restriction of the universal quantifier to applications over totalities generated by a finite process,

2. the restriction of the use of negation over universal statements only to the case where it abbreviates a counterexample, and

3. decidability or calculability of all primitive relations and functions.

This is, Gödel claims, the system that Hilbert wanted to use, but whereas it is very secure, it is also too weak to prove even the consistency of standard arithmetic.

Brouwer's intuitionistic mathematics, however, goes further than this. It fails, says Gödel, to satisfy the first two requirements, as the proof interpretation of the connectives has them ranging over the totality of *all possible proofs*, which is not a finitely generated set of objects. Furthermore, the proof of $\neg\forall x A$ does not necessarily involve a counterexample $\neg A[x/a]$. Therefore intuitionism does not provide a good constructive foundation for mathematics.

Gödel returns to the topic two more times during the 1930s and 1940s. On 29th January 1938, he gave a lecture at a meeting at Edgar Zilsel's home. Zilsel did not have a university position; he was a *Gymnasium* teacher who had a particular interest in history and sociology of science (Sieg and Parsons, 1995, 62). The seminar had probably been meeting for several years before Gödel gave his talk. Zilsel had asked him to give a lecture on the "consistency question," and Gödel had initially suggested to present the Cambridge lecture in German. However, he changed his mind because the lecture was too general (ibid., 63). Instead, he gave a talk that went into more detail about the question of constructive foundations and possibilities of constructive consistency proofs for mathematics.

The Zilsel talk presents essentially the same conditions for constructivity as those listed in the Cambridge talk, here called simply the framework, *die Rahmendefinition*, for constructivity. Whereas Hilbert's system is one of the simplest theories that satisfy these conditions, Gödel now believes that it is possible to extend Hilbert's finite system in a way that still fulfils all of the conditions (Gödel, 1938, 95). Heyting's provability interpretation is not such an extension, as it fails all of the criteria;[24] likewise, Gentzen's 1936 consistency proof, although a better alternative, is inadequate. Gödel's positive thesis is that a system based on functionals of finite types (and perhaps also transfinite

[24] The reason why HA fails the condition of decidability and calculability is that Gödel introduces intuitionistic logic in terms of the modal interpretation, and the provability predicate cannot necessarily be considered as decidable.

ones) could be an extension that would satisfy all of the criteria, but he does not give any details as to how such a system is to be constructed.

Gödel did fill in the details in a lecture presented in Yale on 15th April 1941. The title of the lecture, "In what sense is intuitionistic logic constructive?", suggests a slightly different viewpoint than that of the 1930s lectures. Namely, the question implies that, first of all, intuitionistic logic *is* constructive, and that Gödel will show the sense in which it is so. This sense is given by the functional interpretation, now given in full.

1.2.3 The functional interpretation

Gödel's system Σ interprets HA in terms of functionals, and it more or less corresponds to the system called **T** in Gödel's *Dialectica* article 17 years later (Gödel, 1958). Σ extends Primitive Recursive Arithmetic with the introduction of primitive recursive functionals of higher types, i.e., the extension of definitional schemes to functions over lower-type functions. Each term in our language is assigned a type so that:[25]

1. Computable functions over natural numbers have type 0.
2. A function g whose arguments are variables $f_1, f_2, \ldots f_k$ of types $\tau_1, \tau_2, \ldots \tau_k$ and whose values are of type τ_{k+1} is of type $(\tau_1 \times \tau_2 \times \ldots \times \tau_k) \mapsto \tau_{k+1}$.

We may define the *level* of a term so that a term of type 0 has the level 0, and a term of type $\tau_1 \times \tau_2 \times \ldots \times \tau_k \mapsto \tau_{k+1}$ has the level $\max(\max(\text{level}(\tau_1), \ldots \text{level}(\tau_k)) + 1, \text{level}(\tau_{k+1}))$, i.e., the level of a term is either the level of its greatest-level argument $+1$, or the level of its value, whichever is larger. We may also define the level of a formula as the level of its highest-level term.

The functional translation of Heyting Arithmetic proceeds as follows. To each formula φ of HA is associated a formula φ^D in our functional system which has the form $\exists x_1 \ldots x_n \forall y_1 \ldots y_n A(x_1, \ldots x_n, y_1, \ldots y_n)$, A quantifier-free. The D-translation is given recursively as follows:

1. For atomic A, $A^D = A$.

Let $\varphi = \exists x \forall y A(x, y)$ and $\psi = \exists u \forall v B(u, v)$, where A and B are quantifier-free formulas of an arbitrary level. Then

[25] Here I will use a notation similar to Gödel (1958), which is slightly different from the 1941 presentation.

2. $(\neg\varphi)^D = \exists f \forall x \neg A(x, f(x))$

3. $(\varphi \wedge \psi)^D = \exists x, u \forall y, v(A(x,y) \wedge B(u,v))$

4. $(\varphi \vee \psi)^D = \exists x, u, z \forall y, v((A(x,y) \wedge z = 0) \vee (B(u,v) \wedge z = 1))$

5. $(\varphi \supset \psi)^D = \exists f, g \forall x, v(A(x, g(x,v)) \supset B(f(x), v))$

6. $(\forall z \varphi(z))^D = \exists f \forall z, y A(f(z), y, z)$

7. $(\exists z \varphi(z))^D = \exists x, z \forall y A(x, y, z)$

Because the rules for the universal quantifier as well as implication (and negation, which is simply the *D*-translation of $A \supset \bot$) increase the level of the expression, one will need all finite types to deal with expressions of arbitrary complexity.

In 1938, Gödel suggested that the functional system could be used to give a consistency proof for arithmetic. In 1941, he also presents it as an alternative to the proof interpretation as well as a proof of constructivity of Heyting Arithmetic. Nowadays, the functional interpretation is often presented as a form of a proof interpretation, although it is another question whether Gödel thought of it as such.[26]

Another question, which was of high importance to Gödel, is whether the functional interpretation satisfies his own criteria of constructivity. In 1941, he states that a proof of computability for higher-type functionals is needed to show that the interpretation satisfies the conditions. Proofs for computability have been given for Gödel's system (see, e.g., Howard, 1970), but Gödel never presented one. In the Yale lecture, he mentioned that the proof is rather complicated and that he will not discuss it (Gödel, 1941, 195). In the 1958 article, identity of higher-type functionals is taken as a primitive and computability is simply assumed as a given that cannot be made more evident by a proof. Gödel remarks that the (informal) notion of computability by a Turing Machine cannot be made any clearer by a formal interpretation, as the formal interpretation would not make any sense if the informal notion were not already intelligible (Gödel, 1958, 244).[27]

[26] There are now functional interpretations for various systems, from subsystems of arithmetic to full analysis. For a concise introduction to the topic, see Avigad and Feferman (1998); Kohlenbach (2008) is a thorough presentation of realisability and functional interpretations as well as related results. Here, we will only discuss the formal details of functional interpretations where it is necessary for understanding Gödel's thought.

[27] Page numbering refers to the English translation in Gödel (1990).

1.3 Methods and sources

In addition to the published works referred to above, my research also leans on Gödel's notes and other materials from his *Nachlass*, most of which have been left almost unnoticed. There are two main reasons for why Gödel's papers have not yet been thoroughly studied before. The first is simply the sheer volume of the material. When Gödel died, he left behind a horde of manuscripts, drafts, notes, and other papers, greatly outnumbering his published articles and lectures. These were transferred to the Institute for Advanced Study at Princeton, where, in 1982, John Dawson took up the formidable task of cataloguing the papers. This took him two years to accomplish. The Gödel Editorial Project, which first consisted of Dawson, Stephen Kleene, Gregory Moore, Robert Solovay and Jean van Heijenoort, then started to prepare some of this material for publication.

The *Nachlass* contains boxes and boxes of Gödel's unpublished manuscripts, notes, and correspondence – but also unlabelled loose notes, miscellaneous notebooks on various topics, and personal items. Gödel seems to have never thrown anything away: the endless folders contain not only his scientific notes and manuscripts, but also workbooks from primary school, hotel receipts, and library slips. It would almost be easier to ask what *cannot* be found there than what can – a question still (almost) impossible to answer.

The second reason for why many of Gödel's notes have remained unexamined is their format. Gödel wrote most of his personal notes in a shorthand script called *Gabelsberger* after its inventor, Franz Xaver Gabelsberger. The script was developed in the 19th century and slowly became obsolete in Germany and Austria after the introduction of *Deutsche Einheitskurzschrift* in 1924. Gödel learned the script in school, and he came to use it almost extensively in his notes whenever he wrote in German. This preference is natural, as shorthand writing is much faster than normal handwriting, and a writer trained in shorthand would then easily grow frustrated with the comparably slower pace of longhand writing. Shorthand systems were originally developed for simultaneous transcription of speech that is much faster than writing: the figure for Germanic languages, which are spoken relatively slowly, is 4–6 syllables per second for normal speech, which translates to some 90 to 130 words per minute (see Gebhard, 2012, 17–20). This is several times more than the speed of ordinary longhand writing.

The Editorial Project managed to prepare some of the shorthand material, which was mainly transcribed by Cheryl Dawson who had managed to learn the Gabelsberger system, for publication.[28] The third volume of the *Collected Works* (Gödel, 1995) consists of a collection of Gödel's unpublished lecture notes. Also a great part of the correspondence was published in the fourth and fifth volumes of the *Collected Works* (Gödel, 2003a,b). However, none of the materials in Gödel's notebooks were published (see Dawson and Dawson, 2005).

I started learning to read shorthand in 2017 when given the task to transcribe parts of Gödel's notes for the 1941 Princeton lecture course on intuitionistic logic, which I co-edited with Jan von Plato. The notes were written in English, but many additions were in shorthand German, and moreover, the notebook series called *Arbeitshefte* contained many remarks related to the Princeton course, all written in shorthand. Although my progress was slow at first, by 2019, I was able to read a variety of different texts at a decent speed and accuracy. Most of the notes used for this thesis are my own transcriptions, with some invaluable help from my fellow shorthand readers Jan von Plato and Tim Lethen.

As shorthand reading is very dependent on context – shorthand notes are usually written for the writer themself and not for others, and every writer's style is different – the system is harder to master for a non-native speaker of German, whose vocabulary is not as rich as a native speaker's. In general, transcription work is relatively slow and depends heavily on the richness of the language used. Gödel's notes on logic as well as his proof sketches can be read relatively quickly, as the vocabulary used is very narrow, making them easier to transcribe. On the contrary, his philosophical notes discuss a range of topics from quantum physics to theology, which makes transcription more difficult.

Naturally, both the vastness of the material in Gödel's *Nachlass* and their shorthand format puts limitations on the scope of my data. I have not been able to go through everything Gödel has written in the 1930s and early 1940s. This would be impossible already for the simple reason that Gödel did not write dates on many of his notes, and apart from the notebook series which I will describe below, they are difficult to date. There are also some sources that I have not investigated in full detail because of the complexity of the material and the fragmentary nature of some of the formal notes. I have tried to

[28] For more detail on the history of the Editorial Project, see Feferman (2005).

optimise the situation by an extensive initial review of Gödel's notes before selecting the materials that seemed most relevant for transcription; see **section 1.3.2** below.

There are four major categories in Gödel's *Nachlass* that are of interest to a historian: 1) his scientific and personal correspondence, 2) drafts of articles or lectures, 3) notes for lecture courses, and 4) notebooks and loose notes on various subjects. Because I had to rely on microfilm copies of Gödel's *Nachlass* which do not include his correspondence, I have been able to use only his published correspondence in the *Collected Works*. In what follows, I will briefly describe some of the most important sources as well as choosing and dating the material used.

1.3.1 Lectures and notebook series

As for the lecture notes, there are twelve sets of notes in Gödel's papers, all dated before 1942, of which I chose to investigate only the "Hahn folder" (040025
–040031.5), which contains the lecture notes for a logic seminar led by Hans Hahn, as well as the Princeton Lectures on Intuitionism of 1941 (*PLI*, 040407–040409), which have now been edited for publication and will appear soon in print (Gödel, 2020).[29] Moreover, I have used the draft versions for the "Present situation" of 1933 (040113) as well as those of the 1941 Yale lecture (040263). Most of the notes for the Zilsel lecture of 1938 can be found translated in the *Collected Works*, and the rest of the notes (a preliminary plan for the Zilsel lecture titled "Konzept") have been taken into account in the introduction to the lecture (Sieg and Parsons, 1995). These I have not studied in more depth.

There are four main series of notebooks which are relevant to logic and the foundations of mathematics: *Maximen und Philosophie* (*Max Phil*, 030086–030100), *Resultate Grundlagen* (*RG*, 030115–030119), *Arbeitshefte* (*AH*, 030016–030034), and *Logik und Grundlagen* (*LG*, 030066–030073).

The *Max Phil* series originally contained sixteen notebooks, dated approximately between 1934 and 1955, one of which has unfortunately been lost. The total page count runs at about 1500. The work of publishing these notes was begun only relatively recently by the group at the Aix-Marseille University

[29] In what follows, I will refer to the most often used sources by the abbreviations given here. Item codes for these sources will not be repeated in the text unless necessary for greater precision. Notebooks usually have page numbering, but loose notes do not, and therefore in the latter case, only the item code or abbreviation will be given.

led by Gabriella Crocco. The aim of the project, which lasted from 2009 to 2013, was to investigate the notebooks and study in particular Gödel's idea of concepts and his view of physics. Some results of the project are presented in the 2016 book *Kurt Gödel, Philosopher-Scientist* (see Crocco and Engelen, 2016), which comprises a collection of articles discussing the different themes that arise in Gödel's philosophical notebooks. The first notebook (Gödel, 2019) was published shortly before this dissertation went to print, and so far, only the transcriptions of *Max Phil* 9 and 10 have been made publicly available.

The *Arbeitshefte*, "workbooks," consist mainly of sketches of formal proofs, exercises, and notes. The topics vary from logic to set theory and theory of relativity. The first notebook has the date 1934, and the last *Hefte* are probably from 1943 with some later additions. The *AH* are easier to read than *Max Phil* because they have mostly been written in a rather formal language, and the vocabulary is very limited ("let x be a variable...", "assume that for an arbitrary set A...", etc.). On the other hand, the formal results are often difficult to interpret because their sketchy and fragmentary character, and a reconstruction of many of these results would require special expertise in the given field of mathematics. *Arbeitshefte*, as well as other notes of Gödel's, also contain lists titled "Progr." These are Gödel's *Programme*, plans for things to be investigated and results to be proven, which are useful for getting an idea of the projects which Gödel was working on at a given time.[30] An index of the subheadings in the *AH* can be found in Dawson and Dawson (2005).

The *Resultate Grundlagen*, results on foundations, is also devoted to proofs of logical and mathematical results. In contrast to *AH*, these are carefully prepared, full proofs, and their early sketches can be often found in the *AH*. There are only four notebooks and about 370 pages in total. Only Gödel's finished results ended up in the *RG*, as opposed to many false starts and failed proof attempts in the *AH*. It appears that for the most part, Gödel wrote the notebooks in a very short timespan in 1940–1942, although the last notebook contains some notes that are clearly from a later period.

The six notebooks belonging of series *Logik und Grundlagen* (about 440 pages in total) are also formal in nature, but they contain mainly results of others and sometimes applications of or corrections to them. The *LG* notebooks are sometimes referred to as "Exc." for *Excerptenhefte*; e.g., in the series *RG*, we find several references to "Exc." where the page numbers clearly refer to *LG*.

[30] Sometimes "Progr." seems to be better interpreted as *Progress*, as some of the items in the lists have been marked completed.

Although I have browsed through Gödel's own index for *LG* (030066), I chose not to include these notebooks in my study because they appear to contain little material useful for my purposes.

The category of loose notes is equally vast in scope. The loose notes contain thousands of pages of summaries of the books and articles that Gödel had read, as well as notes from the lectures of others, everything written in shorthand. In addition, there are fragmentary notes on various topics, some several dozen pages long, some written on the back of library slips or receipts. These notes are difficult to interpret and often nearly impossible to date. The folders of loose notes that I have browsed through is a small part of all that there is, and there is no guarantee that some important remarks on intuitionism or constructive foundations from Gödel's early career could not be hiding in one of those unexplored folders. This is not a problem unique to my research; the project of transcribing and studying Gödel's notes has been going on for some decades, and will likely take some decades more. However, I have tried to do my best to identify what seems most relevant to my research.

1.3.2 Choosing the sources

In the first phase of my research, I had to identify the sources most relevant to my study. For the most part, the catalogue of Gödel's papers only indicates the date – if there is any – and the theme of a folder, but little can be said about the actual contents. A large part of the data was previously unexamined territory, of which no documentation exists. Therefore I had to make my way through a considerable amount of material before I could start the transcription work.

In the initial review of the material, I have first examined the series *AH* and *RG*. For *RG*, some of the transcription work had already been undertaken by Jan von Plato, and being able to use his transcriptions made the work easier. Gödel's own index to the *AH* (030016) helped to get an idea of the contents of the notebooks. Moreover, von Plato had also put together a more detailed list of subheadings of the *AH*. Based on these lists, I have read and transcribed parts of *Hefte* 7–9 as well as some material in *Heft* 10. As mentioned, the challenge with this material is not the transcription work but rather reconstructing Gödel's incomplete proofs. Much of the material I encountered went beyond the scope of my expertise and will have to be left for others to reconstruct and interpret.

The challenge in the *Max Phil* notebooks was entirely different: first of all, the rich language and the variety of different topics slows down the transcription work, and moreover, unlike for *AH* and *RG*, there is no index to the notebooks. Together with Tim Lethen, we searched through the notebooks 3–5, written between late 1940 and mid-1942 looking for possibly relevant remarks. I have repeated this procedure several times during the writing of this thesis. Altogether ca. 250 remarks were transcribed, mostly from *Max Phil* 3 and 4, half of it by Lethen and the other half by me. We ended up transcribing plenty of remarks that turned out irrelevant for my purposes; on the positive side, especially the notebooks 3 and 4 have been read through with a tight comb. Nevertheless, we cannot be sure that we might not have missed some relevant passage.[31]

I have mostly excluded the loose notes except for the folders titled "Alte Lit⟦eratur⟧ Grund⟦lagen⟧" (050001–050011), "Alte Literat⟦ur⟧ Math⟦ematik⟧" (050046–050055), "Foundations" (050066), " Logic and Foundations (before 1952)" (050135), "Logic and Mathematics: 'laufend' " (050136–050142), "Memoranda Books" (050236–050240), and "Questions and remarks on foundations of mathematics ('alte')" (*Questions and Remarks*, 060574), none of which were dated when they were catalogued ("before 1952" is Gödel's own addition), but which seem to date mostly to the early 1940s and earlier. Not all turned to be useful; e.g., the Memoranda books are mostly mathematical calculations and exercises, apparently from Gödel's student days.

Of the loose notes, I ended up using only both foundations folders as well as *Questions and Remarks*. The latter turned out particularly fruitful: it is a set of notes containing over 200 questions and comments on various topics including set theory, logic, and foundations. The notes contain no dates, but based on the references and themes – the beginning of the list is mainly set theory, whereas in the end, there is a good handful of notes on intuitionism related to the Princeton course – they seem to be written between the late 1930s and early 1940s, most of them probably dating from 1940 and 1941. The notes are also written on squared pages with punch holes, the type of paper Gödel used in some of his other notes dating from this time. Because most of the remarks were relatively short, I ended up transcribing all of them. Also Gödel's library request slips from 1928–1940 (050173–050183) were used to get a rough picture of what Gödel had read and when. Unfortunately, only a handful of slips from

[31] As for the latter notebooks, after a quick scan it seems that there are increasingly few remarks related to intuitionism in *Max Phil* 6 and beyond. Most of the remarks on the topic are in notebooks 3 and 4, with a handful more in notebook 5.

1941 and after has been preserved.

Spotting the relevant passages in *AH* and *RG* was relatively easy because of Gödel's own lists of contents, as well as the fact that parts of *RG* were already translated. Moreover, *AH* and *RG* contain longhand subtitles, which greatly facilitates the task. Where subtitles are missing, e.g., in *Max Phil* or the loose notes, what has helped in determining relevant parts of Gödel's notes is that proper names, as well as some technical terms, are conventionally written in longhand. For example, "Brouwer" or "Hilbert" would be nearly always written out, and longhand "Int," "Intuit," or "Intuitionismus" were usually used to denote *intuitionistische* or *Intuitionismus*, although sometimes also *Intuition*. There are exceptions, of course: when editing the Princeton Lectures in 2017, we spent a good time trying to decipher what turned out to be the word *Brouwersche* written promptly in shorthand against all conventions. "Finit" or "konstruktiv" have generally not been written out, although "Finitismus" has often been written in longhand by Gödel. Longer shorthand words such as "konstruktiv," "Widerspruchsfreiheit," or "Intuition" are also relatively easy to spot. Another helpful aspect is that Gödel also tended to label his remarks in *Max Phil* ("Bem. Grundl," "Bem. Theol.," etc.), and thus we could mostly skip, e.g., theological remarks, which were highly unlikely to concern intuitionism or constructive mathematics.

Dating the notebooks was rarely an issue. Although Gödel did not often write down dates – e.g., most of the *AH* contain only one or two dates, often to indicate when a notebook was started or the change of a year – most of the notebooks were written in relatively short timespans, often only a few months.[32] *Max Phil* contain dates more often, and it was relatively easy to date the remarks with accuracy up to the range of one or two months. With the loose notes, dating was much more difficult, and one had to consider the themes and possible synchronicities to the notebook series. To give an example, the topics of *Questions and Remarks* had a partial overlap with *AH* 4–10, suggesting that most of them were written between 1940 and 1941. In some cases, such as Gödel's notes on Brouwer's Bar Theorem (050066), one could

[32] Assuming that Gödel filled out the notebooks one after another, e.g., that he did not start filling *AH* 8 before *AH* 7 was finished. This does not always seem to be the case, as Gödel also organised some of his notebooks by theme: e.g., *AH* 9 is devoted to intuitionistic mathematics and the functional interpretation, whereas *AH* 8 is mostly about the theory of ordinals. However, they seem to be loosely chronological: the first entries in *AH* 7 are dated 1st January 1941, the first entries of *AH* 8 are from February, and the most of *AH* 9 seems to date from spring 1941.

only speculate that they had been written in the 1950s since most of the notes in the same folder are from that time; however, I was not able to arrive at a precise date. This type of notes, of which there was a handful, I have not included in my study.

Chapter 2

Formalism and constructivity: the 1930s

Gödel's works on intuitionism and constructive foundations can be divided roughly into three phases. He developed his first logical results between 1931 and 1932, and they shaped his understanding of the nature of intuitionistic logic. In particular, they shaped his belief that intuitionistic logic could and should be interpreted through classical logic, as the differences between the two systems – at least in the case of propositional logic and arithmetic, the two systems Gödel worked on – are so small. The logical works were followed by two lectures in 1933 and 1938, where Gödel introduces what is now known as the Extended Hilbert Programme, a revised and a relativised form of the original programme of constructively justifying non-constructive modes of mathematical reasoning. He criticises intuitionism for having elements that are not ideally constructive and suggests that another alternative for constructive foundations must be found. In 1938, it is mentioned that a finitistic system extended with finite-type functionals fulfils all the criteria for a finitistic or "properly constructive" system. The third phase took place in 1941 when Gödel finally introduced the functional system in a new light as an interpretation and a proof of constructivity for intuitionistic arithmetic. At this point, however, Gödel became more conscious about the fact that his interpretation could also be seen as problematic from the ideally constructive viewpoint presented in 1938. Towards the end of 1941, Gödel had become more open to more natural interpretations of intuitionism as well as the need for irreducibly intensional and possibly non-formal concepts. This marks the beginning of another shift between the viewpoint of the 1941 lectures and that of the *Dialectica* article of 1958.

In this chapter, I will discuss Gödel's views on intuitionism and constructivity in the 1930s, and suggest that they developed into a kind of formalism not too distant from Hilbert's own standpoint. What I mean by "formalism" here can be summarised in four statements, all of which Gödel accepted to some degree:

1. Classical, non-constructive mathematics needs to be justified by a constructive proof of consistency.
2. Consistency is adequate as a criterion of mathematical correctness.
3. Formalisation is necessary for constructive justification, and meaning should have a small role or no role in it.
4. The methods by which the consistency proof should be carried out ought to be as close as possible to original finitism, defined by the finitistic ideals of intuition and evidence.

In particular, steadfast belief in formalisation for the sake of precision and evidence characterises Gödel's early stance on intuitionism, which he accuses of having undesirable elements of "vagueness," "unsurveyability," or "absoluteness." It also affected his understanding of the relationship between intuitionistic and classical logics and their interdefinability. It is also fitting to call Gödel's position formalism in comparison to his later view (see **Chapter 4**), in which also the informal and the intensional – as opposed to the formal and the extensional – have their place in the foundations of mathematics.

Gödel's interest in constructive foundations has sometimes been understood to conflict with his frequent statements that he was a mathematical Platonist from the beginning. In his commentary to Gödel's lecture of 1933, Solomon Feferman characterises Gödel's attitude as surprising, as it "does not seem to square with Gödel's unequivocal assertions [. . .] that he had held a Platonistic philosophy of mathematics since his student days in Vienna" (Feferman, 1995, 39). In which sense he understood Platonism is in itself a complex question which has been discussed quite extensively in the literature (see Parsons, 1995), and one would need to answer it in some way before stating that Gödel's early views were at odds with his self-ascribed Platonism.

Admittedly, there is a pronounced difference between his early viewpoint and his later works, and the still unpublished draft version of "The present situation" contain remarks that are even more hostile towards Platonism. Whereas these remarks do throw some shade on Gödel's own statement that he was a Platonist since 1925, they also underline the fact that Gödel's philosophical focus at the time was primarily in epistemology and not metaphysics. If we read his remarks in the 1933 lecture literally, what Gödel says is that that classical mathematics and its underlying Platonistic metaphysics cannot be assumed without question. No mathematical realist, however, would argue that *because* Platonism is true, we do not need to pay any attention to questions of epistemic justification. Quite the contrary, questions of epistemology – how we can obtain knowledge about mathematical objects – are of great importance to the Platonist because exactly those questions provide the most pressing arguments against mathematical realism.

That said, there is also no explicit endorsement of the realist viewpoint, unless one considers practising classical mathematics as such. Neither can Gödel be called a constructivist, unless one considers seeking epistemic justification for mathematics as a constructive position. One should keep in mind that Hilbert's own attitude towards constructivism was problematic at the very least: his Programme was, after all, set up in order to justify all of classical mathematics so that one could return to practising it as usual, and he was openly hostile towards the idea that mathematics should be restricted to its constructive part. Hilbert believed that finitism was a collection of epistemically responsible methods which could be set aside once this goal was obtained. For now, in light of all of this, I want to suspend the judgement on whether Gödel was a Platonist in the 1930s or not.

Gödel shared Hilbert's idea that epistemically responsible justification was needed, and he also shared the idea that all of mathematics should be practised unrestrictedly. This is apparent when one looks at everything else that Gödel was doing in the 1930s and early 1940s: while calling Platonism unsatisfying and Brouwer's intuitionism not constructive enough, he was happily working on problems of set theory, which he clearly saw as meaningful in a larger foundational context.[1]

Of course, how one defines what is epistemically responsible reveals in itself some philosophical preferences. In order to carry out Hilbert's Pro-

[1] In fact, one reason why it has been challenging to find any notes on Gödel's views on constructivism from 1938–39 is that based on his notes, most of his attention was directed at set theory. In 1938, he proved the consistency of the Axiom of Choice, and in 1939, the consistency of the Continuum Hypothesis. Many philosophical remarks on set theory can still be found from the list *Questions and Remarks*, written around 1940–1941.

gramme, one would also have to believe that a formal reduction and a proof of consistency are enough to prove that classical mathematics is adequate. More importantly, one would need to believe that this formal reduction, or "abstracting away all meaning," preserves what is mathematically relevant. For an intuitionist, this is far from obvious, because even if one can capture most intuitionistic principles by formalisation, formalisations are in principle never complete. Although Gödel understood the intuitionist position, he did not agree with it, which explains the way in which he criticised intuitionism.

I wish to draw attention to one more terminological issue. In the present chapter, when referring to formal intuitionistic systems, I often use "intuitionism" and "intuitionistic logic" interchangeably, which is, especially in the historical context, not entirely appropriate. However, Gödel himself tended to do this, partly because his works were mostly concerned with the logic of intuitionism, and partly because, as mentioned, his early approach placed form over content, at least in the case of intuitionism. In contexts where the difference is relevant, I will use the terms with their usual meanings.

2.1 Early impressions and logical works

Gödel's three formal results were obtained in a relatively short timespan between mid-1931 and early 1932. The two results on propositional logic, i.e., the modal interpretation and the proof that intuitionistic logic does not have an interpretation in finite-valued semantics, were first introduced in Hans Hahn's logic seminar in January 1932. It is not known if Gödel came up with the negative translation at the same time or slightly later, but he presented it for the first time in Karl Menger's Colloquium in July 1932.

Menger and Hahn likely influenced Gödel's views on intuitionism, as they both criticised intuitionism on several occasions. Menger criticised it for making too much out of purely terminological distinctions, and suggested that intuitionistic logic was not actually any more secure than its classical counterpart. Hahn, on the other hand, emphasised the problems with the vague concept of "intuition," which, in his opinion, could not provide a solid foundation for mathematics. Both of these ideas are also present in Gödel's early works, where we can see their development into the criticism of intuitionistic logic in the Cambridge lecture of 1933.

In order to understand this development, one needs to start from the very beginning. Even before the formal works, Gödel was sensitive to questions of constructivity, and he mentions intuitionism in the context of both his dissertation and the incompleteness theorem. It appears that he did not differentiate between finitism and intuitionism at first – as mentioned in **Chapter 1**, this was very common in the 1920s – but soon after came to view the non-formalisability of intuitionistic proofs as the distinguishing factor between the two. After Gödel had started his studies on intuitionistic logic, he became convinced, first by Glivenko's theorem and then by his own negative translation, that intuitionistic logic was closer to classical logic than previously assumed, and that it definitely went beyond finitism.

2.1.1 Non-formalisability of intuitionistic reasoning

Gödel first touch with intuitionism was probably in March 1928, when L.E.J. Brouwer gave two lectures in Vienna (Wang, 1987, xx). The two talks were also Brouwer's last public lectures for quite a while; in the 1930s, he wrote little on intuitionism. Gödel was likely present during these lectures, as he knew their content to some extent (see also Carnap's diary entry below), although he could also have discussed them with someone who attended them.

The lecture of 10th March, "Mathematik, Wissenschaft und Sprache" (Brouwer, 1929) is an elucidating description of the intuitionistic view of the construction of mathematics and the function of (mathematical) language. Brouwer pays particular attention to the nature of language as a tool of will-transmission and the failure of philosophers and mathematicians to see it as such, leading them to believe in an underlying reality of concepts and relations. Whereas classical reasoning led to paradoxes in set theory, the formalists' attempt to fix the situation was to prove that the *language* of set theory and classical mathematics is, in general, free from contradiction – which, of course, does nothing to ensure its correctness. Finally, Brouwer gives a few examples of *fliehende Eigenschaften* such as the one given in **Chapter 1**. He ends with a more optimistic note, saying that a correct and consistent formal theory of intuitionistic mathematics can be developed independently of the problematic Principle of the Excluded Middle. The second lecture, given on 14th March, is titled "Die Struktur des Kontinuums" (Brouwer, 1930) . In the lecture, Brouwer criticises the former approaches to the theory of real numbers and gives a summary of the properties of the intuitionistic continuum. Here, too, counterexamples based on fleeing properties are used to justify the intuitionistic principles of analysis.

Brouwer's influence can be seen in some aspects of Gödel's early works. The idea of mathematics as necessarily incomplete and intuitionistic proofs as open-ended and non-formalisable appear for the first time in his dissertation. Brouwer is, in fact, mentioned already in the opening paragraph of the thesis (Gödel, 1986, 60).[2] Gödel frames the completeness proof around the debate between Hilbert and Brouwer and appears to side with Brouwer. The implication from formal consistency of a system to the existence of a model is not obvious, as it not only presupposes that the there is no proof of unsolvability of any mathematical problem but also that there is no proof that some problem is unsolvable by means of a given formal system.

Gödel remarks that the completeness proof does use classical modes of inference, in particular the Principle of Excluded Middle, which an intuitionist would identify with the principle of solvability of every mathematical problem. Indeed, the completeness proof would not go through if the underlying logic was interpreted in an intuitionistic sense. This is because if the the concept of proof replaced the concept of truth, the completeness proof would become equivalent to a proof of decidability, which is naturally not possible.

There are more fundamental reasons for why there is no way to obtain such a proof for intuitionistic logic. Namely, the intuitionistic concept of provability does not refer to provability within a formal system, but provability "by any means imaginable" (Gödel, 1986, 65). This is to say that some intuitionistic proofs could fail to be contained in any formal system. This was the view Gödel also took in the later works, although in the dissertation he does not express the same dissatisfaction with intuitionistic provability than in his lectures of 1933 and 1938. In general, Gödel's earliest remarks about intuitionism are all of neutral character, as opposed to the more critical stance of his logical works as well as the lectures.

It should also be mentioned that Gödel employed the idea of a Brouwerian counterexample in his refutation of Heinrich Behmann's thesis that every constructive proof can be transformed into a non-constructive one.[3] His example is a sequence of rational numbers which depend on Goldbach's conjecture: let the nth member of the sequence be $\frac{1}{n}$ if all even numbers less than n are sums

[2] Unfortunately, the introduction, which discusses the philosophical motivation of the dissertation, is not included in the subsequent publication in the *Monatshefte für Mathematik und Physik*.

[3] See chapter 5 of Mancosu (2010) for a detailed presentation of Behmann's idea and Gödel's critique.

of two primes, and $1 - \frac{1}{n}$ otherwise. It easily follows that the sequence has an accumulation point of either 0 or 1, but without a solution to Goldbach's conjecture, it cannot be decided which. He uses the same example also in the 1941 Princeton lecture course on intuitionism.

Rudolf Carnap wrote in a diary entry of 23rd December 1929 that he and Gödel had discussed the "inexhaustibility of mathematics." Carnap named Brouwer's Vienna lectures as an influence behind this idea and added: "Mathematics is not completely formalizable. He [Gödel] appears to be right" (Wang, 1987, 84). Formal incompleteness is a necessary consequence. Gödel first presented the idea of the first incompleteness theorem in the Königsberg meeting in September, 1930. In the presentation, only the idea of the theorem is sketched, and there are no philosophical remarks whatsoever.

On 20th November, Johann von Neumann wrote to Gödel that he has discovered that the incompleteness theorem could also be used to prove that the consistency of a formal system is unprovable within the system itself (Gödel, 2003b, 336). A short exchange of letters ensued in which the implications of the theorem to Hilbert's Programme and intuitionism were discussed. In the beginning, Gödel was uncertain of whether the incompleteness theorems applied to constructive mathematics. In an undated letter draft to von Neumann, written in reply to the letter of 20th November, he writes that he is "fully convinced that there is [cancelled: a finite] an intuitionistically unobjectionable proof of freedom of contradiction for classical mathematics [added above: and set theory], and that therefore the Hilbertian point of view has in no way been refuted" (von Plato, 2019, 4051). At this point, neither Gödel nor von Neumann seems to distinguish between Hilbert's finite viewpoint and intuitionism.

Gödel presented both incompleteness theorems for the first time at the meeting of Schlick's Circle in Vienna on 15th January 1931, and the conversation that took place after the presentation has been recorded by Rose Rand (see Mancosu, 2010, 235–236). Unsurprisingly, the same question about the applicability of the theorems to constructive systems came up at several places. In contrast to the letters to von Neumann, here Gödel makes some distinctions between different forms of constructivity. Von Neumann's conviction that Gödel's proof undermines Hilbert's Programme is mentioned, although Gödel is again cautious. The weak point of von Neumann's argument, he says, is that there is no guarantee that finitary methods find a formalisation inside a single formal system.

This is not to say that finitary proofs are not formalisable, or that they are not formalisable in Peano Arithmetic. However, the fact that the set of PA-proofs is recursive, i.e., one can list all PA-proofs, does not mean that one can single out the set of finitarily acceptable theorems. A PA-theorem A could have an arbitrarily complex proof that will appear *somewhere* in the list of proofs, but we have no way of telling where. If a PA-theorem has a finitarily acceptable proof, therefore, it will appear on the list, but if it does not, there is no way of telling. It should be noted that in 1931, although Gödel and others knew of Skolem's Primitive Recursive Arithmetic, the system was not generally equated – as it now is – with Hilbertian finitism.

As for the formal system of Heyting's arithmetic, Gödel answers that Heyting's system is narrower than the system of the *Principia Mathematica* considered in the incompleteness proof and that if Heyting Arithmetic is ω-consistent, then the result applies.[4]

Gödel does seem to distinguish Brouwer's mathematics from both Hilbert's and Heyting's systems. At the end of the discussion notes, we read that (Mancosu, 2010, 236)

> Finally to a remark by Kaufmann Gödel expresses his opinion: that Intuitionism according to Brouwer's conception is not touched by his work because it does not want to be contained in any formal system.

The non-formalisability of Brouwer's mathematics is stated with certainty that Gödel did not attribute to the case of finitism. However, he could not yet pronounce where intuitionism differed from finitism, save from finitism being possibly formalisable and the whole of intuitionistic mathematics in principle non-formalisable.

In the published incompleteness article, there is a only a short mention of the issue of the second incompleteness theorem and constructive mathematics, this time a reference to Hilbert (Gödel, 1931, 195):[5]

> I wish to note expressly that Theorem XI [...] [does] not contradict Hilbert's formalistic viewpoint. For this viewpoint presupposes only the existence of a consistency proof in which nothing

[4] An arithmetical theory is called ω-consistent if it is not the case that for some A, the theory proves $A(0), A(1), \dots A(n), \dots$ for all natural numbers n and that it also proves $\neg\forall x A(x)$.

[5] Page numbering refers to the English translation in Gödel (1986).

> but finitary means of proof is used, and it is conceivable that there exist finitary proofs that cannot be expressed in the formalism of P [i.e., Peano Arithmetic] [. . .]

Here Theorem XI refers to the Second Incompleteness Theorem. Intuitionism is not mentioned anywhere, although the reason is probably that the theorems carried more weight in the context of the Hilbert Programme, whose main objective was to obtain a consistency proof for classical mathematics.

Gödel soon changed his mind about Hilbert: in the "Lecture on undecidable propositions" (040405–040406) that he gave in Bad Elster on 15th September 1931, the final remark has changed into a negative one (p. 8):[6]

> The formalistic school searches for a proof of freedom from contradiction of classical mathematics by finite means, i.e., there must occur in the proof only decidable properties and computable functions, and what is called the existential way of inference must not be applied anywhere. But all such finite ways of inference are easily formalisable in the system of classical mathematics and it is not at all foreseeable today how one could find ones that are not formalisable, even if one cannot exclude this with absolute certainty.

It seems that at this point, it was clear to Gödel that Hilbert's finitistic methods are different from intuitionism, and that the methods of the latter go beyond what is finitistically acceptable. At the time of the Bad Elster lecture, he had already started to study intuitionistic logic more closely, and he soon began to believe that intuitionistic logic has a close relationship to classical logic and was, in some contexts, equally strong. This was proven for arithmetic in 1932, showing concretely that intuitionistic methods surpass finitary ones: classical and intuitionistic arithmetic are equiconsistent, but one cannot prove the consistency of classical arithmetic by the methods used by the finitists in the 1920s.

2.1.2 First studies in intuitionistic logic

By the summer of 1931, Gödel's focus was on intuitionistic logic. From his library slips (050176), we can see that he read both Glivenko's works of 1928 and 1929 as well as Heyting's 1930 articles on the axiomatisation of intuitionistic

[6] Translation by von Plato (2020).

propositional and predicate logic, all of which he borrowed in July 1931.[7] He presented his results on intuitionistic logic for the first time in Hans Hahn's logic seminar in January 1932, where he gave two lectures on intuitionistic logic. In these lectures, Gödel introduces both the modal interpretation of intuitionistic propositional logic and the proof that intuitionistic logic does not have a finite-valued semantics. The negative translation must have been developed around the same time or slightly later, as Gödel's gave his talk at the Menger Colloquium in June 1932.

One could say that the three concise works contain little philosophical or foundational commentary; especially the papers on propositional logic seem more like formal exercises than pondering upon the meaning of intuitionistic logic. Whereas Gödel's early works were leaning more towards mathematics than philosophy, his unpublished notes show that he did think about his results in a larger context of looking for interconnections between intuitionistic logic and classical reasoning. However, contrary to his earliest remarks about intuitionistic reasoning, Gödel reads intuitionism mainly through its logic. This would then lead him to dismiss intuitionism as a constructive foundation for mathematics, as its formalism comes too close to classical logic.

There are two sets of typewritten lecture notes for the Hahn seminar in Gödel's *Nachlass*, probably written down by Paul Boschan, whose surname is written on one of the envelopes which contain the lecture notes. The notes on Gödel's lectures on intuitionistic logic show that he presented both of his results on propositional logic. In addition, there are two handwritten summaries of Gödel's seminar lectures on intuitionistic logic along with several derivations in the intuitionistic formalism as well as sketches of the two results on propositional logic (040028, 040029) as well as 11 pages of shorthand notes related to intuitionistic logic (040025), clearly written in preparation for the Hahn lectures. As opposed to the formal style of the published papers as well as the polished typewritten notes, the shorthand notes have a more conversational tone. Therein we find Gödel's first thoughts on the motivation behind intuitionistic logic and the nature of intuitionistic reasoning. The notes also help explain why he attacked intuitionistic logic so heavily in his lectures of 1933 and 1938.

Gödel starts with Brouwer's objections to the Principle of the Excluded

[7] In the folder "Logic and foundations (before 1952)" (050135) we find careful notes on Heyting's article series, although these seem to be rather written in mid- or late 1930s, not early 1930s.

Middle, the consequences of which Brouwer has "treated systematically" in his work (040025, p. 1). Heyting, then, takes an axiomatic approach to Brouwer's critique, building a formal system that does not include the problematic principle or its consequences. The intuitionist, Gödel writes, does not deny PEM in the sense of affirming its negation, as this would mean that there are sentences that are *neither true nor false*, which is not what Brouwer means. Rather, Brouwer's opinion is that PEM is neither *evident*, nor does it lead to consequences that are evident. What PEM asserts is that there is a procedure for deciding for every sentence whether it is true or false. Since this is not the case, PEM is not a tool that can be used in every mathematical context.

But the situation, then, is this: intuitionists do not deny anything that the classical logician says, but rather, they choose not to assert some things that the classical logician would assert. Gödel emphasises the difference between these two positions by comparing them to Euclidean and non-Euclidean geometry. In non-Euclidean geometries, one can derive the falsehood of the parallel postulate; a case analogous to that of intuitionism would be Euclidean geometry *without* the parallel postulate.

Because intuitionism does not arise from negating theorems of classical logic, it is not obvious how to come up with the correct axioms of intuitionistic logic (p. 1):

> This task (building such an axiomatic system) is not entirely unambiguous, because one can never say something like, one omits the Principle of the Excluded Middle from classical axioms *like one does with the parallel axiom in non-euclidean geometry*, as the sentence *Tertium non Datur* does not appear in the commonly adopted logical axioms at all but is at first derived from them. Heyting's solution is then only one possibility and why this solution is chosen in particular would only be grounded by Heyting on intuition.[8]

In a sketch of a plan for the lectures on intuitionistic logic (040029), Gödel has written "Heytingsche Aufgabe mehrdeutig," referring to this ambiguity

[8] Diese Aufgabe (ein solches Axiomsystem darstellen) ist durchaus nicht eindeutig, denn man kann nicht etwa sagen, man lässt den Satz vom ausgeschlossenen Dritten aus den Axiomen der gewöhnlichen Logik fort, *wie man dies etwa beim Parallelenaxiom und der nichteuklidischen Geometrie macht*, weil der Satz Tert⟦ium non datur⟧ unter den gewöhnlich angenommenen logischen Axiomen gar nicht vorkommt, sondern aus ihnen erst abgeleitet wird. Die Heyting'sche Lösung ist also nur eine der möglichen, und dass gerade sie gewählt ist, wird bei Heyt⟦ing⟧ bloss durch Intuition begründet.

of coming up with a system of axioms matching Brouwer's ideas. He is correct: as was mentioned in **Chapter 1**, Kolmogorov did not accept Ex Falso Quodlibet, and there are good reasons to believe that the principle does not follow from Brouwer's interpretation of implication, either (Van Atten, 2009, 128–130). Nevertheless, Heyting chose to include this principle.

The picture of intuitionistic logic given here is that one gets an intuitionistic theory by considering the corresponding classical theory and then *leaving out* what does not seem correct. Likewise, Gödel characterises Brouwer's position in negative terms by listing which classical principles he does not accept, even though he does mention the interpretation of PEM through provability. Whereas Brouwer did spend quite some energy on criticising classical logic, Gödel's deflationary account of intuitionism and its logic puts little weight on the reasons why he did so. Instead, he says that the classical principles Heyting chose to omit were based only on Heyting's own intuition.

Gödel then goes on to discuss the principles of Excluded Middle and Double Negation Elimination, both of which hold in intuitionistic propositional logic in slightly weakened forms: $\neg\neg(a \vee \neg a)$ and $\neg a \equiv \neg\neg\neg a$ are both theorems of intuitionistic propositional logic, as already noticed by Brouwer. The rejection of DNE is also central to intuitionism, as it is grounded in Brouwer's critique of indirect proofs, in particular those of existential statements. Gödel also notes that it seems as though a form of "Excluded Fourth" could hold in intuitionistic logic, namely $a \vee \neg a \vee \neg\neg a$ (p. 6), although this turns out not to be the case. Elsewhere (040028), he shows that this is because $a \vee \neg a \vee \neg\neg a \equiv \neg a \vee \neg\neg a$ is derivable in intuitionistic logic. This is essentially the argument used by Barzin and Errera (1927) in their attempted refutation of Brouwerian logic.

Gödel notes that the interpretation of negation is where intuitionistic logic differs the most from the classical, but with negation-free statements, there is little difference: e.g., the classical rules of commutativity and associativity of $\vee$ and &, which contain no negation, are also intuitionistically valid.[9] It turns out, however, that there is a way to classically interpret sentences containing negation so that there is nothing in classical logic that could not be formulated

[9] "Das wäre der Unterschied zwischen dem H und dem gewöhnlichen Aussagenkalkül. Sie zeigt sich besonders in formalen Schlüssen in denen die Negation vorkommt. Dagegen besteht in den anderen Formeln verhältnismässig wenig Unterschied. Es gilt z.B. für $\vee$ und $\wedge$ die com. assoc. und auch die beide distrib. Ferner kann man z.B. Implikationen zueinander addieren und miteinander multiplizieren, d.h. $p \supset q \quad r \supset s$ kann man schliessen auf $p \wedge r \supset q \wedge s$." (040025, p. 5)

in intuitionistic logic as well.

The close relationship between classical and intuitionistic propositional logic is shown by Glivenko's translation, by which for every classically provable A, one has $\neg\neg A$ intuitionistically provable. With Glivenko's theorem, Gödel argues (040025, p. 6–7), we can show that intuitionistic concepts have a classical interpretation. Namely, one first observes that every sentence of the form $\neg A_1 \wedge \neg A_2 \wedge \ldots \wedge \neg A_n$ that is a theorem of classical propositional logic is also a theorem of intuitionistic propositional logic by Glivenko and the intuitionistic theorem $\neg A \equiv \neg\neg\neg A$. Now, every classical sentence can be expressed using only the connectives $\neg$ and $\wedge$. Reinterpreting the other connectives by

$$a \supset b = \neg(a \wedge \neg b) \text{ and } a \vee b = \neg(\neg a \wedge \neg b)$$

one gets a mapping of classical logic onto a fragment of intuitionistic logic.

In a fragmentary passage (p. 6–7), Gödel writes that

> [by] this assignment to the intuitionistic basic concepts, it is proven to be a part of classical logic. From this, it appears that by the intuitionistic criticism, nothing at all from classical logic is lost, but on the contrary, one can understand the issue simply as certain reinterpretations taking place in intuitionism.[10]

Gödel, therefore, deduces from the translation that intuitionistic logic only re-explains classical concepts in another way, reducing the debate between intuitionism and the classical conception to a quarrel over terminology.

Gödel next poses the question whether one can reinterpret intuitionistic logic in classical terms so that the original meaning of intuitionistic connectives is retained. It is possible, he says, that there are classical *concepts* which capture the properties of intuitionistic logic (p. 7):

> I come now to the possible interpretations of intuitionistic propositional logic, i.e., to the question of how one should conceive of the basic intuitionistic connectives so that only the theorems of Heyting's calculus hold and nothing else. One can perhaps come up

[10] Und es erweist sich also bei dieser Zuordnung der Grundbegriffe der intuitionistische Aussagenkalkül als Teil des klassischen. Daraus hervorgeht, dass durch die intuitionistische Kritik gar nichts vom klassischen Aussagenkalkül verloren geht, sondern man im Gegenteil die Sache so auffasst, dass im ⟦add.: statt der⟧ Intuitionismus nur gewisse Uminterpretationen stattfinden.

> with some concepts within *classical* logic which have the properties of the intuitionistic $\vee$ $\neg$ etc.[11]

Gödel first considers an interpretation by many-valued logic, which turns out to be impossible. The presentation follows that of the published paper. Gödel then suggests that there is another alternative which formalises the concept of provability (p. 10). This alternative is his modal interpretation; unfortunately, the notes end here, with only a summary of the modal translation of intuitionistic propositional logic. The full formal presentation of the modal system can be found elsewhere in the Hahn notes (040028); again, there are no essential differences to the published version.

It is a stretch to call **S4** a classical interpretation of intuitionistic logic, although the system has classical semantics. It defines an intuitionistic concept of provability, which Gödel does not explicate any further – all we are told is that it cannot be read as "provability within a system," as this will not hold for some systems via the Incompleteness Theorem. Moreover, he does not count the modal translation as a classical interpretation in his other works.[12] In 1938, e.g., he discusses the proof interpretation of Heyting Arithmetic in terms of the modal system, assuming that the provability predicate has the characteristics of intuitionistic provability, even if he does remark that the predicate can also be considered decidable "with enough good will" (Gödel, 1938, 101). Unlike in the case of the negative interpretation, there is no mention of the modal system in the 1933 and 1938 lectures.

2.1.3 Towards reinterpretation of intuitionistic logic

What can be read from the Hahn lecture notes is that Gödel was somewhat acquainted with Brouwer's ideas, although it is not certain where he had encountered them. Brouwer's critique of language or his approach to mathematics as an incomplete collection of mental constructions – which are both

[11] Ich komme nun zu den möglichen Interpretationen des intuit. Aussagenkalküls, d.h. zur Frage, was kann man unter den intuit Grundbegriffen verstehen, damit gerade die Sätze des Heyt. Kalkül und keine anderen gelten, und ⟦man⟧ kann vielleicht irgendwelche Begriffe innerhalb *der klassischen Logik* angeben, welche die Eigenschaften des intuit. $\vee$ $\neg$ etc haben.

[12] Browsing *Arbeitshefte*, one notices that around 1941–1942 Gödel wrote dozens of pages on the interpretation of Heyting's logic in Lewis' modal system. This is not because he was interested in the modal interpretation for its own sake, however, but it rather appears that he aimed to find a topological model of type theory to prove the independence of the Axiom of Choice. The Gödel Editorial Project did transcribe some of the notes in *AH* 14 and 15, but were left uncertain of their contents (Dawson and Dawson, 2005, 159).

present in the Vienna lectures of 1928 – are nowhere mentioned, hinting that Gödel either did not see these aspects as relevant to the development of intuitionistic logic or that his knowledge of Brouwer's works came from other sources. As Gödel's papers give no evidence that he would have read Brouwer before the 1940s, which is consistent with what Gödel reported in the 1970s (Gödel, 2003a, 447), it is likely that, apart from Brouwer's 1928 lectures that he might have attended, he got his information from secondary sources.

Gödel's library slips and the few notebooks that can be found from 1931–1932 tell us something about his readings on intuitionism. In addition to Heyting, Gödel was acquainted with Hermann Weyl's "Die heutige Erkenntnislage" (Weyl, 1925) as well as the book *Philosophie der Mathematik und Naturwissenschaft* (Weyl, 1926).[13] Weyl had an idea of intuitionistic logic that was not entirely in line with Brouwer's, and neither he nor Heyting discuss Brouwer's mathematics as a thought-construction in very much detail.[14] Gödel had also read Hilbert (see **section 2.2.1**), whose conception of intuitionism came from Weyl, as well.

Of course, there were people around Gödel such as Hahn, Menger, and Bernays, who were interested in intuitionism and constructive logic, and the topic might have come up in informal discussions. Unfortunately, we can find no record of the conversations that might have taken place in this early period, and therefore one can only make educated guesses about the influence that different people had on him.

From the likely sources of influence, Karl Menger is worth discussing in some detail because of the similarity of his views on intuitionism and those of Gödel. Menger was a young professor at the University of Vienna who brought together a group of students as well as younger mathematicians which met regularly to discuss mathematical problems (Sigmund, 2011, 78). According to Wang (1996, 75), Gödel attended Menger's seminar for the first time on 24th October, 1929, and remained an active participant thereafter. Some of Gödel's formal results might well have been originally suggested by Menger, who had a habit of giving the participants of his colloquium problems he

[13] Weyl's "Erkenntnislage" has been taken out of the library in November 1932 and we find notes referring to the article and the book in Gödel's early *Excerptenhefte* (030079, 030085) which are dated around late 1931 and 1932.

[14] In 1928, Gödel had also requested the issues 31 and 33 of the *Revue de Métaphysique et de Morale* which contain Wavre's articles on PEM and intuitionistic criteria of existence (Wavre, 1924, 1926a,b); however, the library slip is not stamped, so one cannot be sure whether he ever saw the papers.

thought would be interesting to solve.[15]

Menger himself had a strong connection to intuitionism, as he worked as Brouwer's assistant in Amsterdam from 1924 to 1927. Whereas originally sympathetic to Brouwer's viewpoint, he later grew suspicious of it, especially because of the notion of constructivity, which, according to Menger, is not made precise in Brouwer's work (Hesseling, 2003, 198–199). In his 1930 article, "Der Intuitionismus," Menger blames Brouwer for putting too much weight on terminological distinctions. In his counterexamples, Brouwer defines a real number r_P dependent on an unsolvable problem P, so that "in his parlance the proposition 'neither $r_P = 0$ nor $r_P \neq 0$' is true today, whereas tomorrow, if P is solved, that proposition will be false and 'either $r_P = 0$ or $r_P \neq 0$' valid" (Menger, 1930, 54).[16] He concludes that "Brouwer thus uses the words 'true' and 'false' in a time-dependent, subjective manner." He continues by remarking that Brouwer's calling an unproven theorem "false" is analogous to his criticism of the phrase "there is," where no instance is given. Menger notes that the debate on the notion of existence could have been avoided had it been noticed that it involves two different conceptions of existence. These are already present in many natural languages: there is/there exists, *es gibt/es existiert*, *il y a/il existe* (ibid.). Failing to appreciate the greater expressive power that being able to distinguish between these concepts gives to intuitionistic logic, Menger reduces the debate to playing semantics.

A more fundamental problem lies within the notion of constructivity, which, says Menger, has never been adequately defined. All that is certain, he says, "is that the constructivity requirements of various mathematicians differ substantially from one another" (Menger, 1930, 56). However, he does suggest that this vague notion could be made precise in each of the different ways that it is used (ibid., 56–57):

> For each of the various versions of constructivity one could develop a corresponding deductive mathematics. For example, it is perhaps possible to give a constructivity principle so strict that it would allow only finite sets, or a somewhat weaker one which would include countable sets, or a weaker one still which would admit analytic sets, or a very general one which would allow arbitrary sets

[15] I thank Matthias Baaz for pointing this out to me.

[16] Page numbering refers to the English translation in Menger (1979).

> of real numbers. The requirement of consistency may in this sense be considered the weakest possible constructivity principle.

Menger sees constructive mathematics as a spectrum of stronger and weaker views, and he criticises Brouwer for dogmatically acknowledging only one. Moreover, he holds the belief that it is possible and recommendable to make the notion precise individually in each of the different contexts where it is used.

Another critic of intuitionism was Hans Hahn, who was also close to Gödel in the early years. Hahn's criticism extended to all forms of constructivism which used intuition as a primary source of mathematical knowledge. In a 1933 lecture titled "The crisis in intuition" (Hahn, 1933),[17] Hahn argues that Kantian *a priori* intuition contradicts clearly valid geometrical results. Hahn's idea of intuition, however, is not very precisely expressed, and it does not seem to relate to Brouwer's view of intuition, which went substantially beyond Kant.

In "Does the infinite exist?"(Hahn, 1934), published a year later, he takes a slightly different position. Like Menger, he accuses the intuitionists of ambiguity, claiming that the intuitionists do not give a clear answer to the question of which kinds of infinities are allowed. The "intuitionistic viewpoints" that he considers, though, represent finitism and intuitionism. The first viewpoint, according to which arbitrarily large numbers exist but neither the set of all number nor any non-denumerable infinities exists, is Hilbert's. Objecting to the second statement, Brouwer represents the second viewpoint that Hahn mentions, accepting that some non-denumerable sets exist. (Naturally, one can start debating over the meaning of "exists," but this would be missing the point: the intuitionist can, in any case, *operate* with real numbers whose equality is undecidable, whereas the finitist will not.) Hahn concludes that, as even the intuitionists do not agree on the extent of intuitionism, "the intuitionistic doctrine is thus seen to rest on very uncertain ground; in glaring contrast to this uncertainty is the gruffiness with which the supporters of this position declare meaningless everything that in their opinion is not constructible by pure intuition" (Hahn, 1934, 117). Given that Brouwer was never uncertain between the two viewpoints, half of the criticism is misplaced, although the accusation of dogmatism is justified to a degree.

[17] Page numbering in Hahn (1933, 1934) refers to the English translations in Hahn (1980).

Whichever way Hahn interpreted or misinterpreted Brouwer's view of intuition, it is clear that he did not see *any* kind of intuition as an adequate foundation of mathematics. In the "Diskussion zur Grundlegung der Mathematik" which took place in the Königsberg conference on September 1930, he states that "Mr. Heyting in his paper (Heyting, 1931) started out from a primitive intuition of the number series; to me there is something mystical about this primitive intuition, as there is about pure intuition or intuition of essences, and it is not therefore a suitable starting-point for the foundations of mathematics" (Hahn, 1980, 32). It is no surprise that a logical empiricist would deny the possibility of intuition of essences or abstracts, which does not belong to the realm of the analytical *a priori*.

Gödel's early view on intuitionism has elements from both Hahn and Menger. He was critical towards intuitionism not only because it is vaguely defined, but like Hahn, he believed that it possesses inherent vagueness in its open-ended notion of a proof or a construction. However, like Menger, Gödel believed that intuitionistic ideas could be made more explicit at least within certain realms of mathematics. The idea that there are different "layers" of constructivity is also strongly present in his philosophical lectures of 1933, 1938, and 1941. Moreover, Gödel's view of intuitionistic logic as a reinterpretation of classical logic is of a similar vein as Menger's statement of the debate between classical and intuitionistic logic being terminological. In a sense, Gödel's early logical works can be viewed as "making Menger's impressions [on intuitionism] more precise," as Kreisel (1987, 146) put it.

If one sees intuitionism as a product of leaving something out of classical logic, as opposed to a formal approximation of mathematics starting with concepts entirely different from the classical ones, then it makes sense to ask how far exactly do we get with the removal of this something. Indeed, Heyting did (falsely; see footnote 13 in **section 1.1.2**) state that he got his axioms from going through *Principia Mathematica* and omitting those that seemed unacceptable. But it then makes sense to ask whether the omission actually makes a big difference. Gödel saw the translation of intuitionistic connectives in classical terms as a proof that, at least in the case of propositional logic, the difference is smaller than expected. Next, he would show the same for arithmetic, presenting a classical interpretation similar to Glivenko's interpretation for propositional logic.

The negative interpretation of Heyting Arithmetic was a result that Gödel would mention in all of his lectures on intuitionism, even developing a new variant of it in his Princeton Lectures of 1941. The Menger Colloquium paper (Gödel, 1933c) is slightly longer than the published versions of the two other logical results. What appears to be an early draft of the published paper can be found in the Hahn folder (040027.5) with the note "Wien 3./XII Dr. Kurt Gödel." The draft contains the formal result presented nearly identically to the published version (see **section 1.2.1**); however, in the latter, the following passage has been added at the end (Gödel, 1933c, 295) :

> Theorem 1 [of intertranslatability of HA and PA] [...] shows that the system of intuitionistic arithmetic and number theory is only apparently narrower than the classical one, and in truth contains it, albeit with a somewhat deviant interpretation. The reason for this is to be found in the fact that the intuitionistic prohibition against restating negative universal propositions as purely existential propositions ceases to have any effect because the predicate of absurdity can be applied to universal propositions, and this leads to propositions that formally are exactly the same as those asserted in classical mathematics.

The negative translation result was undoubtedly to Menger's liking, because as Golland and Sigmund (2000, 39) put it, "Gödel managed to prove that intuitionistic mathematics is in no way more certain, or more consistent, than ordinary mathematics." The addition quoted above underlines that this interpretation matched, at this point, Gödel's own.

Whether Gödel's turn to questions of constructive foundations was motivated by this discovery of the apparent inadequacy of intuitionism is not certain. However, the "Present situation" of 1933 is a natural continuation of Gödel's early arguments. In a sense, here the early impressions on intuitionistic proofs and on the connection with classical logic take the form of a programme. If one considers intuitionistic logic from the point of view of constructive foundations, then the result that it mirrors classical logic closely is not very reassuring when it comes to the constructivity of intuitionism. Likewise, the non-formalisability of intuitionistic proofs turns into a critique of the standard interpretation of intuitionistic operations in terms of their provability conditions. The incompleteness theorems had shown that Hilbert's original programme could not be carried out, but if intuitionistic logic is almost no better than the classical variant, what options are left?

2.2 What is a constructive logic?

After the logical works, Gödel started to consider more foundational questions concerning intuitionism and constructive mathematics. In his lectures of 1933 and 1938, he talks about consistency proofs and constructivism in the post-incompleteness era. First of all, he presents what he calls a strict framework for constructivity, a list of properties that a logical system has to satisfy in order to be properly constructive. In 1938, after criticising both Heyting Arithmetic and Gentzen's 1936 consistency proof for arithmetic, he suggests that there is a more constructive way to prove the consistency of classical arithmetic. This alternative, a system based on finite type functionals, would not be presented in detail until 1941.

The two foundational lectures of 1933 and 1938 are framed in the context of what is nowadays called the Extended Hilbert Programme. In the Extended Programme, one has to let go of the idea of finding a universal solution and seek for *localised* solutions instead. Namely, the aim should be to find constructive foundations for particular mathematical systems. In order to make sense of such a task, one has to consider, as Menger suggested, constructivity as a spectrum of weaker and stronger views. On Gödel's spectrum, the original Hilbertian finitism is, as an ideally constructive standpoint, at the very bottom of this hierarchy of constructive systems: it is the "gold standard" against which other constructive approaches are evaluated. The aim of the Extended Programme, then, is to find, for each specific classical theory, a theory as constructive as possible, in which the consistency of the classical theory can be proven.

2.2.1 Formalistic ideals of constructivity

"The present situation on the foundations of mathematics" sketches a picture of what is effectively the Hilbert Programme in the post-incompleteness era. The incompleteness theorem had not only shown that the methods used so far by the finitists could not even prove the consistency of arithmetic, but also that there is no single formal system to prove the consistency for all of formalised mathematics. Therefore one needs to either give up the idea entirely or revise the goals of the Programme.

Gödel's earliest works show that he was well acquainted with Hilbert's ideas. Additionally, he had read "Über das Unendliche" (Hilbert, 1926) carefully and made 14 pages of notes (050135) on the paper in the 1930s,[18] which carefully explicate the finitist standpoint on real and ideal objects, intuition, and quantifiers. Half of the pages are dedicated to Hilbert's proof sketch for the Continuum Hypothesis, which Gödel has carefully reconstructed. In the same folder, there are also notes on the 1930 reprint of the 1927 lecture "Die Grundlagen der Mathematik" (in Hilbert (1930)). Gödel has heavily underlined the quote (p. 3):

> Taking the principle of excluded middle from the mathematician would be akin to denying the astronomer the telescope, or the boxer, the use of fists. The prohibition of existential statements and the principle of excluded middle amounts to the renouncement of the science of mathematics in general.[19]

as well as the final paragraph of Hilbert's article (p. 5):

> Mathematics is a presuppositionless science. For its foundation, I do not need God, like Kronecker, or a special faculty of our understanding attuned to the principle of complete induction, like Poincaré, or the Brouwerian primordial intuition, and neither do I need, unlike Russell and Whitehead, axioms of infinity, reducibility, or completeness, which are, in fact, actually contentful, and moreover not at all plausible, presuppositions.[20]

---

[18] This is in the same folder, "Logic and the foundations (before 1952)," with the notes on Heyting (1930a,b,c). The notes on Hilbert (1926, 1930) are written on the squared "punch hole paper" that Gödel used in the mid to late 1930s, as opposed to the ruled paper that he used in almost all of his 1940–1942 notes. Filed with his library slips of 1937–1938, we find a note which refers to "Ann⟦alen⟧ 95, 99" – Hilbert (1926) was published in volume 95 of *Mathematische Annalen* – and "Hilbert Grundl. d. Geom letzte Aufl." (Hilbert, 1930) have been returned on 4th February 1938. Apparently, Gödel had loaned Hilbert's works in preparation for the Zilsel lecture.

[19] Dieses Tertium non datur dem Mathematiker zu nehmen, wäre etwa, wie wenn man dem Astronomen das Fernrohr oder dem Boxer den Gebrauch der Fäuste untersagen wollte. Das Verbot der Existenzsätze und des Tertium non datur kommt ungefähr dem Verzicht auf die mathematische Wissenschaft überhaupt gleich. (Hilbert, 1930, 306)

[20] [Die] Mathematik ist eine voraussetzungslose Wissenschaft. Zu ihrer Begründung brauche ich weder den lieben Gott, wie Kronecker, noch die Annahme einer besonderen auf das Prinzip der vollständigen Induktion abgestimmten Fähigkeit unseres Verstandes, wie Poincaré, noch die Brouwersche Urintuition und endlich auch nicht, wie Russell und Whitehead, Axiome der Unendlichkeit, Reduzierbarkeit oder der Vollständigkeit, die ja wirkliche inhaltliche überdies gar nicht plausible Voraussetzungen sind. (Hilbert, 1930, 312)

It is clear that Gödel sees Hilbertian ideals as worth pursuing. He is unhappy with the uncritical acceptance of classical logic with its underlying Platonistic interpretation and calls for a constructive justification in the form of a consistency proof.

What remains of mathematics after eliminating the problematic assumptions of non-constructive existence proofs and impredicative definitions is "the so-called intuitionistic mathematics" (Gödel, 1933b, 50). Like Menger, Gödel argues that intuitionistic mathematics is not a monolith but rather a hierarchy of more or less constructive systems (ibid., 51):

> As we ascend the series of these layers, we are drawing nearer to ordinary non-constructive mathematics, and at the same time the methods of proof and construction which we admit are becoming less satisfactory and less convincing.

Gödel then gives three criteria that define the strictest form of constructive mathematics:

A1. All primitive relations should be decidable and all primitive functions calculable.

A2. The use of negation over universal statements should be restricted only to the case where it abbreviates a counterexample.

A3. Universal quantifiers should be applied only over totalities generated by a finite process.

Condition A2 is motivated on the grounds that negating universal statements would "give existence propositions" (Gödel, 1933b, 51). Because Gödel identifies negated universal propositions with existential propositions, A2 just states that an existence claim should be accompanied by a construction.

According to Gödel, the three conditions define the finitistic methods that Hilbert wanted to use to prove the consistency of mathematics. However, the systems that satisfy these conditions are too weak for this purpose. The inadequacy of the strictly constructive systems suggests that one needs to seek a solution from higher up the hierarchy of constructive mathematics. Although such a solution has not yet been found, Gödel remains hopeful that an ideally constructive system could be constructed (Gödel, 1933b, 53).

The Zilsel talk presents essentially the same conditions for constructivity as those listed in the Cambridge talk, here called simply the framework, *die Rahmendefinition*, for constructivity. It is formulated as follows (Gödel, 1938, 91):

B1. All primitive relations should be decidable and all primitive functions calculable.

B2. The use of existential quantifier should be restricted only to the case where it abbreviates an instance. No propositional connectives should be used over a universal quantifier.

B3. The objects of the theory should be surveyable (*überblickbar*).

B4. The rules and axioms that are to be allowed are all of propositional logic, recursive definitions, the rule of substitution, and complete induction.

Gödel's formulation of the criteria is well in line with Hilbert's position. The rejection of quantifiers is characteristic to Hilbert's finitism, whose base logic is classical. Decidability then follows from the lack of quantifiers and calculability from the definition of admissible (primitive recursive) functions in Hilbert's system.

The criterion of referring to totalities generated by finite processes (A3) is essentially the same as the criterion of surveyability (B3), which can be viewed as a finitary characterisation of "real" objects. What "finitely generated" means is that one has an effective process which can be seen to operate on concrete objects. Such processes have the "intuitive evidence" that Hilbert and Bernays required for the basic objects and procedures of a finitistic theory. A paragon finitistic object is a natural number, as the sequence of natural numbers starting from 1 and the continuation of the number sequence by iteration of the operation of adding one is entirely graspable. Numbers are *concrete* objects that can be represented in, e.g., stroke notation. Here $||||$ would represent the number 4; the next natural number, 5, is wholly graspable as the addition of one more stroke $|$ to the previous string, producing the figure $|||||$. Therefore the concept of a natural number is graspable and the set of natural numbers is finitely generated.

Not all processes are of the good kind. Gödel states that the surveyability condition is problematic "because of the concept of function" (Gödel, 1938, 91). From the Hilbertian viewpoint, a function, in general, cannot be considered a finitary object because it does not represent a concrete object but rather a procedure. Some operations are finitary – e.g., addition, multiplication, and primitive recursion in general – although these are not considered to be basic objects but basic *rules* (which are included in Gödel's list in criterion B3).[21] Moreover, all of these are essentially operations on numbers, which are finitary objects, and thus, e.g., functions over functions could not be acceptable, which would become a problem for Gödel's own functional interpretation.

Intuitionism, which extends "considerably" beyond finitism (Gödel, 1938, 93), does not satisfy all of the criteria. Strictly speaking, intuitionistic arithmetic satisfies only the first criterion A1 or B1.[22] The first half of the criterion B2 states that existential quantifiers should be introduced by definition

$$\exists x A(x) := A(a) \text{ for some term } a$$

Later on, Gödel will characterise a constructive system as one that possesses the existence property, i.e., where every existential theorem $\exists x A$ is accompanied by an instance $A[x/t]$, and a system that satisfies B2 makes the existence property transparent. However, he does not talk about this issue when criticising intuitionism, but says merely that B2 is not satisfied because logical operations are applied to all statements (ibid., 103). The main problems are, according to him, application of propositional connectives and in particular negation over universal quantifiers (criteria A2 and B2) and introduction of objects that are not surveyable (criteria A3 and B3), namely, proofs.

It should be mentioned that another common objection against the constructivity of the proof interpretation is that the clause for implication is impredicative. The proof explanation of an implication $A \supset B$ reads "any proof of A can be transformed into a proof of B," but nothing rules out the fact that the proof of A might contain $A \supset B$. However, Gödel did not criticise the proof interpretation on this account until much later, when he was working on a revised edition of the 1958 *Dialectica* paper (Van Atten, 2018, 10).

[21] Zach notes that "most of Hilbert's remarks deal with objects of finitism, and leave the finitistically admissible forms of definition aside" (Zach, 2003, 227). This is the why it is difficult to make absolute claims about the formal system Hilbert believed to correspond to the finitistic standpoint, even if it is nowadays widely agreed that finitary methods are contained in Primitive Recursive Arithmetic. It is often said that finitism excludes abstract objects, but there are some processes which are not concrete in the same way that numbers are concrete, but that are graspable when applied on concrete objects.

[22] Condition B1 is, at least, satisfied for the primitive functions. However, as in the 1938 talk, Gödel examines the modal interpretation of HA (see **section 1.2.1**), it is not certain if the condition on basic relations is satisfied because the **B**-predicate is not necessarily decidable. Nevertheless, this criticism does not apply directly to HA itself.

2.2.2 Interlude: an earlier draft of the "Present situation"

Gödel's sympathies towards Hilbert come through even more clearly in the draft version of the 1933 lecture that can be found in the *Papers* (040113). It is written in longhand English and is, for the most part, identical to the version that was published in the *Collected Works*. However, there are some differences which, I think, elucidate Gödel's early thought better than the posthumously published paper.

The conditions of constructivity are more or less the same as A1–A3, although the restriction on negated universal quantifiers (A2) is stated as a consequence of the general rule that existence statements should be witnessed (p. 18), showing clearly that Gödel, like Hilbert and Weyl, did not see any difference between the two. Gödel also expresses some doubt as to whether the incompleteness theorem holds for Hilbert's finitism, although he says that it is very likely that it does (p. 19–20):

> The reason why this statement [that finitistic methods are encompassed in the system defined by A1–A3] cannot be made with absolute certainty is this: all functions of integers which can be actually called calculated for each particular integer are allowable in the system A but it is impossible, to describe all the procedures for the construction of such functions and therefore it is also impossible to give a (rigorous proof) that all of them are expressible in classical arithmetic although it can be made very plausible that nobody will ever be able to construct any such function for which this would not be the case.

The reasoning appears similar to the discussion in Vienna in January 1931, where Gödel suggests that the totality of finitary methods would be formalisable in a single system because one can maybe diagonalise the collection of constructive proofs.

As in the published paper, Gödel states that the intuitionistic notion of absurdity is doubtful because one can reinterpret classical mathematics using this notion. In fact, he has originally written that the negative translation is a *constructive* interpretation of classical mathematics but later cancelled the word "constructive," perhaps because he was aware of the deviant character of the interpretation. On the contrary, he does not say that much about intuitionistic proofs, only noting that criterion A3 is violated because the totality of all proofs is not finitely generated. Nothing is said about vagueness of such totalities or absoluteness of intuitionistic proofs. This suggests that, unlike often assumed, the main motivation for Gödel's critique was not the proof interpretation but rather his (classical) understanding of intuionistic quantifiers in terms of the negative translation.

However, Gödel comes off as less critical towards intuitionism, concluding that (p. 21)

> [therefore] you may be doubtful as to the correctness of the notion of absurdity and as to the value of a proof for freedom of contradiction by means of this notion. But nevertheless it may be granted that this formulation is at least more satisfactory than the ordinary platonistic interpretation [. . .]

The above quote leaves no doubt about the fact that Gödel did see Platonism as problematic. The milder statement that classical mathematics presupposes "a kind of Platonism" which is unsatisfying is here stated in a much stronger form. Gödel states that as soon as meaning is given to classical mathematics, "we become entangled hopelessly kind of Platonism or which obviously doesn't give any guarantee against contradiction" (p. 15).[23] This should be enough justification for calling Gödel's early views formalistic in nature.

2.2.3 Non-surveyability of intuitionistic proofs

The criticism of the intuitionistic conception of provability is expressed in several different ways in the two lectures. In the "Present situation" Gödel states that intuitionism breaks condition A3 because intuitionistic proofs cannot be generated by a finite process. He argues that "[totalities] whose elements cannot be generated by a well-defined procedure are in some sense vague and indefinite as to their borders" (Gödel, 1933b, 53). In the Zilsel lecture, where Gödel introduces the proof interpretation in his modal framework, he says that the proof interpretation involves provability "in the absolute sense" (Gödel, 1938, 101). In fact, the axioms of the modal interpretation would be false for any formalisable provability predicate: if the operator **B** denoted prov-

[23] In his drafts, Gödel left empty spaces for references to others' work or when he could not find a correct English word.

ability within a system, the incompleteness theorem would be violated by the theorem $\mathbf{B}(\mathbf{B}\bot \supset \bot)$ (ibid., footnote w).

Now, the intuitionistic notion of absurdity, which is supposed to account for the difference between intuitionistic logic and classical logic, refers to to the totality of all proofs. This is because $\neg A$ is defined through implication as $A \supset \bot$, read as "for *any* proof of A, one can derive a contradiction." Thus, Gödel concludes, "this foundation of classical arithmetic by means of the notion of absurdity is of doubtful value" (Gödel, 1933b, 53).

In which sense are intuitionistic proofs problematic? I will discuss Gödel's critique of intuitionistic proofs in detail in **Chapter 3**, but for now, we should distinguish between two types of problems: those related to the surveyability condition and those related to non-formalisability of the totality of intuitionistic proofs (provability by "any means imaginable," as Gödel put it in his dissertation). In the second case, intuitionistic provability is a vague concept because one cannot give formal criteria for what it is to be an intuitionistic proof. The primary reason to make this distinction is that Gödel's 1933 and 1938 conditions of constructivity contain the surveyability condition, but the 1941 conditions do not, even though Gödel still criticises the proof interpretation on account of vagueness or absoluteness. Here, I will focus on the first issue.

As mentioned above, a proof in the sense of a procedure is not a finitary object by default, although in order to do meaningful mathematics, one must accept some procedures as finitary. The status of natural numbers, which were characterised as textbook finitary objects, as finitary requires us to understand some procedures as finitary for the following reason. To grasp an object such as a number is to "see it" in its entirety. In this sense, we can see the number four as a series of that many strokes: ||||. We understand what "the successor of 4" means because the operation of adding one is itself fully graspable. Practically, a number such as $10^{(10^{100})}$ cannot be directly visualised, but we can understand its *construction* because we understand the operation of exponentiation m^n on natural numbers (Bernays, 1930, 347). The understanding of exponentiation, on the other hand, is based on our understanding multiplication $m \times n$ on natural numbers, which is simply writing the string representing m n times. Therefore one needs to allow for intuition of some procedures to justify an *arbitrary* natural number as finitary.

Which procedures are surveyable cannot be given a definite answer, at least from the viewpoint of historical finitism. The first reason is that, as mentioned, Hilbert and Bernays never gave a definitive criteria for finitary methods. Moreover, the theory and the practice of finitism were sometimes two different things (see Zach, 2003). In his 1936 proof of consistency for Peano Arithmetic, Gentzen provides a rather long explanation of the way in which his construction of transfinite ordinals should be seen as finitary. Gentzen constructs his transfinite ordinals as arithmetically defined well-orderings. He admits that the weak point in the proof is the sense in which such well-orderings can be called finitarily constructed: whereas an ordering of size ω or even $\omega \times \omega$ can be grasped as a sequence of numbers and a sequence of ω-sequences, respectively, an ordering of type ω^α is certainly not graspable in the same way for a large α (Gentzen, 1936, 559).

Gentzen argues that what one needs to grasp is but the transition from α to ω^α. Certainly, ω is graspable; we can also grasp the transition from ω to ω^n because this is just a generalisation of $\omega \times \omega$, which we can understand if we understand what a sequence of type ω is and the replacement of every number in the sequence by a sequence of type ω. From there we can reach a sequence of type ω^ω.[24] Generalising to higher ordinals, there is nothing fundamentally new in the process that one uses to transform a sequence α to α^n, even if the ordering α might not in itself be graspable *as a well-ordering*.

However, if what is grasped in grasping a well-order is the way in which it was constructed, one is not operating on numbers of any sort but rather procedures. Those procedures that are operated upon could then also be procedures on procedures up to any level of complexity. This idea is much more complicated than Bernays' example of going from natural numbers m and n to a natural number m^n. Gentzen's idea of a procedure, then, seems to transcend what Hilbert and Bernays originally meant by a finitary operation.

This idea of surveyability of procedures is not entirely congenial to Gödel either. He mentioned already in 1938 that the surveyability condition is problematic because of the notion of a function. There seem to be, at least from Hilbert and Bernays' point of view, no problems with ordinary primitive recursive functions. However, Gödel's functional interpretation, which he mentions in the Zilsel lecture, uses the idea of a function, which is a nested

[24] In fact, not quite; but Gentzen would argue that as all infinities must be seen as potential, and that therefore ω^n for an arbitrarily large n is as close to the limit ω^ω as it can be (see Hämeen-Anttila, 2019, 115–120).

procedure like Gentzen's construction of ordinal numbers.[25]

Surveyability of a procedure, then, depends on the complexity of the procedure and the objects on which it operates. A formal proof, in general, is not a finitary object because proofs are one kind of a procedure. In addition, proofs may operate on other proofs, as is the case for, e.g., the implication introduction rule, which assumes that the proof of the antecedent of the implication is given. However, intuitionistic proofs are not just beyond finitary intuition; rather, Gödel also indicates that the notion is not well-defined because it cannot be formalised. In this sense, it is, as Gödel often says, vague.

This conception of an intuitionistic proof is the same that Gödel characterised in his dissertation as provability by any means imaginable. Whereas the set of all proofs of Heyting Arithmetic is recursive, the set of all intuitionistic proofs whatsoever is not, even if for every single proof, there existed some way to figure out if it is a proof or not. It has already been made clear that the proof interpretation cannot refer to HA-proofs only, for it would contradict the incompleteness theorem. Moreover, there is no principled reason to assume, at least from the Brouwerian viewpoint, that every single intuitionistic proof can be given in *a* formalism. One can say, however, that there are still criteria governing constructions that are proofs in the non-formal sense (see Sundholm and Van Atten, 2008), and therefore "provability by any means imaginable" is perhaps not an entirely fair characterisation of intuitionistic proofs. Unlike in the later lectures, Gödel does not even consider whether the intuitionistic conception of proof could have *some* criteria, even if informal. This is because, as we have seen, he was not yet open to the idea of informal provability as a fruitful foundational concept.

This criticism of the notion of proof from the viewpoints of vagueness and unsurveyability is, compared to the logical works, a new aspect. In the early works, Gödel's argued that intuitionistic logic has no real advantage over classical logic because intuitionistic concepts can also be interpreted classically. Nevertheless, the problematic character of intuitionistic proofs is often seen as Gödel's main argument. Whereas it has more philosophical weight, in the lectures of the 1930s, the argument is based on the criterion of surveyability, which no longer appears as a requirement for the constructivity in 1941. In order to understand better the roots of Gödel's critique one must also, I believe,

[25] In the Zilsel lecture, Gödel remarks that Gentzen's proof replaces the "vast notion of proof" by "the equally vast notion of functional" (Gödel, 1938, 109). This suggests that Gödel was already aware of this problem.

take a look at the other aspect of Gödel's critique, namely that of intuitionistic quantifiers, against the background of his logical works on intuitionism.

2.2.4 "Man hat ein Modell"

The previous sections showed that Gödel's early conception of intuitionistic logic is defined by its relationship to classical logic. He often calls intuitionistic logic a "reinterpretation" of classical logic with little to offer as a system of its own. This is in stark contrast to, e.g., Bernays and Gentzen, who, despite standing nominally on Hilbert's side, worked with and appreciated intuitionistic logic for its own sake.

Gödel's essentially classical viewpoint leads him to dismiss intuitionistic logic as non-ideal, at least in its treatment of arithmetic. Although there is no criterion of constructivity that is directly violated by the existence of a negative translation, Gödel uses the translation to explain why the quantifier criteria A2 and B2 apply to intuitionistic logic even if the intuitionistic meanings of the quantifiers differ from those of classical logic. This is why the negative translation is mentioned alongside the critique of intuitionistic quantifiers in all of Gödel's lectures on constructive foundations and intuitionism.

In 1933, Gödel says that intuitionism fails criterion A2 because the notion of absurdity. Brouwer's absurdity, he admits, is supposed to be different from classical negation, but still "it might happen [...] that you can derive a contradiction from the proposition "for every x, $F(x)$ is true" by intuitionistic methods without anyone's being able to give a counterexample [...] so we have a perfect substitute for non-constructive existence theorems" (Gödel, 1933b, 52). This is to say, we have an intuitionistic concept, involving the seemingly special notion of absurdity, that seems to correspond to the classical non-constructive concept of existence.

Indeed, there is a specific way in which intuitionistic $\neg\forall\neg$ corresponds to the classical $\exists$, namely, via the negative translation: "If we investigate the axioms of intuitionistic mathematics as stated by Heyting, Brouwer's disciple, we find that for the notion of absurdity exactly the same propositions hold as do for the negation in ordinary mathematics – at least, this is true within the domain of arithmetic" (Gödel, 1933b, 53). A similar remark is made in the Zilsel lecture, where Gödel says that the "apparently weaker assumptions," i.e., the elimination of PEM, could be replaced by the classical interpretation via the negative translation (Gödel, 1938, 97). Gödel concludes that the only aspect in which intuitionistic arithmetic differs from its classical counterpart is that "the substrate on which the constructions are carried out are proofs instead of numbers or other enumerable sets of mathematical objects" (Gödel, 1933b, 53). Here we run into problems because of the notion of proof, which makes intuitionistic arithmetic, in a sense, even worse than its classical counterpart.

Gödel still placed plenty of importance on the negative translation even in the 1941 lectures. Although intuitionistic logic is no longer said to have non-constructive aspects, Gödel continued to claim that "nothing at all is lost by dropping the law of the excluded middle, but only the interpretation of the theorems has to be changed" (Gödel, 1941, 190), which is almost the exact wording of the Hahn notes from 1931. The more fundamental difference is that intuitionists reject impredicative definitions, and thus intuitionism and classical mathematics differ fundamentally only in the realm of analysis and set theory. This does not seem to be entirely in line with Gödel's idea that the negative translation shows that intuitionistic and classical arithmetic have no essential differences, as the translation can be extended to impredicative systems as well. Moreover, as Troelstra (1986, 285) notes, there are impredicative intuitionistic theories that are frequently considered admissible: e.g., Myhill's (1973) intuitionistic version of ZF has an unrestricted separation axiom. He does remark, however, that Gödel's view of intuitionism seems to have simply excluded impredicative elements.[26] In the Princeton Lectures, which I will discuss in more depth in **Chapter 3**, Gödel is slightly more careful in his wording, saying that this shows that "the restrictions which Brouwer puts on classical methods of proof do not go beyond those of half-intuitionists *as far as the formalism is concerned*" (*PLI*, p. 39.2).

Gödel's interpretation of the negative translation theorem is by no means the only possibility. One can say that whereas relative consistency is provable, the translation obviously fails to preserve meaning. From the intuitionistic point of view, consistency was never enough for correctness, and thus

[26] However, Gödel seems to have thought in 1941 that intuitionistic theories can have impredicative elements. In *Max Phil* 4 (p. 155) he writes that there are impredicative elements in intuitionistic set theory:

> Bem. (*Grundl.*): Impräd⟦ikative⟧ Elemente bei *Brouwer*:
>
> 1. Summenspecies einer beliebigen Species
> 2. *Df.* der Ordinalzahlen (im Wesentlichen als Durchschnitt aller gegen die 2 Erzeugungsoperationen abgeschlossenen Species)

what has been proven is but a formal result, which in no way implies that intuitionistic arithmetic is not constructive. The problem with the negative interpretation of a classical statement $\exists x A$ is that it is, from the intuitionistic viewpoint, ambiguous: it cannot distinguish between cases where A can be instantiated and when the nonexistence of an instance is inconsistent. The translation $\neg\forall x\neg A$ of $\exists x A$ retains meaning only from the classical point of view, in which the equivalence $\neg\forall x\neg A \equiv \exists x A$ is valid. One cannot, however, read the translation intuitionistically in a way that is meaning-preserving without losing some distinctions along the way. Then intuitionistic negated universals are certainly innocent: they do not say anything more than they promise to, even if they express one sense of the classical existential quantifier.

Kreisel (1987, 81–82) notes that Gödel's calling the difference between intuitionistic and classical quantifiers "only apparent" is "like saying that the notions of countable and uncountable structures differ mildly because the same first order formulae are valid classically for both classes of structures. If [...] one wants to dismiss [intuitionistic logic], one needs to find a less hackneyed (metamathematical) property than conservation of classical logic over [intuitionistic logic] for the negative fragment." Elsewhere, Kreisel points out that Gentzen's interpretation[27] of the negative translation is very different: he emphasises the purely formal nature of the result and "seems impressed by what one may call the richness of intuitionistic logic" (Kreisel, 1971, 257).

What is curious is that Gödel seems to acknowledge the purely formal character of the translation, although he also states that it does not matter. In the Zilsel lecture he says that PEM is irrelevant, for even though it does not hold for Heyting's system, its negative translation holds (Gödel, 1938, 97). The "somewhat deviant" character of the interpretation, as Gödel puts it in 1932, makes no difference because "one has a model," *man hat ein Modell* (ibid.). But this argument is indeed akin to concluding that countable and uncountable structures are exactly the same because there are uncountable models of first-order Peano axioms. Another example is that the second incompleteness theorem entails the existence of a model of, e.g., Peano Arithmetic where an inconsistency is "provable." The "proof," however, cannot be a proof in any normal sense, as it would have to be infinite. It should not make us doubt that arithmetic is inconsistent; it simply means that the Peano axioms leave space

[27] Gerhard Gentzen arrived at the result independently in 1932, but withdraw his article for *Mathematische Annalen* when he found out about Gödel's work. Gentzen's paper was published posthumously in Szabo (1969).

for some ambiguity, which is realised in the existence of non-standard models.

We need not doubt the consistency of PA because we have in mind what should be the *intended* interpretation of PA: our objects at hand are the natural numbers, and *only* the natural numbers (as they are commonly understood), on which the arithmetic operations are carried out. It is a fact about first-order logic that it cannot express axioms that would pin down the intended interpretation. Of course, this had led some to think that we should adopt a second-order framework which is categorical, i.e., whose models are all isomorphic and match the intended interpretation. However, here we are dealing with questions about the formalism, not about the meaning we want to give to the Peano axioms.

How we should talk about intended interpretations is in itself an interesting question. The formalism of second-order arithmetic matches more closely its semantics, although it no longer has a complete proof system. One then has to decide between accepting just an informal understanding of the intended interpretation of the Peano axioms and a stronger semantics expressing precisely the intended interpretation. Will an informal description suffice to make clear what one means by the axioms?

In the spring of 1941, Gödel would become more sensitive to these kinds of questions. In the Princeton and Yale lectures, Gödel considers the proof interpretation as the intended interpretation, and his critique arises from the fact that this intended interpretation is not made precise to a satisfactory extent. His functional interpretation is then presented, not only as an *alternative* to Heyting Arithmetic, but also as an intended interpretation in more specific terms. However, this is not how he saw the issue in 1938.

2.2.5 The functional interpretation in the Zilsel lecture

In the Zilsel lecture, Gödel gives three possible ways of approaching the consistency question: a system of functionals of higher types, the "modal-logical route" by which he means intuitionistic arithmetic with the proof interpretation, and Gentzen's consistency proof for Peano Arithmetic by transfinite induction. Of these, only the first approach satisfies all of the constructivity conditions. The basic idea is to extend Hilbert's finitary arithmetic by primitive recursive functionals, although not much is said about the system beyond this.

Unlike in 1941, the functional system is not presented as a reinterpretation of Heyting Arithmetic, but rather, it seems, a system in which a consistency proof for arithmetic could be directly carried out. This was, after all, the explicit aim that Gödel declares at the beginning of both 1933 and 1938 lectures. He considered his system as an alternative approach to intuitionism as well as Gentzen's strategy of extending Primitive Recursive Arithmetic with a restricted principle of transfinite induction. However, Gödel saw extension by functionals as superior to Gentzen's approach.

After a concise description of what functionals are, we find a list of properties of the functional system (Gödel, 1938, 97):

> 4. ⟦The following⟧ are contained in this procedure:
>
>   1. Addition of recursion on several variables,
>   2. Addition of the statement *Wid*,
>   3. Addition of Hilbert's rule of inference.

Here 3 refers to the ω-rule.[28] *Wid* refers to a statement expressing consistency, i.e., the statement that there is no proof of a contradiction.

As for negative results, Gödel states that it is not possible to prove the consistency of arithmetic with only finite types, and therefore extension into the transfinite is necessary. He also conjectures that one cannot prove the consistency of analysis in any functional system. Gödel calls the construction of functional systems "an interesting open problem," because only these systems satisfy the strict criteria for constructivity. The section ends with a remark: "To show for the individual requirements" (Gödel, 1938, 91).

We do not know whether Gödel said something more about how a functional system can satisfy the criteria. As mentioned, the compatibility of functionals with the surveyability criterion is at the very least problematic, and Gödel seems to have acknowledged this. Nevertheless, he claims that the system will satisfy the criteria B1–B4. This is not consistent with his statement in 1933 that one cannot arrive at a consistency proof for arithmetic in a system that satisfies criteria A1–A3. The only difference is in the formulation of criterion of surveyability, which is stated in A3 as the requirement that all totalities which are referred to are created by a "finitary process," but this does not seem any broader than the criterion B3, which says that the objects should

[28] The ω-rule is an infinitary rule of inference, where $\forall x A(x)$ can be derived with all its instances $A(0), A(1), A(2), \ldots A(n), \ldots$ as premises.

be "surveyable (that is, denumerable)." Letting go of the criterion of surveyability, however, would mean a departure from the finitistic viewpoint, which is motivated by the very criterion.

Gödel did not start working on the functional system until some years later. He was mostly occupied with set theory between 1937 and 1939 when he gave consistency proofs for both Axiom of Choice and Continuum Hypothesis (see Dawson, 1997, chapter VI). In 1939 and 1940, he gave two lecture courses on his results at Notre Dame and the Institute for Advanced Study. When exactly he decided to pick up the functional interpretation again is not certain, but this seems to be relatively late: in the Bulletin of the IAS of April 1940, we are told only that in the academic year 1940–1941 Gödel will "probably give some lectures related to the foundations of mathematics" (p. 6). The first sketches of the functional interpretation that our team has been able to find are dated 1st January, 1941, and most of his notes on intuitionistic logic are from 1941, around the time that Gödel was lecturing on the topic in Princeton. Between 1938 and 1941, his view had changed, and the functional interpretation is now presented as an interpretation of Heyting Arithmetic and a proof of constructivity of intuitionistic logic.

Chapter 3

Changing perspectives: 1941 lectures in Princeton and Yale

In the spring of 1941, Gödel gave a lecture course on intuitionistic logic at the Princeton Institute for Advanced Study. In the Princeton Lectures, which probably took place between February and May, Gödel gave a thorough introduction to intuitionistic logic, focused on the interrelations between intuitionistic and classical logic, followed by the presentation of his functional system as a reinterpretation of Heyting Arithmetic. He also introduced the functional interpretation, then called system Σ or $\overline{\Sigma}$, in a lecture titled "In what sense is intuitionistic logic constructive?" delivered in Yale on 15th April 1941.

In contrast to the Zilsel lecture of 1938, where a system based on finite-typed functionals is introduced as an alternative way of carrying out the finitary consistency proof for arithmetic, Gödel's Princeton Lectures and the Yale talk present it foremost as an interpretation of intuitionistic logic. Of course, one still gets a relative consistency proof for classical arithmetic by the negative translation and the functional interpretation, but this is, especially in *PLI*, mentioned on the side. This is not to say that Gödel was not interested in other aspects of the interpretation. He did think about proving the consistency of analysis in a similar way, and he believed that intuitionistic methods could help him in another ambitious goal, a proof of independence for Continuum Hypothesis (see **section 4.3.1**). However, the formalistic framework of the 1930s lectures is no longer present in his 1940s works.

Although the Yale lecture gives a relatively complete presentation of the functional system, the Princeton course helps to give us a better idea of Gödel's views on intuitionism and the nature of his criticism as well as the motives behind the functional interpretation. Gödel still focused on classical interpretations of intuitionistic logic and placed plenty of importance on his negative translation. On the other hand, he was beginning to reconsider some of his earlier tenets, and in general, one can see that he came to see finer distinctions in the field of constructive foundations. This may have been partly due to the challenges Gödel faced when developing Σ: it is not entirely clear whether it can serve as an interpretation or as a proof of constructivity of Heyting Arithmetic. In particular, he was concerned about whether Σ complies with the strict criteria of constructive systems as formulated in his previous lectures. Here the key difficulty lies in the computability of finite-typed functionals.

Apart from the Princeton lecture notes, there is a large amount of material related to the functional interpretation in Gödel's *Nachlass*. The early sketches for *PLI* are contained in the series *Arbeitshefte*, mainly in *Hefte* 7 (030025) and 9 (030027). As mentioned in the previous chapter, the earliest notes can be found from *Heft* 7 ("backwards direction") and they are dated 1st January 1941. The style of the Princeton Lectures shows that Gödel was still working on the functional system when he had already started lecturing. In the notes, we find several errors that Gödel has not corrected – his more finished drafts usually contain many cancellations and additions – until some days later, when he would come back to the error and correct it in a later lecture. The most notable errors concern the existence property for HA and the Σ-interpretation of disjunction, the latter of which made it also to the Yale lecture (see **section 3.2.2**). On 16th March, Gödel wrote to his brother Rudolf that he is very busy preparing the lecture notes, which confirms that he was not quite ready with his functional interpretation when he started the Princeton course (Van Atten, 2015, 200).

In the broader context of the Princeton Lectures and the notes related to the functional interpretation, we get the impression that Gödel was not, in the end, perfectly happy with the lectures. In another letter to Rudolf Gödel, dated 4th May, 1941 (Van Atten, 2015, 201), he wrote that there were only three students left at the end of the course. Part of the reason might have been the wartime circumstances, but a part of it could well have been that Gödel's lecture course got quite formal and difficult to follow towards the end. Gödel later recalled to Hao Wang (1996, 86):

> I obtained my interpretation of intuitionistic arithmetic and lectured on it at Princeton and Yale in 1942 or so [should be 1941].

> [Emil] Artin was present at the Yale lecture. Nobody was interested.

In a draft of an undated letter to a graduate student Frederick Sawyer from perhaps around 1974 or later, Gödel wrote that he chose not to publish the Yale and Princeton lectures on intuitionism because his interest had shifted to other problems; also, he says, "there was not too much interest in Hilbert's program at the time" (Gödel, 2003b, 210–211).

Gödel's dissatisfaction was a symptom of the many challenges he encountered at the time of working with the functional interpretation, mathematical and philosophical. By 1941 it was clear to Gödel that the functional system was not strictly constructive in the sense of the criteria given in the 1933 and 1938 lectures. Indeed, the finitistic criterion of surveyability of objects is no longer present in the Princeton and Yale lectures. Nevertheless, he was still unhappy with the intuitionistic conception of proof, which he thought of as imprecise, and sought to give a more precise interpretation of intuitionistic arithmetic in terms of functionals. Despite having given up surveyability, Gödel faced another difficulty trying to prove that the functionals also satisfy the criterion of computability. He never succeeded in giving a satisfactory proof, and later on, he simply chose to assume computability as primitive.

What we see in Gödel's works and notes of 1941 is a gradual resignation of the formalistic viewpoint that characterised the early lectures. This change made Gödel more sympathetic towards interpretations of intuitionism on its own terms. Even though he never engaged in a deeper study of these interpretations, he did have some ideas as to where that study would lead. Moreover, his mathematical ambitions related to the consistency of analysis and independence of the CH led him to study Brouwer's mathematics more closely.

3.1 The sense in which intuitionistic logic is constructive

Gödel's 1941 lectures place the question of constructivity of intuitionistic logic at the centre of the discussion. The functional system is presented both as a reinterpretation and a proof of constructivity for Heyting Arithmetic: it gives, to answer the question posed in the title of the Yale lecture, "the sense in which intuitionistic logic is constructive." By this more precise interpretation, Gödel claims, one can show that intuitionistic logic really has the characteristics of a constructive system, most importantly, the existence property.

The functional interpretation can also be considered an interpretation of classical arithmetic via the negative translation. As such, it also serves as a relative proof of consistency of classical arithmetic. Although Gödel mentions this result in both Princeton and Yale, it does not, in contrast to the Zilsel lecture, provide a framework for the discussion. Moreover, Gödel no longer explicitly claims that the consistency proof is more constructive than Gentzen's variant, and he does not discuss the Hilbert Programme in this context. I will not, therefore, discuss the interpretation of Gödel's functional system in the context of the Extended Hilbert Programme, even if Gödel himself places it in this context in the *Dialectica* paper of 1958.

Gödel's critique of intuitionism has evolved from the 1930s lectures. To begin with, there is no criterion to match the surveyability condition of 1933 and 1938, which is a clear departure from the Hilbertian standpoint. This means that Gödel has to present his criticism from a different viewpoint than the non-surveyability of intuitionistic proofs. What he puts forward is not an argument for intuitionistic logic being non-constructive, but rather a rhetorical argument to question its constructivity, focused on the negative translation between classical and intuitionistic arithmetic.

The close connection between the two systems is characterised as "suspicious," casting doubt on whether intuitionistic logic is genuinely constructive. To answer the argument from the negative translation involved talking about the intended interpretation, and here the intuitionist is, according to Gödel, in trouble: the proof interpretation is supposed to be the intended interpretation, yet it cannot be specified precisely enough to convince us. Therefore intuitionism needs to be reinterpreted in a way that is both natural, in the sense of being non-deviant, and precise. The functional system, then, is introduced as the desired interpretation.

3.1.1 Criteria without surveyability

In the 1933 and 1938 lectures, Gödel stated that the intuitionistic proof interpretation does not satisfy the strict criteria of constructivity because intuitionistic proofs are not surveyable. It was pointed out that from the formalistic point of view, proofs and procedures, in general, are not finitary objects in the same way that a natural number is a finitary object. The key problem with

Gentzen's consistency proof of arithmetic was to explain how the procedure of transfinite induction, "running through" well-orderings of increasing length, can be intuitively grasped. Many believed that the principle goes well beyond original finitism, and indeed, it is difficult to see how transfinite induction could be compared to the much simpler inference rule of induction on natural numbers. However, the same objection applies to the extension of the finitary system by introduction of functionals of higher types. As mentioned, Gödel's comments in 1938 suggest that he was aware of this problem, and accordingly, we find no criterion of surveyability in the 1941 lectures.

The criteria for the Yale and the Princeton lectures differ only in wording, so it is convenient to present them in one set of conditions:

1. All primitive relations should be decidable and all primitive functions calculable.

2. Existential statements should occur only as abbreviations of their instances.

3. Negated universal statements should occur only as abbreviations of the corresponding counterexamples.

Criterion 3 is expressed somewhat more generally in the *PLI*, where it says that propositional connectives, in particular $\supset$ and $\neg$, should only be applied to decidable statements, and thus never over quantifiers, as "quantifiers destroy decidability" (p. 45.1).

Gödel states that intuitionistic arithmetic does not satisfy conditions 2 and 3 because "the existential quantifier there is not introduced by definition but appears as a primitive symbol and the propositional connectives are applied without restriction" (Gödel, 1941, 192). Again, he is mostly focused on condition 3, which appears counterintuitive from the intuitionistic standpoint. Without the quantifier dualities, one cannot get existential statements from negated universals, which is the real motivation behind the criterion. Obviously, negated universal theorems do not imply the provability of their instances, but this violates the existence property only if we assume $\neg\forall x A \supset \exists x \neg A$. However, here Gödel leans on the negative translation once more, stating that because of the close connection between classical and intuitionistic logic, the question of constructivity of intuitionistic logic is by no means trivial.

3.1.2 The argument from negative translation

Gödel's argument from the negative translation serves as a motivation to take a closer look into intuitionistic logic. Among the notes for the Yale lecture (040263), we find a two-page list written in shorthand and titled "*Vortrag Yale Ist die int. Logik konstruktiv*." It starts with

1. Problematic Ax⟦iomatic⟧[1] relationship to classical logic. Intuitionistic objection against classical logic [intuitionism and half-intuitionism]. But do the intuitionists themselves not fall into the same mistakes?

   A.) Existential statements are still negated

   B.) Result on equivalence

   C.) Questionable if every existential statement is constructible[2]

It is not clear why Gödel has written "existential" in A.) instead of "universal," but this might be simply a mistake on his part. Item B.), the "result of equivalence," refers to the negative translation. A.) – C.) together constitute an argument for doubting the constructivity of Heyting Arithmetic, although they are not meant to prove that HA is not constructive. Instead, Gödel's goal is to interpret HA in a manner that is constructive in a more precise way than the ordinary proof interpretation.

As for item C.), it should be said that in *PLI*, Gödel explicitly states (p. 3) that a system is called constructive if it has the existence property, i.e., the property that every theorem of the form $\exists x A$ is witnessed by an instance $A[x/a]$.[3] Although in the Yale lecture, the criteria **1–3** are said to define a

[1] "Ax." here could as well indicate "Axiom(s)" or "Axiomatisation." However, it seems clear that this passage refers to the negative translation, which Gödel brings up in the beginning of the Yale lecture.

[2]

1. Problemat⟦isches⟧ Ax. Verhältnis zur klassischen ⟦Logik⟧. Int⟦uitionistische⟧ Einwände gegen die klassische Logik [Int⟦uitionismus⟧ und Halbint⟦uitionismus⟧]. Aber verfallen die Int⟦uitionisten⟧ nicht in denselben Fehler?

   A.) Existenzsätze doch negiert

   B.) Resultat über Äquivalenz

   C.) Fraglich, ob jeder Existenzsatz konstruierbar

[3] Not all intuitionistic theories possess the existence property. Intuitionistic projective geometry, to give an example, does not have it (von Plato, 2017a, 172–173), and there are intuitionistic set theories that, too, lack the property (see Crosilla, 2019, section 5.3.2). However, in *PLI*, Gödel expressed his doubt about the constructivity of intuitionistic geometry as well as set theory for other reasons (see **section 4.1.2**).

"strictly constructive or finitistic" system (Gödel, 1941, 191), in *PLI* (p. 45.1) it is said only that a system satisfying the criteria is "constructive in a more *precise* and stronger sense" (my italics). Namely, it is fully transparent that a system which has existence quantifiers defined by

$$\exists x A(x) := A(a) \text{ for some term } a$$

and contains no negated universal quantifiers has the existence property. Criteria **2** and **3**, therefore, serve to make constructivity wholly transparent.

Now, a system which does not restrict quantifiers in this manner might still have the existence property. In the case of intuitionistic arithmetic, this seems very plausible, given the intuitionistic interpretation of the connectives. However, one cannot still see at once that the existence property is satisfied, and, according to Gödel, there are reasons to suspect that it is not.

Gödel's earliest logical works focused on the relationship between classical and intuitionistic systems. It seems that in 1941, he still views intuitionistic logic through classical logic. In the Yale lecture, Gödel calls intuitionistic logic "rather a renaming and reinterpretation than a radical change of classical logic" (Gödel, 1941, 190). This is seen, according to Gödel, in the fact that although intuitionists reject nonconstructive existential statements and thus negations of universal statements "as meaningless," they still apply *absurdity* to universally quantified sentences, and "the axioms which intuitionists consider as evident about this predicate [of absurdity] lead, with suitable definitions of the other terms, to exactly the same calculus as classical negation, provided the other logical notions are suitably defined" (ibid.).

By "suitable definitions," Gödel refers to the interpretation given by his negative translation. He quickly puts aside the problem of deviancy discussed in the previous chapter, and states that the prohibition of non-constructive existence proofs becomes meaningless if one can simply use the translation to obtain $\neg\forall x \neg A$ from $\exists x A$. This is precisely the argument Gödel gives in the 1933 and 1938 lectures. However, here he does not present this as a proof that intuitionistic logic is less than ideally constructive, but rather as a source of suspicion. According to him, such a tight relationship between intuitionistic and classical arithmetic should make us doubtful about the true nature of Heyting's logic.

In the Princeton Lectures, Gödel devotes a fair amount of time to the topic of interconnections between intuitionistic and classical logic. There are two new theorems that he presents, one for propositional logic, which merits to be called the "truth table theorem" (see **section 4.1.1**), and one for arithmetic. The latter is a new version of the negative translation. Gödel first presents the original translation which, according to him, shows that we can deduce "[in a] sense the whole classical logic" (p. 38). He then says that we can come up with a new translation based on the concept of a "strong negation" which is more constructive than absurdity (p. 39).

What Gödel calls the "positive form" of a formula is formed by first replacing implications $A \supset B$ by disjunctions $\neg A \vee B$ and then pushing the negations as far inside the formula as possible by the classically valid De Morgan laws and quantifier dualities. He gives an example: $\forall x \exists y R(x,y) \supset \neg \forall x K(x)$ has the positive form $\exists x \forall y \neg R(x,y) \vee \exists x K(x)$. The positive form of A is denoted by $\overline{A}$.

The constructive negation translation of A is the negation of $\overline{\neg A}$. Apart from the above explanation, Gödel does not give the details of the translation, but it can be reconstructed as follows:

$$A^{con} := \begin{cases} A & \text{for atomic } A \\ \neg A_{con} & \text{otherwise} \end{cases}$$

$$A_{con} := \neg A \text{ for atomic } A$$

$$(B \& C)_{con} := B_{con} \vee C_{con}$$

$$(B \vee C)_{con} := B_{con} \& C_{con}$$

$$(B \supset C)_{con} := (\neg B)_{con} \& C_{con}$$

$$(\forall x A)_{con} := \exists x A_{con}$$

$$(\exists x A)_{con} := \forall x A_{con}$$

$$(\neg A)_{con} := \begin{cases} A & \text{for atomic } A \\ (A_{con})_{con} & \text{otherwise} \end{cases}$$

By the above translation, the translated formula will contain negations at most only at the beginning of the whole formula and in front of atomic formulas.

Below is a reconstruction of Gödel's rather more concise proof sketch (*PLI*, p. 42–44):

Theorem. *PA proves A if and only if HA proves A^{con}.*

Definition. We will use the 1932 translation, which is defined as follows. Implication is not needed, for no positive form contains an implication.

$(A)' = A$ where A is atomic

$(\neg A)' = \neg A'$

$(A\&B)' = A'\&B'$

$(A \vee B)' = \neg(\neg A'\&\neg B')$

$(\forall x A)' = \forall x A'$

$(\exists x A)' = \neg\forall x \neg A'$

Lemma. $\overline{A} \vdash_{HA} A'$

Proof. For atomic sentences, $\overline{A} = A'$.

Assume that $\overline{B} \vdash_{HA} B'$ and $\overline{C} \vdash_{HA} C'$. For $A = B\&C$, $\neg B$ or $\forall x B$, the result follows immediately.

For $\overline{A} = \overline{B \vee C} = \overline{B} \vee \overline{C}$, we have $(B \vee C)' = \neg(\neg B'\&\neg C')$. Then the derivation of $(B \vee C)'$ proceeds as follows:

$$\dfrac{\dfrac{\overline{B} \vee \overline{C} \qquad \dfrac{\begin{matrix}[\overline{B}]_2\\ \vdots \\ B'\end{matrix} \qquad \dfrac{[\neg B'\&\neg C']_1}{\neg B'}\,\&E}{\bot}\,\&I \qquad \dfrac{\begin{matrix}[\overline{C}]_3\\ \vdots \\ C'\end{matrix} \qquad \dfrac{[\neg B'\&\neg C']_1}{\neg C'}\,\&E}{\bot}\,\&I}{\bot}\,\vee E, 2, 3}{\neg(\neg B'\&\neg C')}\,\neg I, 1$$

The proof for $\overline{A} = \overline{\exists x B} = \exists x \overline{B}$ is slightly less complex but similar.

Negation is the only elaborate case because the transformation into positive form pushes negations inwards. For $\overline{A} = \overline{\neg B}$, we need to consider each possible form of B separately.

For an atomic B, $\overline{\neg B} = \neg B = (\neg B)'$. Assuming that $\overline{\neg C} \vdash_{HA} (\neg C)' = \neg C'$ and $\overline{\neg D} \vdash_{HA} (\neg D)' = \neg D'$, we can prove the other cases.

For $B = \neg C$, we have $\overline{\neg\neg C} = \overline{C}$, and the result follows from the first inductive hypothesis.

We then have cases $\overline{\neg(C\&D)} = \overline{\neg C} \vee \overline{\neg D}$, $\overline{\neg(C \vee D)} = \overline{\neg C}\&\overline{\neg D}$, $\overline{\neg\forall x C} = \exists x \overline{\neg C}$, and $\overline{\neg\exists x C} = \forall x \overline{\neg C}$ left. All of these are relatively straightforward; as an example, the proof for $\overline{\neg(C \vee D)}$ proceeds as follows.

We construct the derivation $\overline{\neg(C \vee D)} \vdash_{HA} (\neg(C \vee D))' = \neg\neg(\neg C'\&\neg D')$:

$$
\dfrac{\dfrac{\dfrac{\begin{array}{c}\dfrac{\overline{\neg C}\&\overline{\neg D}}{\overline{\neg C}}\,\&E\\ \vdots\\ \neg C'\end{array} \qquad \begin{array}{c}\dfrac{\overline{\neg C}\&\overline{\neg D}}{\overline{\neg D}}\,\&E\\ \vdots\\ \neg D'\end{array}}{\neg C'\&\neg D'}\,\&I \qquad [\neg(\neg C'\&\neg D')]_1}{\bot}\,\&I}{\neg\neg(\neg C'\&\neg D')}\,\neg I,1
$$

This concludes the proof of **Lemma 1**. *QED*

Proof of **Theorem**. Assume A is a classical theorem. Then so is $\neg\neg A$, which is classically equivalent to $\neg(\overline{\neg A}) = A^{con}$. It follows that $(\neg(\overline{\neg A}))' = \neg(\overline{\neg A})'$ is intuitionistically provable (Gödel, 1933c). But by **Lemma 1**, we have $\overline{\neg A} \supset (\overline{\neg A})'$, and by contraposition, $\neg(\overline{\neg A})' \supset \neg(\overline{\neg A})$. Therefore A^{con} is intuitionistically provable.

QED

Gödel's translation differs from the 1932 version in that although it gets rid of implication, it retains existential quantification and disjunctions. A summary of the result, dated January 1941 or slightly later, can also found in *RG* (p. 203–204). The fact that it can be found with Gödel's finished results in *RG* suggests that he thought it had individual interest. We also get the impression that he thought this translation to be a stronger result than the original one, as in *PLI*, he calls the positive form of A "the most constructive statement equivalent with A" (p. 41). In *RG*, the proof summary can be found in the index (030115) under the title "$\overline{A} \to A$ beweisbar im Heyt⟦ing⟧ Kalkül wenn A die "konstruktivste" Formulierung von A" ist" (p. 3).

Nevertheless, the translation itself is no more constructive than the previous one: what it shows is that if A is classically valid, then $\neg(\overline{\neg A})$ is intuitionistically valid. However, whereas the negation of the constructive negation (i.e., the positive form of a negative statement) of a classically valid A can be proven intuitionistically valid, this is not necessarily "the most constructive statement" corresponding to $\neg\neg A$ anymore. The translation does have some interesting properties: in addition to preserving disjunction and existential quantification, it also gets rid of double negations and restricts negations to the front of the formula and the front of atomic propositions. The positive form, therefore, has the lowest negation complexity possible, although the translation does not: e.g., the 1932 translation would translate $\forall x A(x)$ into $\forall x A(x)$, whereas A^{con} gives $\neg\exists x\neg A(x)$.

In any case, Gödel uses the negative translation to motivate his reinterpretation of intuitionistic logic as a proof of its constructivity (p. 45.1):

> The results obtained have been pretty much surprising in so far as they show that in a sense the whole classical logic is contained in the intuitionistic logic. Of course it is contained only formally, i.e., the same formulas can be proved but the meaning of these formulas is completely different. (e.g. $\neg(x)\phi(x)$ ⟦and⟧ $\sim (x)\phi(x)$) ⟦$= \overline{\neg(x)\phi(x)}$⟧. But this difference of meaning makes the result still more surprising since this means that the non-constructive classical logic has a constructive interpretation. And this makes one doubtful whether intuitionistic logic really is constructive or if not. Perhaps some non-constructive elements are hidden in the axioms which is quite possible regarding the great complicatedness in the primitive terms.

The wording is very similar to the Yale lecture, where Gödel says that the negative translation "makes one doubtful whether the intuitionists have really remained faithful to their constructive standpoint in setting up their logic or if not perhaps they have allowed some non-constructive elements to creep unnoticed into their axioms" (Gödel, 1941, 190).

Gödel's earlier remarks on the negative translation, discussed in **Chapter 2**, seem rather odd from the intuitionistic viewpoint. What can the purely *formal* translation show us about the *meaning* of the intuitionistic operators? However, the above passages from the Princeton and Yale lectures constitute a more coherent argument. If the negative translation is a deviant interpretation from the intuitionistic viewpoint, then one must be able to explain what the intended interpretation is. This leads us to the proof interpretation of intuitionistic logic, which, suggests Gödel, is not precise enough for us to see that it is really the intended one. In particular, it is not precise enough for us to see at once that it is as constructive as it should be.

This is not to say, however, that the argument is entirely intuitive. Kreisel (1987, 148–149) recalls that "[on] several occasions Kleene has coyly referred to a 'well known logician', evidently meaning Gödel, and his doubts about the disjunction and existence properties." Perhaps it is not entirely fair to say that Gödel actually *doubted* that the intuitionistic logic has the existence property, even if his remarks about the negative interpretation making the constructivity of intuitionistic logic "doubtful" do suggest this. But Gödel's project was to interpret intuitionistic logic more precisely and in so doing, make it completely clear that it has the existence property. Therefore for him, the fault is not that intuitionistic logic is not constructive, but rather that it is imprecise.

3.1.3 The argument from vagueness

The negative translation does not prove that intuitionistic logic is not constructive, but according to Gödel, it is enough to make us question whether intuitionistic logic *is* really constructive. Here the intuitionist is, according to Gödel, in trouble. For in order to show why the negative interpretation is indeed deviant, one needs to be able to specify the intended interpretation – and intuitionists, says Gödel, cannot do this. The next point in the Yale summary reads

2. Answer previous interpretation[4]

   A.) What does constructive (without existential quantifiers) mean in the strictest sense?

   B.) ⟦Intuitionistic logic⟧ can be interpreted constructively through the unclear concepts of "proof" and "procedure;" examples for axioms

   C.) Why is that unsatisfying?

   D.) Another interpretation (but only its application to specific mathematical systems)

Without the surveyability criterion, there is no immediate reason why intuitionistic proofs should not be viewed as constructive. However, whether constructive or not, they cannot be (in a formal sense) immediately recognised as so, which is why Gödel sees the proof interpretation as inadequate. In Yale, Gödel states that "the primitive terms of intuitionistic logic lack the complete

[4]

2. Antwort vorherige Interpret⟦ation⟧

   A.) Was heisst konstruktiv (ohne Existenzquantoren) im engsten Sinn?

   B.) Sie kann konstruktiv interpretiert werden mittels der unklaren Begriffe des "Beweises" und des "Verfahrens", Beispiele für Axiome

   C.) Warum ist die nicht befriedigend?

   D.) Andere Interpretation (aber nur ihrer Anwendung auf bestimmte math⟦ematische⟧ Systeme) perspicuity and clarity which should be required for the primitive terms of an intuitionistic system" (Gödel, 1941, 190). Similarly, in *PLI*, Gödel criticises intuitionists for not giving enough justification for the behaviour of intuitionistic operations (p. 7):

> The attitude of the intuitionists themselves concerning the meaning of these logical notions is this, that they take them as primitive and therefore cannot give any justification for their axioms but evidence.

The imprecision in intuitionistic proof seems to be a similar kind of an argument that Gödel makes about their *unsurveyability* in the 1930s lectures: there, he also characterises intuitionistic proofs as "vague." However, there are a few points to note about his critique of 1941. First of all, he does not consider this imprecision to mean a disqualification of intuitionistic logic, but rather criticises Heyting's proof interpretation *as an interpretation of* intuitionistic logic. Secondly, Gödel's critique here is directed against the notion of proof in itself, not just the intuitionistic notion of proof. In the Yale and the Princeton lectures, Gödel often talks about "proofs *or* procedures," in a more general sense, as opposed to intuitionistic proofs. Moreover, he uses the term *vague* as the opposite of *precise*, as in precisely defined, not of surveyable or evident.

An indication that Gödel did not so much consider the finitary unsurveyability of intuitionistic proofs as a problem at this point can be found from Gödel's notes (030078) on Bernays' "Sur le platonisme dans les mathématiques" (Bernays, 1935b). Bernays' article discusses the tendencies of Platonism and intuitionism as well as their "synthesis" of sorts in Hilbert's proof theory. The two pages of notes, written around mid-1941, are mostly summarising what Bernays said, but we do find a few of Gödel's own notes as well. His point 3 states:

> 3.) Difference between a "number" and a "proof" is that one is concretely evident and the other abstract. (Own remark: one precise and the other imprecise – "procedure" plays the same role as "proof.")[5]

[5] 3.) Unterschied zwischen einer "Zahl" und einem "Beweis" ist: Das eine ist konkret anschaulich, das andere abstrakt. (Eigene Bemerkung: Das eine präzise das andere unpräzise. "Verfahren" spielt dieselbe Rolle wie "Beweis".)

In his article, Bernays uses only the terms evident and abstract; the word "precise" (or "vague") is nowhere mentioned in this context. On the contrary, "finitely generated" or "surveyable" totalities are not, as far as I know, referred to in any of Gödel's 1941 notes. He does use the term "evident" in the passage quoted above, but this seems to mean intuitiveness rather that the concretely evident quality of finitistic objects. Certainly, "finitely generated" can also be read as "precise," by a rather natural interpretation of mathematical precision, but what is relevant is that the latter term does not carry any Hilbertian connotations of graspability, intuitive evidence, and so on. This suggests that Gödel saw imprecision as the main problem with proofs, and finding a more precise concept as a solution.

One way of making the proof interpretation precise would be to limit the notion of proof to a proof within Heyting Arithmetic, but as mentioned, that would not go through because of the incompleteness theorems. Therefore one is bound to seek another way around the problem. Gödel's solution was the functional system, which serves both as a more precise interpretation of intuitionistic arithmetic and a proof of its constructivity.

3.2 The functional system Σ

Gödel's functional interpretation can be looked at from two different angles. On the one hand, it is a contribution to the "Extended Hilbert Programme" of extracting constructive content out of non-constructive systems; in the same vein, it gives a relative consistency proof for Peano Arithmetic. On the other hand, the functional interpretation is an interpretation *of* Heyting Arithmetic, and as such, a contribution to intuitionism itself. The former viewpoint is the one that Gödel took in 1938. In 1941, he put more weight on the latter project: showing the sense in which intuitionistic logic *is* constructive. From this point of view, Gödel's functional system need not be thought of as a competitor to Brouwer and Heyting's logic, but rather as an (attempted) improvement of it. Fittingly, in modern proof theory, functional interpretations are often seen as a special kind of a proof interpretation, whether or not Gödel intended his interpretation as such.

In *Arbeitsheft 9* (030027), we find a list of objectives for the Princeton course written in early 1941. Item $2'$ summarises Gödel's general viewpoint (p. 2):

> On the basis of the intuitionistic axioms formulated by Heyting,

> criticism against them [especially the availability of negative universal statements.] *What is a properly intuitionistic* system [in particular, existential statements superfluous]. *Then also classical number theory derivable. This would perhaps be a reason against ⟦Heyting's logic⟧, but not correct, because the Brouwerian concepts are expressible in a system where no such unclarities occur. That is the goal of the lectures. It results also in a consistency proof for number theory.* First, however, the intuitionistic Heyting system and its properties.[6]

In contrast to the 1930s lectures, the consistency proof is indeed presented only as a corollary, not as the main result. Hilbert's name is not mentioned anywhere in *PLI*, and it occurs only once in the Yale lecture (Gödel, 1941, 191), where Gödel characterises systems that satisfy the strict conditions of consistency "finitistic" and closely related to "what Hilbert called the 'finite Einstellung'." As mentioned, in the Princeton course (p. 45.1) Gödel characterises such systems as "constructive in a more *precise* and stronger sense."

We saw that Gödel defined constructive logic as a logic that has the existence property. A logic that is strictly constructive in the sense that it satisfies Gödel's criteria **1–3** has this property trivially, as existential quantifiers are "defined" through their instances. Whether Heyting Arithmetic has the property cannot be seen at once because of the vagueness of the concepts involved. Gödel then sets as his goal to clarify these notions in order to secure their constructivity (*PLI*, p. 7):

> I think that these notions *can* be defined in terms of much simpler and clearer ones, at least in their application to definite mathematical theories, e.g., number theory or analysis. To give such a definition and a consequent proof of the intuitionistic axioms is the chief purpose of these lectures.

In the Yale lecture, Gödel mentions that it is "perhaps not altogether hopeless" to extend Σ to transfinite types to cover the consistency of analysis (Gödel, 1941, 200). However, he remarks that it is not certain whether the system will

[6] Aufgrund dieser intuit⟦uitionistischen⟧ Axiome formuliert [von Heyting] Kritik dagegen [insbesondere Vorhandensein der Negationen von Allaussagen]. *Was ist ein wirklich intuit⟦ionistisches⟧ ⟦System⟧?* [Insbesondere Existenzaussagen überflüssig]. Daher auch klassische Zahlentheorie ableitbar. *Das ⟦wäre⟧ vielleicht ein Grund dagegen, aber nicht richtig, denn die Brouwer'schen Begriffe ⟦sind⟧ ausdrückbar in einem System, in welchem keine solchen Unklarheiten vorkommen. Das ist der Zweck der Vorlesungen. Ergibt auch Widerspruchsfreiheitsbeweis für Zahlentheorie.* Zunächst aber intuit⟦ionistisches⟧ Heyt⟦ing'sches⟧ System und seine Eigenschaften.

still be constructive, perhaps because of impredicative elements in analysis. There is more evidence in the notebook series *Arbeitshefte* as well as *Resultate Grundlagen* that Gödel wanted to extend the functional interpretation to analysis (see, however, **section 2.2.4**).

The next item in the plan for the Yale lectures reads:

3. In which system New interpretation
   - A.) Exposition of the system
   - B.) Show in which sense constructive (A. by construction, B. by reduction procedure)
   - C.) Df. of basic concepts and proof of some axioms[7]

The "new interpretation" refers, of course, to the functional system Σ. Item 3B refers to the sense in which *intuitionistic* logic is constructive, not Σ. What "by construction" means is not entirely clear, but the Princeton Lectures suggest that there are two ways of proving the constructivity of Heyting Arithmetic: first, by showing that it can be *reduced* to or reinterpreted in a more constructive system Σ, and secondly, that it can be shown via the functional interpretation that HA has the existence property, a necessary condition for a constructive logic.

In the following sections, I will first discuss Σ as an interpretation of Heyting Arithmetic. First of all, the interpretation is not intuitionistically valid in the sense that it has been shown to assume non-intuitionistic principles, and therefore Σ is not particularly satisfying as an interpretation of intuitionistic arithmetic. The second aspect, Σ as a proof of constructivity, is not entirely unproblematic either. Gödel does not succeed in proving the existence property for Heyting Arithmetic directly, although one can still understand the interpretation as a constructivity proof by reduction. However, the question still arises whether Σ is indeed more constructive, or constructive in a more transparent way, than HA. This was a question Gödel took seriously, although he was never able to answer it in a way that he would have been happy with.

7

3. In welchem System Neue Interpret⟦ation⟧
   - A.) Ausstellung des Systems
   - B.) Aufzeigen, in welchem Sinn konstruktiv (A. Durch Konstruktion, B. Durch Reduktionverfahren)
   - C.) Df. der Grundbegriffe und Beweis einiger Axiome

3.2.1 Σ as an interpretation of Heyting Arithmetic

According to Gödel, systems that satisfy his three conditions of constructivity, are constructive in a precise sense: unlike the vaguely defined proof interpretation of Heyting Arithmetic, the quantifier restrictions of Σ make the constructivity of the system explicit. Seen from this viewpoint, Σ is a more precise interpretation of HA, and as such, a proof of constructivity of HA.

It should be noted that nowhere does Gödel claim that Σ is a kind of a *proof interpretation*; however, in *PLI*, he does set out as his objective to capture the intuitionistic meaning of the connectives in an alternative way. E.g., he motivates the translation of an implication (see below) by its meaning "in a constructive logic," and when he notices that he has misdefined disjunction (see **section 3.2.2)**, he remarks that "[the wrong translation] is a very natural and simple notion of disjunction and it is intuitionistically admissible, but it is not the notion in Brouwer's and Heyting's logic" (p. 89.1). In this sense, it is presented as an intended interpretation that should, at least, agree with the proof interpretation on the principles of inference that are and are not valid.

In general, we can ask three kinds of questions about the constructivity of the functional interpretation:

1. Is the translation obtained in an intuitionistically acceptable way, i.e., can we prove the equivalence of a HA-formula and its D-translation by intuitionistically valid inferences?

2. Is the validity of Heyting's axioms in system Σ proved in an intuitionistically acceptable manner?

3. Are the objects of Gödel's functional interpretation (finite-type functionals) more constructive than the basic objects of the proof interpretation (proofs) that they are intended to replace?

Questions 1 and 2 are related to Σ as an interpretation of Heyting Arithmetic, whereas question 3 is a more general issue related to the constructivity of Σ. I will discuss the last question, which turns out to be much more difficult to answer than the first two, later. The latter questions, on the contrary, are relatively formal and straightforward, and it turns out that both of them have negative answers. It should be noted, though, that Gödel himself did not consider these questions, and they were only posed quite a bit later.

For the first question, the main difficulty lies in the translation of implication.[8] In the Princeton course, Gödel motivates the Σ-translation of an implication $\varphi \supset \psi$ as follows.

First, let $\varphi = \exists x \forall y A(x,y)$ and $\psi = \exists u \forall v B(u,v)$ as usual. We form the implication

$$\exists x \forall y A(x,y) \supset \exists u \forall v B(u,v)$$

Our goal, then, is to transform the expression into a form which begins with a string of existential quantifiers followed by a string of universal quantifiers and finally a quantifier-free formula. Gödel notes (p. 76) that this "is not possible by simply shifting the quantifiers. But we can use the following heuristic argument. This expression means: If there exists an [x] satisfying a certain condition, then there exists a [u] satisfying another condition. In a constructive logic, that will mean: We have a procedure p which allows us to obtain such a [u] if such an [x] is given [...]" so that the implication turns into the form

$$\exists p \forall x (\forall y A(x,y) \supset \forall v B(p(x),v))$$

This inference requires a principle called Independence of Premise (IP):

$$(\forall x A \supset \exists x B(x)) \supset \exists x (\forall x A \supset B(x))$$

Here A is quantifier-free and x is not free in A. This is because we need to first move the existential quantifier in $\exists u \forall v B(u,v))$ to the front to get $\forall x \exists u (\forall y A(x,y) \supset \forall v B(u,v))$. The preceding inference of moving the universal quantifier that binds x to the front is intuitionistically valid.

As for the inner implication $\forall y A(x,y) \supset \forall v B(p(x),v)$, according to Gödel (p. 76–77), "[the] simplest meaning which suggests itself is this: Given a counterexample for the second assertion, one can construct a counterexample for the first," i.e., we get

$$\exists r \forall v (\neg B(p(x),v) \supset \neg A(x,r(v)))$$

which, using the law of contraposition, transforms into

$$\exists r \forall v (A(x,r(v)) \supset B(p(x),v))$$

[8] The Σ-translation of $\forall x A$ does require the Axiom of Choice, but this is usually not seen as intuitionistically problematic.

and the whole formula to

$$\exists p \forall x \exists r \forall v (A(x, r(v)) \supset B(p(x), v))$$

Finally, we can move the second existential quantifier over the universal quantifier (or “skolemize” it) to get the final form

$$\exists p, q \forall x, v (A(x, q(x, v)) \supset B(p(x), v))$$

In the last chain of inferences, the step that allows Gödel to proceed from $\forall y A(x, y) \supset \forall v B(p(x), v)$ to $\exists r \forall v (\neg B(p(x), v) \supset \neg A(x, r(v)))$ requires the equivalence

$$\forall y (\neg \forall v B(p(x), v) \supset \neg A(x, y)) \equiv \forall y (\exists x \neg B(p(x), v) \supset \neg A(x, y))$$

which uses a rule called Markov’s Principle (MP), formulated as

$$\neg \forall x A(x) \supset \exists x \neg A(x)$$

where A is decidable. In fact, we can say that Gödel’s translation uses somewhat strengthened forms of MP and IP because the formulas in question might still contain terms of any finite type.

In 1941, Gödel did not pay attention to the principles required from his translation; in his explanation, he seems to simply assume that they are constructive, as is suggested by his wording “in a constructive logic, that will mean [...]” above. This seems almost careless, as an intuitionist like Heyting would have clearly seen that, e.g., MP cannot hold. Nevertheless, it was Clifford Spector (1962) who first remarked that Gödel’s translation requires these rules of inference. Formally, this means that Heyting Arithmetic extended to contain all finite types – Gödel calls this Σ_I in the Princeton Lectures, but it is nowadays often denoted as HA^ω – does not prove $A \equiv A^\Sigma$. Instead we have only

$$\Sigma_I + AC + MP + IP \vdash A \equiv A^\Sigma$$

Here AC denotes the Axiom of Choice (for finite types), and MP and IP are defined as above.

As for the second question, Troelstra (1990a, 227) notes that one needs to assume the existence of characteristic functions[9] to show that the axiom $A \supset (A \& A)$ is valid. Namely, where we have

$$\exists x \forall y A(x,y) \supset (\exists x \forall y A(x,y) \& \exists x \forall y A(x,y))$$

the translation gives us the very complex-looking

$$\exists q, p_1, p_2 \forall x, y_1, y_2 (A(x, q(x, y_1, y_2)) \supset (A(p_1(x), y_1) \& A(p_2(x), y_2)).$$

Indeed, in order to find the realisers for q, p_1, p_2, Gödel uses the definition by cases (*PLI*, p. 90):

$$p_1(x) = x$$

$$p_2(x) = x$$

$$q(x, y_1, y_2) = \begin{cases} y_1 \text{ if } \neg A(x, y_1) \\ y_2 \text{ if } A(x, y_1) \end{cases}$$

Here q is then a characteristic function for A. Interestingly, Russell also met difficulties when proving the validity of the classical axiom $(A \vee A) \supset A$ for predicate logic in the *Principia Mathematica* (Urquhart, 2016, 508). The rule to which this axiom corresponds to is contraction,[10] which is the main reason for the complexity of the proof of consistency of arithmetic. Contraction-free arithmetic has a very small proof-theoretic ordinal of ω^{ω} as opposed to ε_0 for standard arithmetic (Petersen, 2003).[11]

One can argue that whether the translation is arrived at in an intuitionistically acceptable way is not relevant if the resulting system Σ can be constructively justified. However, this explanation would not be satisfying for an intuitionist because although Σ can be seen as constructive in some sense, it

[9] The conventional definition of a characteristic function is: $f : A \mapsto \{0,1\}$ is a characteristic function for a set A in case $f(x) = 1$ if $x \in A$ and $f(x) = 0$ if $x \notin A$.

[10] The rule of contraction allows us to eliminate duplicate occurrences of formulas. It can be expressed in sequent calculus as

$$\frac{A, A, \Gamma \Rightarrow C}{A, \Gamma \Rightarrow C}$$

[11] Gödel was certainly not aware of any of this in 1941; in a 1970 letter to Bernays, Gödel wrote that he could not understand why characteristic functions are required for the proof of $A \supset (A \& A)$ (Gödel, 2003a, 282). This is particularly curious considering that his proof for the validity of the axiom in *PLI* makes this fact explicit.

would not, strictly speaking, be an intended interpretation of Heyting Arithmetic. Moreover, one can argue that it is not necessary to use these principles to formulate an interpretation like Gödel's: to give an example, Kreisel's realisability interpretation (Kreisel, 1959) does not use MP. Because of these reasons, Σ is not entirely faithful as an interpretation of intuitionistic arithmetic.

3.2.2 Σ as a proof of constructivity of intuitionistic logic

What the above problems show is that it is questionable whether the functional interpretation is a natural or an intended interpretation of intuitionistic logic: the proof interpretation does not validate IP and MP, and thus the Σ-translation is not intuitionistically correct. It can be argued that it is relatively, even if not completely, faithful; and whereas MP and IP are not intuitionistically valid, they can be seen as constructive principles. Therefore the negative responses to questions 1 and 2 do not undermine the functional interpretation as a proof of constructivity for intuitionistic logic, even if it cannot be seen, strictly speaking, as an interpretation of intuitionistic logic.

The last item in the plan for the Yale lecture involves proving three results:

4. Applications

    A.) Proof that ∃ can always be constructed and even more

    B.) Consistency of number theory

    C.) Consistency of the ¬ of the Principle of Excluded Middle and comparison with the Brouwerian theses[12]

In the Princeton Lectures, Gödel defines a system as constructive if it has the existence property. The proof of the existence property for HA, then, is a proof of its constructivity, and this is precisely what Gödel attempts to prove first after presenting the functional system and the translation for HA-formulas. This result is mentioned in the Yale lecture, and a more detailed explanation can again be found in *PLI*. The proof attempt (p. 81–84) goes as follows.

12

4. Anwendungen

    A.) Beweis, dass ∃ immer konstruiert werden kann, und sogar mehr

    B.) Widerspruchsfreiheit der Zahlentheorie

    C.) Widerspruchsfreiheit der ¬ des Satz vom ausgeschlossenen Dritten und vergleichen mit Brouwer'schen Behauptungen

First, extend the quantifier-free Σ with intuitionistic quantifiers and the rules for them to get a system Σ_I.[13] If $\vdash_{HA} \exists x A$, then $\vdash_{\overline{\Sigma}} \exists x A^{\Sigma}$ and so $\vdash_{\Sigma} A^{\Sigma}[x/\alpha]$, where α is a term of any finite type and A^{Σ} is the functional translation of A (see **section 1.2.3**). Since Σ is a subsystem of Σ_I, it also holds that $\vdash_{\Sigma_I} A^{\Sigma}[x/\alpha]$. Gödel argues (p. 84) that this gives "the desired proof for constructivity of intuitionistic logic."

However, this does not yet prove the existence property *for Heyting Arithmetic*; instead, it shows the property only for its extension Σ_I. This happens because there is no guarantee that the realiser α is not a term of higher type, and therefore we cannot, in general, conclude that $\vdash_{HA} A[x/a]$ for an HA-term a. At first, Gödel does not seem to notice that the proof takes us only half-way to the proof of the existence property for HA. Likewise, in the Yale lecture, Gödel states that the proof shows that "for intuitionistic number theory such a procedure [of finding an instance] exists, as shown by the interpretation I explained" (Gödel, 1941, 199). Towards the end of the Princeton course (p. 116), Gödel notices this issue and states that one obtains the proof of existence property only for HA-sentences of the form $\exists x A$ with A quantifier-free.

Usually, the disjunction property, which states that if $A \vee B$ is provable then either A is provable or B is provable, is stated alongside the existence property as a characteristic feature of a constructive logic. However, Gödel does not ever mention this property. In fact, although he explains the meaning of an intuitionistic disjunction correctly in the introduction to *PLI* ("one has a procedure of which one knows that it must lead either to a proof of p or of q," p. 12), he first defines the Σ-translation of a disjunction analogously to the translation of a conjunction (p. 75). This is not corrected until two lectures later on p. 89.1.

Given that Gödel set as his goal the proof of constructivity of intuitionistic logic in this very sense, the fact that the proof will not go through is somewhat unsatisfying. Of course, one could still obtain a more precise formulation of the proof interpretation and with it, a proof of constructivity "by reduction:" satisfying the quantifier conditions **2–3**, Σ has the existence property by default. However, the question arises whether the system itself is constructive enough.

[13] In the Yale lecture, Gödel calls the full quantified functional system Σ. In the Princeton Lectures, however, he uses Σ to refer to the quantifier-free functional system and $\overline{\Sigma}$ to refer to the quantified extension. Here, I will use the Yale notation, as the system of notation used in *PLI* is somewhat nonstandard and replaces existential and universal quantifiers with overlined and underlined variables, respectively. To give an example, $\exists x \forall y A(x, y)$ is denoted in *PLI* by $A(\overline{x}, \underline{y})$. The expression is unambiguous since every sentence of Σ is in the $\exists\forall$-form.

3.3 Justifying the constructivity of Σ

One reason why Gödel's attitude towards intuitionism becomes less critical in 1941 might be that his functional interpretation suffers from the same problems that he accused Heyting's and Gentzen's approaches of in 1938. The notion of a functional – which in 1938, in the context of Gentzen's proof, is called "equally vast" as the concept of proof (Gödel, 1938, 109) – does not seem less complex than the notion of a proof or a procedure, and looks similarly open-ended. Perhaps Gödel dropped the surveyability criterion because he saw that his functionals could not satisfy it. However, this is not where his problems end.

In addition to problems with surveyability, it is not sure whether the system Σ even satisfies the new criteria **1–3**. The functional interpretation of Heyting Arithmetic requires all finite-typed functionals, that is, functions over functions. Whether such objects are constructive in the ideal sense still depends on whether they are computable. An inductive proof of computability is straightforward but requires all of Heyting Arithmetic, which undermines the point of giving an independent interpretation of it, but Gödel had trouble in coming up with a better proof. Seventeen years later, in the first published version of the functional interpretation, Gödel assumes computability as a primitive property which needs no formal proof.

Primitive recursive functions of natural numbers are, of course, computable, but the case for higher-type functionals is not that simple because their arguments are themselves functions. An ordinary function can be thought of as an operation on numbers or a set of ordered pairs of numbers. A function over a function (of type 1), however, operates on these operations or these sets of ordered pairs of numbers. Functionals of arbitrary type operate on arbitrarily complicated operations. The computation of a function over functions is arguably more complicated than the computation of a function over natural numbers, and each increase in type level increases complexity.

In the Yale lecture, Gödel acknowledges that unlike in finitistic systems, the objects of the functional interpretation are not only numbers, but "also functions or, in other words, procedures for obtaining numbers out of given numbers (respectively, for obtaining procedures out of given procedures [...])" (Gödel, 1941, 196). He does not problematise this, however, as "it is contained in the notion of a procedure that it can always be carried through," i.e., that we have calculability. That we get calculable terms from the given schemes of definitions (here, explicit definition and definition by primitive recursion) relies on the fact that any specific procedure we may substitute into the term will be calculable, and therefore the term will be calculable for any substitutions. Nevertheless, Gödel adds that the manner in which finite-typed functions are calculable is "pretty complicated" and that he will not discuss the question now.

3.3.1 Inductive proof of computability

One proof of computability can be found from Gödel's notebook series *Resultate Grundlagen*. Unlike the sketches in the *Arbeitshefte*, the proofs that made it to the *Resultate* are always clean and finished; there are, for the most parts, no cancellations or corrections. Gödel likely copied the proofs that he thought were important enough into the *Resultate* after he had finished working on them. The computability proof can be found on p. 188–191 in *Resultate Grundlagen 2* (030117). It is titled "Jede Fnct. d. eigentl. intuit. Systems ist berechenbar" and dated 1st January, 1941. Below, I will give an outline of the proof, which is as faithful as possible to Gödel's own notes.[14]

Definition. Where $\mathfrak{A}$ is a term and $\overline{A} = A_1, A_2, \ldots A_n$ is a series of terms, $\overline{A}$ is a *complete argument series* (c. a. s.) for $\mathfrak{A}$ if $\mathfrak{A}$ reduces to a term of type 0 (a number) with the subsequent substitutions of $A_1, \ldots A_n$ into $\mathfrak{A}$:

$$\begin{aligned} &\mathfrak{A}(A_1) \\ &\mathfrak{A}(A_1)(A_2) \\ &\vdots \\ \mathfrak{A}(A_1)(A_2)\ldots(A_n) &= k \end{aligned}$$

for an integer k.

Definition. Call a constant expression $\mathfrak{A}$ *computable* iff

1. If $\mathfrak{A}$ is of type 0, then $\mathfrak{A} = k$ is provable for some number k.

2. If $\mathfrak{A}$ is of type $n+1$, then if for any c. a. s. $\overline{A}$ for $\mathfrak{A}$, $\mathfrak{A}(\overline{A}) = k$ is provable for some number k.

[14] As Gödel wrote his notes for himself only, some of his notation is not quite transparent, symbols are not defined, and so on. In some cases, I have had to guess from the context what Gödel means by a symbol; I have added footnotes where this is the case.

Here we will call an expression type $n+1$ when the maximum of the largest types of its arguments $+1$ and the type of its value is equal to $n+1$.

In order to prove that every constant term of Σ is computable one first needs to show that computability is preserved under application and the schemes of definition for new terms.

Lemma 1. *If $\mathfrak{A}$ and $A_1, \ldots A_n$ are all computable, and $A_1, \ldots A_n$ is a c. a. s. for $\mathfrak{A}$, then $\mathfrak{A}(A_1, \ldots A_n)$ is computable.*

The result follows directly from the previous definitions.

Lemma 2. *Every constant expression built in some way out of computable constant expressions is computable.*[15]

Gödel's schemes of definition introduced in *PLI* include the schemes of explicit definition (p. 54–55) and that of primitive recursion (p. 55).

Definition. The following schemes of definition for a term φ are allowed:

Explicit definition. Where $\overline{x}$ is a c. a. s. for φ consisting of mutually distinct variables and A is an arbitrary term of type 0 containing no variables except at most those occurring in $\overline{x}$ and only previously defined constants, we may define

$$\varphi(\overline{x}) \doteq A.$$

Recursive definition. Here we assume that the argument type of φ is 0, the value type is arbitrary, and φ is defined as follows:

$$\begin{array}{rcl} \varphi(0, \overline{x}) & \doteq & A(\overline{x}) \\ \varphi(s(x), \overline{x}) & \doteq & B(\varphi(x), \overline{x}) \end{array}$$

where $\overline{x}$ is a c. a. s. of mutually distinct variables for $\varphi(0)$ which do not include x, A contains apart from $\overline{x}$ only previously defined constants, and B contains only previously defined constants and only contains φ as applied to x, i.e., $\varphi(x)$.[16]

[15] Gödel gives no proof for this lemma; it is simply stated.

[16] In the computability proof, and elsewhere in *AH* and *RG*, Gödel defines the recursion scheme by

$$\mathbf{L}(x, k, n) = \underbrace{k \ldots k}_{n}(x).$$

The symbol $\doteq$ denotes extensional equality, where for A, B of type different from 0, $A \doteq B$ iff $A(\overline{x}) = B(\overline{x})$ for any c. a. s. $\overline{x}$ for A, B.

Lemma 2 needs to be proven for both cases. For explicit definition, assuming $\overline{x}$ and A to be computable, we then have an integer n such that $A = n$, and by transitivity of equality, $\varphi(\overline{x}) = n$.

For recursive definition, the proof proceeds by induction. In the case of $\varphi(0, \overline{x})$, again assuming that $A, \overline{x}$ are computable, then $A(\overline{x})$ is computable by **Lemma 1**, and therefore $\varphi(0, \overline{x})$ is computable as well. Assuming computability for $\varphi(k, \overline{x})$ and $B, \overline{x}$, consider $\varphi(s(k), \overline{x}) \doteq B(\varphi(k), \overline{x})$. Because $\varphi(k)$ is assumed computable, the whole expression is also computable by **Lemma 1**.

Lemma 3. *If $\mathfrak{A} \doteq \mathfrak{B}$ and $\mathfrak{A}$ is computable, then $\mathfrak{B}$ is computable.*

This follows from the fact that if $\mathfrak{A}$ is computable, then for every c. a. s. $\overline{A}$ for $\mathfrak{A}$ there is an integer k such that $\mathfrak{A}(\overline{A}) = k$. But $\mathfrak{A}(\overline{A}) = \mathfrak{B}(\overline{A})$ always, so then also $\mathfrak{B}$ must be computable.

Finally, we may prove the theorem:

Theorem. *Every constant expression of Σ is computable.*

Gödel's proof (*RG*, p. 190–191) proceeds as follows.[17]

Proof. Order all of the constant expressions of Σ $\mathfrak{B}_0, \mathfrak{B}_1, \ldots \mathfrak{B}_n, \ldots$ based on their definitions in increasing order of complexity, so that each $\mathfrak{B}_k$ only contains constants that have already occurred in the list. The proof proceeds by induction on this ordering.

The first constant will be the successor function s. So let $\overline{B}$ be a c. a. s. for s. Clearly, $\overline{B}$ must be type 0, so $\overline{B} = n$ for some natural number n. But then $s(\overline{B}) = s(n)$, which is again a natural number.

Assume that all constants in the ordering are computable up to k and consider $\mathfrak{B}_{k+1}$. Here we have two cases corresponding to our two cases of definition.

1. $\mathfrak{B}_{k+1}$ has been defined by explicit definition, i.e.,

$$\mathfrak{B}_{k+1}(\overline{x}) = \mathfrak{A}$$

[17] This part of Gödel's proof consists of little more than formalism, although the reading is quite clear from the context.

where $\mathfrak{B}$ is composed of the variables in $\overline{x}$ and previously defined constants $\mathfrak{B}_i$ for $i \leq k$. Then for any c. a. s. $B_1, \ldots B_n$ for $\mathfrak{B}_{k+1}$,

$$\mathfrak{B}_{k+1}(B_1, \ldots B_n) = \mathfrak{A}'$$

where $\mathfrak{A}'$ is the corresponding expression, i.e., $\mathfrak{A}$ applied to the c. a. s. $B_1, \ldots B_n$ in $\mathfrak{A}$ (as far as their corresponding variables occur in $\mathfrak{A}$). But $\mathfrak{A}'$ is built from computable expressions and computable constants, and thus it is computable by **Lemma 2**. Since it is also type 0 by definition, there is an integer n such that $\mathfrak{A}' = n$. By **Lemma 3**, we have computability for $\mathfrak{B}_{k+1}$.

2. $\mathfrak{B}_{k+1}$ has been defined by recursion. Then let

$$\mathfrak{B}_{k+1}(a, \overline{b}) = \mathfrak{A}$$

where $a, \overline{b}$ are computable. As a must be type 0, there is an integer n such that $a = n$. We use induction on n:

a) $n = 0$. So then $\mathfrak{A} = \mathfrak{B}_{k+1}(0, \overline{b})$; as $\mathfrak{A}$ contains only $\overline{b}$ and previously defined (i.e., computable) constants, it is also computable.

b) $n = l + 1$ for some l. Then

$$\begin{aligned} \mathfrak{A} &= \mathfrak{B}_{k+1}(l+1, \overline{b}) \\ &= \mathfrak{G}(\mathfrak{B}_{k+1}(l), \overline{b}) \end{aligned}$$

for some previously defined (i.e., computable) $\mathfrak{G}$. But by inductive hypothesis, $\mathfrak{B}_{k+1}(l)$ is computable and then so is $\mathfrak{A}$.

QED

Gödel's mention of the "pretty complicated" proof in the Yale lecture suggests that he is talking about a proof that he already has at hand; because the above proof is dated January 1st, 1941, he would have known of at least this proof. However, the proof is arguably not very complicated. Halfway through the Princeton course, he believed that the inductive proof is not satisfactory because it uses the principle of complete induction in an unrestricted manner. On p. 60–61, he describes a proof of computability that matches the one above, and concludes:

> I don't want to give this proof in more detail because it is of no great value for our purpose for the following reason. If you analyze this proof it turns out that it makes use of the logical axioms also for expressions containing quantifiers and since it is exactly these axioms which we want to deduce from the system Σ.

Tait (2006b, 213–214) expresses this difficulty as follows: consider, e.g., a function $f : \mathbb{N} \mapsto A$, where A is some finite type. We define $f(n) = g^n(a)$, where $g : A \mapsto A$ and g^n signifies the n-fold iteration of g. We may formalise the property of computability as follows: that a term of type A is computable means that there is some set of conditions C_A, where

$C_{\mathbb{N}}(t) :=$ t is an integer,

$C_{A \mapsto B}(h) := \forall x(C_A(x) \supset h(x)$ computes to some y such that $C_B(y))$

Now, to prove that f is computable, we assume that g, a are computable and show that for any n, $f(n) = m$ for some integer m. For the inductive case, we need to assume that $C_A(f(k))$ for some k. However, $C_A(f(k))$ could have an arbitrary number of quantifiers in it, and therefore the computability proof requires complete induction over statements of arbitrary complexity.

As the Princeton Lectures do not contain dates, we cannot know whether Gödel wrote the above passages on the alternative proofs before or after the Yale lecture of 15th April. However, it seems likely that he was aware at this time that the straightforward inductive proof would not be satisfying from the point of view of giving a constructive interpretation of Heyting Arithmetic. Instead, one needs to come up with something that does not have to rely on HA.

3.3.2 Alternative ways of proof

Midway through the Princeton Lectures, then, Gödel was aware that the proof in the *Resultate* was not optimal for his purposes. However, he did not know how to go about finding a successful proof. After the remark that the inductive proof is inadequate, there is an incomplete passage that has been cancelled (p. 61–62):

> So our attitude must be this that the axioms of Σ (in part the schemes of definition) must be admitted as constructive without proof and

> it is shown that the axioms of intuitionistic logic can be deduced from them with suitable definitions. This so it seems to me is a program

It appears that Gödel changed his mind very quickly. The cancelled passage ends a lecture, and at the beginning of the next lecture, Gödel states that there exists another proof by transfinite induction (p. 62–63):

> The idea is the following: In a derivation to show that every function is calculable ⟦cancelled: comes to the same thing as⟧, it is sufficient to show that every constant term of type *I* ⟦integer⟧ can be transformed into a numeral by replacing in it successively all defined symbols by their definiens, and in order to show that this process of replacing comes to an end after a finite number of steps, you can associate an ordinal $< \varepsilon_0$ with each term and then show that this ordinal is diminished by every replacement.

The idea, then, would be to use transfinite induction within the basic framework of Primitive Recursive Arithmetic to construct the computability proof. Essentially, one assigns an ordinal number to each term and shows that this ordinal diminishes in every transformation towards the normal form. Gödel does remark, though, that one might think that the system Σ is in itself simpler than the calculus of ordinal numbers (p. 63^{iv}), and then the justification of the former by the latter would again make no sense.

Indeed, Gödel had doubted the constructivity of transfinite induction. In the Zilsel lecture, Gödel was suspicious of the use of the principle in Gentzen's proof, admitting that it "has a high degree of intuitiveness" (Gödel, 1938, 109) but is still problematic, mentioning that the definition of ordinal numbers in terms of induction on ordinal numbers is an impredicative procedure. In 1977, when discussing the 1941 lectures with Wang, Gödel remarked that "[the] consistency proof of [classical] arithmetic through this interpretation is more evident than Gentzen's" (Wang, 1996, 83).

However, the situation is not entirely straightforward, and it seems that to some degree, Gödel had changed his mind about Gentzen's use of transfinite induction. His functional interpretation itself was clearly inspired by Gentzen's ideas, which is supported by the fact that he titled his early sketches of the functional system in *AH* "Gentzen Wid." The two logicians could have met in Göttingen in December 1939, when Gödel gave a lecture on his results on CH. Helmut Hasse, the head of the mathematics department, had written to Gentzen and suggested that he attend Gödel's lecture (Dawson, 1997, 147). Although there is no record of a meeting, it seems that Gödel did communicate with Gentzen at some point: in a list of questions, stacked between the pages of *AH* 1 but written around 1941, Gödel has written: "General Gentzen Theorem on definability of ordinal numbers and comparison of my and his "ordinal number" of analysis" (p. 11).[18] How else could he have an idea about Gentzen's ordinal number of analysis, which – as Gentzen's notes on the topic have been lost – is not even known to the modern historians? However, nowhere does he mention Gentzen publicly, and before his early notes on the functional system have been thoroughly examined – a task that is well beyond the scope of this thesis – we can only speculate.[19]

In any case, it seems that in 1941, Gödel no longer saw the principle of transfinite induction as problematic. Strictly speaking, with the absence of criteria of permissible rules of inference[20] and surveyability given in 1938, the introduction of the principle would not outright fail any of the criteria 1–3, although the question of computability of the definition of the ordinal numbers could arise in the same manner as with functionals of higher types.

Gödel's remark about the proof by transfinite induction has been mentioned in the literature. Troelstra points out Gödel's comment that he will speak of the proof later, but remarks that "it is not likely that Gödel had actually carried through such a proof in detail" (Troelstra, 1995, 189). Van Atten (2015, 202) also notes that Gödel's cancelled passage about accepting computability as a primitive shows that he was already in 1941 open to this option. This implies that Gödel was considering alternatives to the formal proof to justify the use of functionals. The cancelled passage represents his stance in 1958, where he refers to Turing's *informal* idea of computability as a parallel, irreducible yet entirely informative concept (see **section 4.3.2**).

Looking at the Princeton Lectures only, it is difficult to figure out Gödel's aims and beliefs. His personal notes on the topic, which outnumber the pages of the lecture notes, can help us gain more clarity on the questions left unan-

[18] Allgemeines Gentzen Th⟦eorem⟧ über Definierbarkeit von Ordinalzahlen und Gleichheit meiner und seiner "Ordinalzahl" der Analysis.

[19] The folder *Logic and foundations* (050135) also contains 14 pages of notes on Gentzen's 1936 article, but they cannot be given a definite date. They appear to be written on a different type of paper than most of the 1930s and early 1940s notes – blank A4s as opposed to ruled or squared notebook pages – which suggests that they could be from later.

[20] These included (see **section 2.2.1**) propositional rules, recursive definitions, the rule of substitution, and complete induction.

swered. In the following section, I will consider some further themes in Gödel's works on intuitionism and constructive foundations read through his notes mainly from 1941. In addition to the mathematical notes on the functional system and the computability proof, which I will briefly summarise, we can also find rather more philosophical remarks. It turns out that Gödel thought deeply about the relationship between classical and intuitionistic logic, not only from the viewpoint of the classical logician but also from that of the intuitionist.

Chapter 4

Afterthoughts: beyond the lectures

In the spring of 1941, Gödel's attitude towards intuitionism had switched from mostly critical to curious, and he was open to studying new aspects of intuitionistic mathematics and its philosophy. His study of the connections between intuitionistic and classical logic also led to more meaningful insights into intuitionistic logic. In this chapter, we will encounter a theorem that connects classical and intuitionistic logic by the assumption of decidability and shows that intuitionistic logic can express any classical propositional proof. I will also discuss Gödel's views on the expressive power of intuitionistic logic, the value of the negative interpretation, and the status of the Principle of the Excluded Middle in intuitionistic logic.

Many of Gödel's early works – the negative translation and even the more natural modal interpretation of intuitionistic logic – were attempts to give classical meaning to intuitionistic logic. Despite these translations being, as Gödel put it, deviant, he took them to suggest that perhaps intuitionistic logic cannot be trusted. This, together with the critique of the proof interpretation as overly vague, motivated him to develop yet another interpretation to clarify the sense in which intuitionistic logic is constructive.

However, as seen in **Chapter 3**, the functional system Σ is neither a very natural interpretation of intuitionistic arithmetic nor a fully satisfactory proof of its constructivity. Because Σ validates principles, such as Markov's Principle, which are not intuitionistically acceptable, it cannot be said to be entirely faithful to Heyting Arithmetic. There were two different problems with the proof of constructivity. The minor problem was that the proof of the constructivity of HA in the sense of proof of the existence property does not go through *for HA*. The major problem concerns the constructivity of the system Σ itself.

In order to satisfy the strict criteria of constructivity, Gödel needed to find a proof of computability of his finite-typed functionals. This turned out to be a more complicated task than he expected it to be.

In this chapter, I will attempt to tie up loose ends from both of these aspects, namely Gödel's changed perspective on intuitionism and its logic, and the struggles with the computability proof and other mathematical goals. In the notes written during and after the Yale and Princeton lectures of 1941, Gödel began to consider interpretations of intuitionism from its own viewpoint rather than the classical one. As for the latter topic, there is a large amount of material on the functional interpretation and intuitionistic mathematics in the series *Arbeitshefte*. This material is not only large in size but relatively fragmented and requires specialist expertise in areas where I do not possess it. I do not aim to present a study of Gödel's mathematical results; my goal is to give an idea of the projects that occupied Gödel shortly before he gave up on his mathematical goals and turned towards philosophy, as well as point out possible directions for future research.

4.1 Subtle differences

In *Questions and Remarks*, likely written around 1941, Gödel writes:

> 89. Intuitionism seems to consist in that certain classical theorems are left out, in reality in that certain classical concepts are left out – in particular, the concept of "objective truth." Intuitionism can be likewise treated in two different ways:
>
> **A.** What one can say about and within intuitionistic mathematics avoiding the scorned concepts.
>
> **B.** Using the scorned concepts – in sense B intuitionism is a subsystem of classical mathematics (all intuitionistic concepts have a classical sense but not the other way around).[1]

[1] 89. Der Intuit⟦ionismus⟧ besteht scheinbar darin, dass gewisse klassische Sätze weggelassen werden, in Wirklichkeit darin, dass gewisse klassiche Begriffe weggelassen werden – insbesondere der Begriff "objektive Wahrheit". Der Intuit⟦ionismus⟧ kann dementsprechend in zweifacher Weise behandelt werden.

A. Was kann man in und über die intuit⟦ionistische⟧ Mathematik aussagen mit Vermeidung der verpönten Begriffe?

The first viewpoint defines intuitionistic logic negatively in terms of what it lacks compared to classical logic, whereas the second viewpoint gives a positive characterisation of intuitionism, as it were, on its own terms.

Around the same time,[2] Gödel had read and made notes (050135) on Heyting's 1936 article titled "De ontwikkeling van de intuitionistische wiskunde." Heyting begins with the critique of PEM in the form of a counterexample. However, after presenting the counterexample, he writes (Heyting, 1936, 131):

> It is understandable that once one has taken note of this criticism, one often asks the question: "But what is then left over of classical mathematics?" Nevertheless, the question is formulated completely incorrectly.[3]

Intuitionism is not as simple, states Heyting, as to "simply work through the Encyclopedia of Mathematical Sciences page by page and bracket out what cannot stand the test of criticism in order to obtain an Encyclopedia of Intuitionistic Mathematics" (Heyting, 1936, 131–132).[4] To write such an encyclopedia, one has to start from a scratch, based on the properly intuitionistic picture of the foundations of mathematics in terms of mental constructions. Gödel has written (050135):

> It is incorrect to consider intuitionistic mathematics as a part of classical ⟦mathematics⟧. There are entirely new fields, e.g., choice sequences, and in general the "positive" statements that, in comparison to the negative, are of independent interest.[5]

B. Mit Verwendung der verpönten Begriffe. Im Sinn B ist der Intuit⟦ionismus⟧ ein Teilsystem der klassischen Mathematik (alle intuit⟦tionistische⟧ Begriffe haben klassischen Sinn aber nicht umgekehrt).

[2] The notes in the "Logic and foundations" folder (050135) are not dated, although one can make educated guesses. During 1940–1941, Gödel used the same kind of ruled notebooks for *AH* and *RG* notebooks as well as *PLI* and the notes for the Yale lecture. The notes on Heyting, as well as notes on Brouwer's dissertation (see **section 4.2**), the latter of which can be dated to early 1942, have been written on the same paper. The later notes, on the other hand, are usually written on blank paper.

[3] Het is begrijpelijk, dat men, na van deze kritiek kennis genomen te hebben, dikwijls de vraag stelt: "Maar wat blijft er dan van de klassieke wiskunde over?" Toch is die vraag geheel onjuist geformuleerd.

[4] Er spreekt immers duidelijk de voorstelling uit, dat men slechts de Encyclopaedie der Mathematische Wetenschappen bladzijde voor bladzijde zou behoeven door te werken en alles tusschen haakjes te zetten wat de toets der kritiek niet kan doorstaan, om een Encyclopaedie der Intuitionistische Wiskunde over te houden.

[5] Es ist falsch, die intuitionistische Mathematik als Teil der klassischen zu betrachten. Es

The example of choice sequences comes directly from Heyting, who uses them as an example of a concept that has no corresponding classical notion. Heyting suggests that a positive approach might also produce concepts that are radically different from their classical counterparts, giving as examples the notion of convergence and the theory of virtual order (ibid., 133).

From early on, though, Gödel stuck to the negative standpoint. In the Hahn lectures of January 1932 (see **section 2.1.2**), he wrote that "by intuitionistic criticism, nothing at all from classical logic is lost, but rather only certain reinterpretations take place." Whereas an interpretation through many-valued logic does not work, he suggests that there could be other ways to interpret intuitionistic concepts classically so that "only the theorems of Heyting's calculus hold and nothing else." One gets intuitionistic logic by removing something from classical logic.

It is understandable that Gödel thought this way: in his 1930 three-part article on the formalisation of intuitionistic logic, Heyting claims that he went through the axioms and theorems of *Principia Mathematica* and dropped those that were not intuitionistically acceptable. But this is precisely the "Encyclopedia method" that Heyting himself sets aside as erroneous! Although Brouwer gave a philosophically richer account of intuitionism, his 1920s articles on intuitionistic logic focus on the negative thesis of criticising classical logic instead of the positive thesis of motivating the choice of intuitionistic axioms by the basic principles of intuitionism. Moreover, Gödel mainly focused on arithmetic, where the differences between intuitionistic and classical logic are relatively mild.

Gödel was undoubtedly not the only one under the impression that intuitionistic logic is defined by the classical theorems that it leaves out. These kinds of misconceptions led many to believe that intuitionists are only playing semantics. Other misconceptions include the idea that certain classical statements, such as negated universals, should lack intuitionistic meaning, something that also Gödel appears to have believed for quite some time.

Gödel saw the interconnection between classical and intuitionistic arithmetic, made explicit in his negative translation, as suspicious. As late as in 1941, he stated that the negative translation is "pretty much surprising in so far as [it shows] that in a sense the whole classical logic is contained in the intuitionistic logic" (*PLI*, p. 45.1). Whereas Gödel calls this sense, both in the

---

gibt ganz neue Gebiete, z.B. die Wahlfolgen, und überhaupt die "pos⟦itive⟧" Sätze welche im Vergleich mit der negative ein eigenes Interesse haben.

1930s and in the Princeton and Yale lectures, "only formal" or "deviant," he continued to believe that it casts doubt on whether intuitionistic arithmetic is constructive in the sense that it has the existence property. The functional interpretation, then, is introduced not only as a consistency proof but – and this is the primary motivation in the Princeton Lectures – as a proof of constructivity for HA.

Sense **A**, on the other hand, leads to what I have been calling natural or intended interpretations of intuitionistic logic. These interpretations arise from the foundations of intuitionistic reasoning, which is, in the end, a creation of different kinds of mental constructions from simpler constructions and elements of intuition by the Creating Subject. This is a radically different position from the realist metaphysics underlying classical, bivalent logic. Mathematics, for Brouwer, is a mental activity and has little to do with formalised mathematical language. The interpretation of the logical connectives in terms of proofs not restricted to the formalism itself is in line with the idea of mathematics being a partially nonformalisable process of construction.

What we see in Gödel's notes from spring 1941 onward, especially in *Max Phil* and *Questions and Remarks*, is a more refined view of intuitionistic logic, which now incorporates sense **A** as well. In what follows, I will discuss the evolution of Gödel's views on intuitionism and intuitionistic logic during and after the Princeton and Yale lectures. Although his focus was still mostly on the relationship between classical and intuitionistic logic, he had developed a more respectful attitude towards intuitionism and its philosophy.

4.1.1 Intuitionistic interpretations of classical logic

In sense **B**, intuitionistic concepts have a classical sense but not the other way around. Intuitionistic "truth" is mostly identified with provability, whereas classical logic distinguishes between truth and proof, and therefore it appears that classical logic has two concepts which intuitionistic logic flattens into one. However, at the level of meaning, the situation is reversed: whereas intuitionistic logic has less deductive power than classical logic, and in this sense, is more restricted, it has nevertheless more expressive power than classical logic.

In classical logic, one cannot express both "there exists" and "there is," to use Menger's example (see **section 2.1.3**) because $\neg\forall x\neg A$ and $\exists xA$ are equivalent. In Gödel's negative translation, both $\neg\forall x\neg A$ and $\exists xA$ get mapped into

$\neg\forall x \neg A$. As a consequence, some existence statements become weaker in the sense that we could have made a claim about something that *is* there, i.e., derived an instance, rather than claiming only existence in the sense of impossibility of the nonexistence of an instance. Likewise, the intuitionistic disjunction expresses more than the classical disjunction that can be defined in terms of negation and conjunction. In classical logic, $A \vee B$ does not indicate whether either of A and B can be known to be true, whereas intuitionistic logic differentiates between $A \vee B$ in the sense of "I can decide which of A and B holds" and $\neg(\neg A \& \neg B)$, "it cannot be that both A and B are false." One obtains finer distinctions within intuitionistic logic than are possible in classical logic.

However, as tends to happen, the higher expressive power of intuitionism comes with a price. Whereas intuitionistic logic can say more than classical logic, it can prove less; and even where the same result is both classically and intuitionistically provable, the intuitionistic proof is often more complicated. Whether this sacrifice is justified depends on one's philosophical standing. Brouwer believed that it is, at least in principle, and Hilbert thought that the renunciation of classical principles would be a disaster similar to depriving a boxer of using his fists (Hilbert, 1927, 80).

Yet there is a way in which intuitionistic logic can imitate classical proofs without changing the meaning of intuitionistic connectives. Because intuitionistic logic can make explicit the classical, non-intuitionistic assumptions made in a proof of any theorem, it is often possible to produce a similar proof where those assumptions are taken into account. In *PLI*, Gödel presents a theorem, not published elsewhere, that gives an intuitionistic translation like this of classical propositional proofs. The *truth table theorem* (p. 24-27) proves that classical and intuitionistic propositional logic coincide in decidable contexts. The presentation of the theorem below is a reproduction of Gödel's proof using his own terminology and notation wherever possible.

Definition. A *primitive conjunction* of a propositional formula A is of the form $(A_1 \& A_2 \& \ldots \& A_n)$ such that A_i is either p_i or $\neg p_i$, where $p_1, p_2, \ldots p_n$ are all the atomic subformulas of A.

Any primitive conjunction of A expresses one row in the truth table of A:

p	q	$p \vee q$
t	t	t
t	f	t
f	t	t
f	f	f

We may read off the table four different primitive conjunctions for $p \vee q$ or any formula that has the same atomic subformulas: $p\&q, p\&\neg q, \neg p\&q$, and $\neg p\&\neg q$. Moreover, each row of the table can be assigned another formula based on whether the truth value assignment given by the row satisfies the formula or not:

$$(p\&q) \supset (p \vee q)$$
$$(\neg p\&q) \supset (p \vee q)$$
$$(p\&\neg q) \supset (p \vee q)$$
$$(\neg p\&\neg q) \supset \neg(p \vee q)$$

Theorem. *Given a propositional formula A, for any primitive conjunction A' of A, either intuitionistic logic proves $A' \supset A$ or intuitionistic logic proves $A' \supset \neg A$.*

Proof. If A is an atomic formula p, then the theorem holds because $p \supset p$ is an intuitionistic theorem. Assuming the theorem holds for B, C, we have two cases: $A = \neg B$ or $A \circ B$ where $\circ \in \{\&, \vee, \supset\}$. In the case of negation, choose any primitive conjunction B' for B. Then either $B' \supset B$ or $B' \supset \neg B$. In the second case, we are done; from the first case, it follows that $B' \supset \neg\neg B$. As for a binary connective, let us consider $\vee$ for an example. Either $B' \supset B$ or $B' \supset \neg B$ and either $C' \supset C$ or $C' \supset \neg C$. Assume, e.g., that we have $B' \supset B$ and $C' \supset \neg C$. Then $(B'\&C') \supset B$ from which it follows that $(B'\&C') \supset (B \vee C)$. As $(B'\&C') = (B \vee C)'$, we have the result. The proofs for all the other cases are similar. *QED*

Corollary. *For a theorem A of classical propositional logic, A is intuitionistically provable if all atomic subformulas of A are decidable.*

Proof. If A is a classical tautology, then it has the value true on each row of the truth table; that is to say, for any primitive conjunction A', $A' \supset A$. Now let the atomic subformulas of A be $p_1, p_2 \dots p_n$. Because every conjunction of these formulas or their negations implies A,

$$(p_1 \vee \neg p_1)\&(p_2 \vee \neg p_2)\&\dots\&(p_n \vee \neg p_n) \supset A$$

is provable in intuitionistic logic. *QED*

Gödel notes that in general, one can prove in intuitionistic logic only that $(B \supset A)\&(\neg B \supset A) \supset \neg\neg A$, and therefore, $\neg\neg A$ will hold in intuitionistic logic for any classical theorem A. However, in order to conclude that A holds one needs the axiom $\neg\neg A \supset A$. From this it also follows that $A \vee \neg A$ is provable in intuitionistic logic if and only if A is decidable.[6]

The truth table theorem illustrates the fact that in decidable contexts, intuitionistic logic agrees with classical logic. Classical propositional logic becomes the special case of intuitionistic propositional logic where all basic properties and relations are assumed decidable; e.g., the logic of numerical equations. Classical logic is, so to speak, the logic of finite situations, which is why statements like Excluded Middle or Double Negation Elimination seem to make so much sense to us. Brouwer wrote in 1948 that the "long belief in the universal validity of the principle of the excluded third in mathematics" has persisted because "firstly the obvious non-contradictority of the principle for an arbitrary single assertion; secondly the practical validity of the whole of classical logic for an extensive group of simple every day phenomena" (Brouwer, 1949, 492).[7] Especially because of the latter reason, PEM has become "a deep-rooted habit of thought" which became not only acceptable in the everyday logic of finite phenomena, but also as an *a priori* truth (Brouwer, 1949, 492).

It is, however, an error of thought to generalise from the finite to the infinite without reviewing one's logical principles. Whereas basic propositional logic is indeed a logic of finite situations, the same no longer holds for quantified logic, whose domains are not necessarily finite any longer. Brouwer noticed this already in 1928: in "Intuitionistische Betrachtungen über den Formalismus," he notes that whereas the "simple excluded middle" for finitely many properties is consistent, this is not true of the more complex case of an arbitrary species of properties (Brouwer, 1928, 377–378).[8] This is to say that whereas by the above, $\neg\neg((p_1 \vee \neg p_1)\&(p_2 \vee \neg p_2)\& \ldots \&(p_n \vee \neg p_n))$ holds in intuitionistic logic, $\neg\neg\forall x(A(x) \vee \neg A(x))$ does not. Therefore intuitionistic arithmetic cannot be decidable.

[6] Jan von Plato (1999) has proved the same result from the same insight, decidability of atomic formulas. The result is also mentioned in Negri and von Plato (2001, 27).

[7] Page numbering in Brouwer (1928, 1949) refers to the reprints in Brouwer (1975).

[8] If Gödel had not read this article, he had at least planned to: in a list titled "Literatur alt" (050013), dating back to the early 1930s, we find several articles of Brouwer's, including the *Intuitionistische Betrachtungen*, as well as his dissertation.

As the assumptions made in classical proofs can be explicitly formulated in intuitionistic logic, it follows from the truth table theorem that whenever the nonconstructive inferences used in a classical proof are propositional, intuitionistic logic is capable of reproducing it in a conditional form. As Vidal-Rosset (2012, 13)[9] puts it, the non-constructive elements in classical proofs become intuitionistic hypotheses or subproblems that need to be solved before concluding the classical theorem. He takes as an example the theorem "there are at least two irrational numbers x, y such that x^y is rational." The classical proof proceeds by cases: either $\sqrt{2}^{\sqrt{2}}$ is rational or it is not. In the first case, we are done because $\sqrt{2}$ is irrational; but in the second case, we set $x = \sqrt{2}^{\sqrt{2}}$ and $y = \sqrt{2}$, which gives us the result because $\sqrt{2}^{\sqrt{2}}$ is irrational and $\sqrt{2}^{\sqrt{2}^{\sqrt{2}}} = 2$ is rational.

The first assumption, "either $\sqrt{2}^{\sqrt{2}}$ is rational or it is not," is not intuitionistically acceptable, and therefore the proof is not intuitionistically valid. However, one can prove, by the above theorem, the statement "if $\sqrt{2}^{\sqrt{2}}$ is rational or it is not, then there are at least two irrational numbers x, y such that x^y is rational" (Vidal-Rosset, 2012, 12). The theorem has the form of an implication: if the condition is satisfied, then the (classically valid) result follows. In this sense, we can always work with the propositional part of a classical proof in intuitionistic logic.

In *Resultate Grundlagen*, a more general version of the truth table theorem, listed in the index (030115, p. 3) under the title "Klassische Logik $\subseteq$ intuit. Logik ohne Annahme der Entscheidbarkeit der ausserlogischen Begriffe (nebst Hilfssatz über $\neg\neg(p \supset q)$ und Beispiel einer Heyt⟦ingschen⟧ unbeweisbaren Formel). "Ohne Annahme ..." here is slightly awkward, but apparently means that intuitionistic logic *with* the assumption of decidability equals classical logic, or that classical logic without the assumption of decidability equals intuitionistic logic. The proof can be found in *RG* 3 on p. 199–203, written around January 1941.

Gödel first notes that (p. 199)

> If one translates the basic concepts of the intuitionistic propositional calculus[10] in the following way,

[9] Page numbering refers to the archived version at `https://hal.archives-ouvertes.fr/hal-01241316/document`.

[10] The translation for the universal quantifier has been added later, and therefore this should read "intuitionistic predicate logic;" however, there is no translation for $\exists$ either.

> $\neg' p = \neg p$
>
> $p \supset' q = \neg\neg(p \supset q) \quad [= \neg\neg p \supset \neg\neg q = \neg(p.\neg q)]$
>
> $p \vee' q = \neg\neg(p \vee q) \quad [= \neg(\neg p \,\&\, \neg q)]$
>
> $p \,\&'\, q = p \,\&\, q$
>
> $(x)' = (x)\neg\neg$
>
> then each identity of the classical propositional calculus (expressed through $\neg' \quad \supset'$ etc) is provable in the intuitionistic one.[11]

The formulas given in the translations of $\supset, \vee$ can be proven equivalent, and therefore we can use the formulations with just $\&, \neg$. No proof is given for the propositional part, as Gödel states that the propositional axioms can be formulated only with negation and disjunction. The proof for the introduction and elimination rules – Gödel uses the axiomatic versions – for universal quantifier go through with relative ease.

However, one additional result is needed, namely a lemma that states that if, in our restricted calculus, $\neg\neg A$ is provable, then so is A. If we consider the result using natural deduction, it becomes clear why this is so. Assuming the translations of the premises have been proven, then we must derive the translation of the conclusion. For &, the result is immediate, as the translation does not affect the formula. The introduction rules can be easily proven sound under the translation, here, e.g., implication, where we assume that a proof $A' \vdash B'$ is given:

$$\dfrac{\dfrac{\begin{array}{c}\dfrac{[A'\&\neg B']_1}{A'}\,\&E \\ \vdots \\ B'\end{array} \qquad \dfrac{[A'\&\neg B']_1}{\neg B'}\,\&E}{B'\&\neg B'}\,\&I}{\neg(A'\&\neg B')}\,\neg I, 1$$

[11] Umsetzt man die Grundbegriffe des int⟦uitionistischen⟧ Aussagenkalküls folgendermassen:

$\neg' p = \neg p$
$p \supset' q = \neg\neg(p \supset' q) \quad [= \neg\neg p \supset' \neg\neg q = \neg(p.\neg q)]$
$p \vee' q = \neg\neg(p \vee q) \quad [= \neg(\neg p \,\&\, \neg q)]$
$p \,\&'\, q = p \,\&\, q$
$(x)' = (x)\neg\neg$

so ist jede Ident⟦ität⟧ des klassischen Aussagenkalküls (in $\neg' \quad \supset'$ etc ausgedrückt) im int⟦uitionistischen⟧ beweisbar.

The problem occurs with elimination rules. E.g., for the (special) elimination rule for implication, only the following derivation can be constructed. We are again given the premises A' and $(A \supset B)' = \neg(A' \& \neg B')$.

$$\dfrac{\dfrac{\dfrac{A' \qquad [\neg B']_1}{A' \& \neg B'}\,\&I \qquad \neg(A' \& \neg B')}{(A' \& \neg B') \& \neg(A' \& \neg B')}\,\&I}{\neg\neg B'}\,\neg I, 1$$

In order to get rid of the extra double negations, one then needs the lemma. It is indeed proven on p. 200, although Gödel does not explain why. The proof is relatively simple. For atomic formulas, the result holds trivially because no formula of the form $\neg\neg A$, A atomic, is a theorem of pure first-order logic. Assuming the result for B and C, one needs to show that it also holds for $B \& C, B \vee C, B \supset C$, and $\forall x B$. For the case of $\neg, \supset$ and $\vee$, which have negation as the main connective, the only task is to get rid of the two additional negations. Likewise, the proof for universal quantifier is easy. The only case in which the inductive assumptions are needed is conjunction:

$$\dfrac{\begin{array}{c}\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{[B \& C]_2}{B}\,\&E \qquad [\neg B]_1}{B \& \neg B}\,\&I}{\neg(B \& C)}\,\neg I, 2 \qquad \neg\neg(B \& C)}{\neg(B \& C) \& \neg\neg(B \& C)}\,\&I}{\neg\neg B}\,\neg I, 1 \\ \vdots \\ B\end{array} \qquad \begin{array}{c}\dfrac{\dfrac{\dfrac{\dfrac{\dfrac{[B \& C]_3}{C}\,\&E \qquad [\neg C]_4}{C \& \neg C}\,\&I}{\neg(B \& C)}\,\neg I, 3 \qquad \neg\neg(B \& C)}{\neg(B \& C) \& \neg\neg(B \& C)}\,\&I}{\neg\neg C}\,\neg I, 4 \\ \vdots \\ C\end{array}}{B \& C}\,\&I$$

Gödel then notes that there are formulas in classical predicate logic with only $\neg, \&, \forall$ that do not hold in intuitionistic predicate logic: his example is $\neg(\neg\forall x \neg\neg A \& \neg\forall x A)$ (p. 203). However, he remarks that

Theorem. *If all basic relations are decidable, then each classical theorem*[12] *in* $\neg, \&, \supset$, $\forall$ *holds intuitionistically.*

Because our system excludes disjunction, the decidability condition boils down to the somewhat weaker property that if $\vdash \neg\neg A$ for an atomic A, then $\vdash A$.

Strictly speaking, this is not proven by Gödel, as he leaves out implication. However, the result holds indeed: Gentzen already noticed this in a

[12] Gödel has accidentally written "each classical formula" here.

manuscript from January 1933 (see von Plato, 2017b, 186–188), and it was proven in Prawitz (1965).

This result extends the truth table theorem to a fragment of first-order logic. We expressed the connection between intuitionistic and classical logic in terms of decidability. The generalised theorem could be interpreted as follows: in an intuitionistic theory without $\vee$ or $\exists$, if any indirect proof can be transformed into a direct one, then the theory is equivalent with the corresponding $\vee$- and $\exists$-free classical theory.

4.1.2 The expressive power of intuitionistic logic

What the above shows is that intuitionists do not so much "leave out" classical theorems – they simply refuse to give *universal* validity to them. The fact that some classical theorems do not hold in intuitionistic logic is a result of a more refined definition of the logical operators. Menger's idea (see **section 2.1.3**) that the difference between classical and intuitionistic mathematics is, in reality, simply a fight over semantics is not entirely true. The distinction between the two logics is a product of the richness of the intuitionistic semantics that allows for finer distinctions which are lost in classical logic. Therefore there is a genuine argument for choosing the intuitionistic semantics, which can *also* express the classical meanings of connectives, over the classical one.

If classical logic is the logic of finite domains, Heyting Arithmetic can then be thought to consist of a classical (decidable) propositional part and intuitionistic quantificational part. Demanding that the quantificational part should not affect decidability makes sense only as a criterion of constructivity when one assumes bivalent logic as the underlying interpretative framework. Hilbert happily accepted Excluded Middle, and therefore whatever cannot be decided must be omitted from a finitistic theory. But this is exactly Gödel's demand when he motivates the ban of applying propositional connectives to quantifiers – the special case of which is negating a universal quantifier – "because quantifiers destroy decidability" (*PLI*, p. 45.1). Hilbert, too, required this restriction that arises from a predominantly classical understanding of logic.

The fallacy that undecidable concepts are meaningless to intuitionists is also based on a classical bias. Gödel was still confused over this in early 1941. In the Yale lecture, he states that the negative translation works for any intuitionistic system as long as its primitive functions are calculable and primitive relations decidable. He adds that this is a necessary condition, but that "it is perhaps a requirement of a sound intuitionistic system" that its functions are calculable and primitive relations decidable (Gödel, 1941, 195). On p. 39.3 of *PLI*, he writes that this is not the case with some intuitionistic theories:

> As to the question ⟦of⟧ whether the intuitionistic ⟦theories⟧ actually always satisfy this requirement that the primitive functions must be calculable and the primitive relations decidable, I am afraid the answer must be no since Heyting uses the set-theoretical ε-relation as a primitive and a set in the intuitionistic sense (which is called species) ⟦which⟧ may contain an arbitrary series of quantifiers in its definition.

In a passage in *Max Phil* 4 (p. 181) written around mid-1941, Gödel makes the same claim about intuitionistic geometry, making clear that he believed that undecidable relations do not belong in intuitionistic mathematics:

> Remark (*Foundations*): ε is not an intuitionistic basic concept because not decidable. Neither are the basic concepts of geometry. In their place approximation, e.g., $F_n(ab)$ = distance between a and b with the approximation $\frac{1}{n}$.[13]

Gödel had likely read Heyting's "Zur intuitionistischen Axiomatik der projektiven Geometrie" (Heyting, 1927); we find a reference to the article in the list "Literatur alt" (050013).

The basic geometrical relations of equality and incidence are, indeed, not decidable, and the same holds for equality and set membership in analysis. In general, as Gödel remarks, any continuous property cannot be decided but up to a finite approximation. Therefore, e.g., to figure out if two real numbers are equal, we can start approximating with increasing precision. If they are distinct, we will eventually find a decimal place where they differ. Their equality, however, cannot be decided in this way.

Why should undecidable concepts be meaningless to the intuitionist, who does not accept PEM? In the intuitionistic framework, meaning is connected with construction: this is what the often heard constructivist slogan "the meaning of a statement is its proof," is trying to express. A question arises, then,

[13] Bem. (*Grundl.*): ε ist int⟦uitionistisch⟧ kein Grundbegriff, weil nicht entscheidbar. Ebenso wenig die geometrischen Grundbegriffe. An ihrer Stelle Approx⟦imation⟧, z.B. $F_n(ab)$ = Dist⟦anz⟧ zwischen a und b mit der *Approx.* $\frac{1}{n}$.

about stating things which cannot be proven. One reading is that a statement that has no proof has no meaning. A weaker "possibilist" reading would be that a statement that *cannot* have a proof is meaningless. The strong reading has to be false, unless one wants to claim that, say, the statement of Fermat's theorem had no meaning before it was proven. On the weaker reading, any objectively unprovable statement, such as $\bot$, will still lack meaning.[14] Moreover, we run into epistemological problems with statements that have not yet been proven or disproven. If a statement has meaning only when it is provable, then we cannot know whether such statements are meaningful or not.

On Heyting's interpretation (see **section 1.1.3**), a *proposition* expresses an intention to produce a construction; however, the object of an intention is not a proof, which, if nonexistent, would leave the proposition itself without meaning. We are to rather understand Heyting's definition as one's intention being directed towards the conditions that need to be fulfilled in order for the construction to go through. To make a statement A is to claim that one has such a construction (proof) of A, and in this sense, the statement that A, in case that A is unprovable, lacks sense. However, it does not ban us from using the proposition A in other ways. In particular, to *assume* A is to derive consequences *from* the conditions under which A can be constructed, in the restricted context of this derivation of consequences. This distinction must be made in order to have a reasonable proof system.

Naturally, an unprovable statement cannot be justifiably asserted, and in this sense, it is mathematically pathological; however, it can still be assumed. This is why intuitionistic implications can have "ideal" antecedents, and it is the point in which, as Bernays has noted (Bernays, 1941, 147), intuitionistic logic transcends finitism, whose underlying logic is classical. Sundholm and Van Atten (2008, 70) express this relationship between the antecedent and succedent of an implication as follows:

> The proof-condition for an implication $A \supset B$ requires a relation between the proof-conditions for the propositions A and B; neither the condition itself nor its fulfillability presupposes any information concerning the fulfillability of the conditions for A and B, that is, whether these propositions really are true. In order to under-

[14] It follows from the incompleteness theorems that there are non-standard models containing non-standard numbers, some of which code infinitely long "proofs" of $\bot$, but it should be safe to say that these are not proofs in any natural sense of the word, just as non-standard numbers are not numbers in any natural sense of the word.

> stand, and even to know that an implication is true, it is only necessary to know the conditions for the truth of A and B, but not whether these conditions are, or can be, fulfilled.

The conditions themselves might be unsatisfiable, either because one does not know how to fill them or because they cannot be fulfilled, but this does not imply that the statement has no meaning. Put more simply, what a statement expresses is a construction task, giving *requirements*, which may or may not be satisfiable, for its proof. We might not even know if they are satisfiable or not, but this does not prevent us from expressing them. The intuitionist, therefore, does not need to restrict themself to the realm of the provable or the decidable.

Even though Gödel claims in the Princeton and Yale lectures that undecidable concepts have no place in intuitionism, it appears that he had already started to reconsider. In the list of improvements for the Princeton Lectures, probably written towards the end of the lectures and filed with the lecture notes, Gödel has added in shorthand that "leave out that a system of axioms is intuitionistically meaningful only if the basic concepts are decidable."[15] Unfortunately, we find no other remarks on the issue of decidability and meaningfulness of intuitionistic concepts.

However, it seems that Gödel did see the richness of intuitionistic logic also as a strength and not only as a weakness. The following passage can be found from the notebook *Max Phil* 3 (p. 138), written in April 1941, when Gödel was finishing his lecture course on intuitionistic logic and the functional interpretation:

> Remark (*Foundations*): The classical logic is in so far weaker than the intuitionistic one that it leaves out certain concepts (nondefinable out of others) ($\vee, \exists$, i.e., the concepts of "decision" and "construction"). Thereby a range of stages of approximation to a problem (which is perhaps essential for the solution) gets omitted.[16]

That this is the case is illustrated by the results in the previous section. Classical logic leaves out the "subproblems to be solved," although these could

[15] Weglassen, dass ein Axiomsystem *nur* dann int⟦uitionistisch⟧ sinnvoll ⟦ist⟧, wenn die Grundbegriffe entscheidbar sind.

[16] Bem. (*Grundl.*): Die klassische Logik ist insofern schwächer als die int⟦uitionistische⟧, als sie gewisse (aus den anderen undefinierbare) Begriffe weglässt ($\vee, \exists$, d.h. die Begriffe der "Entscheidung" und der "Konstruktion"). Dadurch fallen eine Reihe von Approximationsstufen an ein Problem (welche vielleicht für die Lösung wesentlich sind) fort.

affect the end result, as Brouwer's counterexamples of "fleeing" real numbers show.

From the above viewpoint, the negative translation shows that intuitionistic logic has more expressive power than classical logic. This is the view that Kreisel ascribed to Gentzen (Kreisel, 1971, 257). The negative translation – in fact, any of the usual translations – utilises the interdefinability of classical operators to create a mapping to what is intuitionistically equivalent to the "negative fragment" of intuitionistic arithmetic, i.e., the fragment that does not contain $\vee$ or $\exists$. That is to say, something is left out, as Gödel put it in the above remark.

The way in which classical arithmetic is, as Gödel states it in the Princeton Lectures (p. 45.1), a subsystem of intuitionistic arithmetic, is very different from the way in which intuitionistic arithmetic is a subsystem of classical arithmetic. In the second case, for any theorem A of intuitionistic logic, the *same* statement A is a theorem of classical logic. In the first case, one has to first give a mapping $'$ from classical to intuitionistic logic in order to prove that whenever A is a classical theorem, then A' is an intuitionistic theorem. A and A' only mean the same thing from the classical viewpoint, where $A \equiv A'$ is valid. The mapping will be neither one-to-one nor onto, which shows that intuitionistic logic is semantically richer than classical logic. This is something that Gödel did not pay attention to until in 1941, almost a decade after he originally came up with the translation theorem.

4.1.3 PEM and undecidable statements

Brouwer's critique of the principle of the Excluded Middle is based on the idea that there are what he called "fleeing properties," decidable properties of natural numbers of which we cannot prove whether they have an instance or not. He constructed several counterexamples to PEM in the form of such properties. Gödel was inspired by Brouwerian counterexamples and had used them before in his refutation of Behmann's thesis that any non-constructive proof can be made constructive. At the beginning of the Princeton Lectures, he motivates the existence property as a criterion of constructivity by presenting one of these counterexamples, a sequence of rational numbers which depends on solving the Goldbach conjecture. The sequence does have a condensation point, although we do not know where it lies.

In his early works, Brouwer suggests that there are absolutely unsolvable problems, i.e., statements that not only have remained undecided but also cannot be decided now or ever, and that such a problem could actually be formulated (Van Stigt, 1990, 253–254). He weakened this statement later, stating that the Subject cannot reach absolute impossibility, only the impossibility within the Subject's present knowledge.[17] However, as intuitionistic mathematics is never complete, there will always be undecidable statements, and therefore absolute *undecidability*, even if one cannot formulate absolutely undecidable statements.

On the side of formalised mathematics, a parallel result follows from the first Incompleteness Theorem: for any formal system, one can construct a statement that is undecidable yet perhaps solvable by stronger methods. Therefore even formalised mathematics can never be complete, and like in intuitionistic mathematics, there is absolute undecidability even if we cannot find absolutely undecidable sentences. In intuitionistic logic, where PEM is not equivalent to the law of noncontradiction $\neg(A\&\neg A)$, this is expressed in the failure of the universal PEM, $\forall x(A(x) \vee \neg A(x))$ for all A. The denial of this principle is a second-order formula, although we can still express that $(x)(A(x) \vee \neg A(x))$ is not intuitionistically valid by proving that HA is consistent with $\neg(x)(A(x) \vee \neg A(x))$ for some A.

One of the goals in Gödel's plan for the Yale lecture is (4C, see **section 3.2.2**) "Consistency of the $\neg$ of the Principle of Excluded Middle and comparison with the Brouwerian theses." He does mention this in the lecture, saying that "you can construct a number-theoretic propositional function $\varphi(x)$ for which it is free from contradiction to assume in intuitionistic mathematics that $\neg(x)(\varphi(x) \vee \neg\varphi(x))$" (Gödel, 1941, 199).

In *Questions and Remarks*, probably dating from around the same time, Gödel writes:

> 117. For the intuitionists, every mathematical problem is a construction task, and it is possible that both the construction tasks A and $\neg A$ have demonstrably no solution. Is that not already the case for the problem: $(x)(\exists y)(z)\varphi(x\ y\ z) \equiv yBx \vee \sim zBx$?[18]

[17] It should be remembered here that impossibility, here, does not refer to absurdity: if one could prove the absurdity of a statement, then it would not be undecidable, but merely false.

[18] 117. Für den Intuit⟦uitionisten⟧ ist jedes mat⟦hematische⟧ Problem eine Konstruktionsaufgabe, und es ist möglich, dass die beiden Konstruktionsaufgaben A und $\neg A$ nachweislich beide keine Lösung haben. Ist das nicht schon der Fall für das Problem: $(x)(\exists y)(z)\varphi(x\ y\ z)$ wobei: $\varphi(x\ y\ z) \equiv yBx \vee \sim zBx$?

Here B denotes the proof predicate; xBy is read "x is (a Gödel number of) a proof of the formula (with the Gödel number) y." The formula

$$(x)(\exists y)(z)(yBx \vee \sim zBx)$$

expresses, then, the sentence that for any statement x, either there is a proof of x or that no proof is a proof of x. That is, the statement itself claims that every problem is decidable.

A related short passage in *Arbeitsheft 7* (p. 8, backward direction) has been written in the beginning of 1941 – which appears to be when Gödel wrote the Yale notes as well as the note in *Questions and Remarks* – and is titled "Intuit. absol. unentsch. zahlentheor. Satz". The whole passage reads:

$$(x)(\exists y)(y \text{ Bew Formel } x \vee x \text{ unbeweisbar}) \tag{4.1}$$
$$(x)(\exists y)[yBx \vee (z) \sim zBx] \tag{4.2}$$
$$(x)(\exists y)(z)[yBx \vee \sim zBx] \tag{4.3}$$
$$(\exists x)(y)(\exists z)[\sim yBx \cdot zBx] \tag{4.4}$$
$$\sim yBa \cdot zBa \tag{4.5}$$
$$\sim 0Ba \cdot z_0Ba \tag{4.6}$$
$$\sim z_0Ba \cdot z_0'Ba \tag{4.7}$$

(4.1)–(4.3) formulates the undecidable statement explained above first in natural language and then formally. The negation of (4.3) is classically equivalent to (4.4), which states that there is a statement such that no derivation is a proof of it and some derivation is a proof of it. This is shown by the "derivation" (4.5)–(4.7) to be contradictory, and because of the equivalence, $\neg\neg(4.3) \equiv (4.3)$ is a classical theorem.

Intuitionistically, $\neg(4.4)$ can be easily proven. However, it is not equivalent to $\neg\neg(4.3)$, and moreover, (4.3) is not provable. Classically, we have the following sequent proof of (4.3):

$$
\begin{array}{cl}
\mathbf{z_0Ba} \to \exists y\forall z(yBa \vee \neg zBa), z_0Ba, \neg z_1Ba, 0Ba, \bot & \\
\hline
\to \exists y\forall z(yBa \vee \neg zBa), \mathbf{z_0Ba}, \neg\mathbf{z_1Ba}, \mathbf{0Ba}, \neg\mathbf{z_0Ba} & R\supset \\
\hline
\to \exists y\forall z(yBa \vee \neg zBa), \mathbf{z_0Ba} \vee \neg\mathbf{z_1Ba}, 0Ba \vee \neg z_0Ba & R\vee x2 \\
\hline
\to \exists y\forall z(yBa \vee \neg zBa), \forall\mathbf{z}(\mathbf{z_0Ba} \vee \neg\mathbf{zBa}), 0Ba \vee \neg z_0Ba & R\forall \\
\hline
\to \exists y\forall z(yBa \vee \neg zBa), \mathbf{0Ba} \vee \neg\mathbf{z_0Ba} & R\exists \\
\hline
\to \exists y\forall z(yBa \vee \neg zBa), \forall\mathbf{z}(\mathbf{0Ba} \vee \neg\mathbf{zBa}) & R\forall \\
\hline
\to \exists\mathbf{y}\forall\mathbf{z}(\mathbf{yBa} \vee \neg\mathbf{zBa}) & R\exists \\
\hline
\to \forall x\exists y\forall z(yBx \vee \neg zBx) & R\forall
\end{array}
$$

Intuitionistically, however, the proof will terminate at the fourth row from below: application of rule $R\forall$ will destroy rest of the context, leaving us with $\to 0Ba \vee \neg z_0Ba$.

Gödel's example expresses a fleeing property: although yBx is decidable, one cannot give, for an arbitrary x, a proof of an instance (i.e., the existence of a proof) or a proof of no instance (i.e., the existence of no proof). This is consistent with the metatheory of classical arithmetic. As PA is consistent, it cannot be proven that there is a proof of $\bot$. However, it still cannot be proven, within arithmetic, that no proof is a proof of $\bot$. Nevertheless, classical arithmetic cannot express the notion of decidability because it does not have an independent concept of disjunction.

$\neg$ (4.3) is naturally not provable, either, as it would then also be classically provable and contradict the fact that (4.3) is classically provable. Therefore (4.3) is undecidable in intuitionistic logic.

In *PLI*, we find a semantic proof of the consistency of the negation of the universal PEM in the system Σ; this part (p. 107–117) constitutes the last lecture of the course. Gödel uses what are today called *Hereditarily Effective Operations* (see Troelstra, 1973, section 2.4.11) to build a model in which the desired number-theoretic function is constructed. In short, we code into $\varphi(\ulcorner x \urcorner)$ the statement "x is a non-total recursive function." Then $\forall x(\varphi(x) \vee \neg\varphi(x))$ means that the Halting Problem is solvable, which cannot be the case, and therefore we obtain the result.

This kind of a concrete failure of PEM is a difference between intuitionistic propositional logic and predicate logic. The truth table theorem shows that although $p \vee \neg p$ is not a theorem of intuitionistic propositional logic, its weakened form $\neg\neg(p \vee \neg p)$ is, showing that $\neg(p \vee \neg p)$ is inconsistent with intuitionistic propositional logic. However, the above counterexamples show that the same does not hold for the quantified version of PEM. Intuitionistic analysis goes even further in that it not only abstains from asserting the universal PEM but, in fact, denies it: in Brouwerian analysis, $\neg\forall x(A(x) \vee \neg A(x))$ can be proven for a suitably chosen A. Therefore $\neg\neg\forall x(A(x) \vee \neg A(x))$ cannot be a theorem of intuitionistic logic. Brouwer showed this in 1928 on the grounds that the intuitionistic continuum cannot be split into two (nonempty) parts (Brouwer, 1928).

4.2 Gödel on the nature of intuitionistic reasoning

Although the view **A**, "what one can say about and within intuitionistic mathematics avoiding the scorned concepts," suggests a programme of interpreting intuitionism from a natural viewpoint, i.e., interpreting it based on Brouwer's positive characterisation of intuitionism, this is not a programme that Gödel explicitly contributed to. However, even if the functional interpretation is not entirely faithful to Heyting Arithmetic, it is nowadays often seen as one kind of a proof interpretation, and thereby a contribution to intuitionism as it is practised in its modern form, more inspired by Heyting than Brouwer. In this section, I will discuss Gödel's views on intuitionism beyond its logic as well as the interpretation of the functional system based the notes in *Max Phil* and *Questions and Remarks*. It is a historical stretch to call it a proof interpretation, notwithstanding that Gödel's notes, as we will see, suggest that the goal of the functional system to make intuitionistic concepts more precise is in line with what he believed is the purpose of intuitionistic mathematics.

Gödel was not as well acquainted with Brouwer as he was with Heyting's works. In 1975, he stated that he read Brouwer first in 1940 (Gödel, 2003a, 447). He probably meant that he did not study Brouwer's works thoroughly before that, as it seems that he at least knew the contents of some of Brouwer's 1920s articles. He worked intensely on Brouwerian analysis in the autumn of 1941, although he did not seem quite as interested in Brouwer's philosophical views. There is one exception: Gödel read and made extensive notes (030135) on Brouwer's dissertation in March 1942.[19] The 14 pages of notes show that

[19] There is no date in the notes, and they have been stuck in the same folder with notes mostly from the 1950s. However, there are several pieces of evidence suggesting that the notes on Brouwer's dissertation are from the early 1940s. First of all, Gödel wrote to his brother in September 1941, asking if he could obtain a copy of Brouwer's dissertation (Van Atten, 2015, 190–191). Secondly, in the *Max Phil* 4, one can find a margin note "Beginn Lektüre Brouwer ca. Ende März" on p. 243 and a corresponding "Ende Lektüre Brouwer Ende März 42" on p. 258. In the front page of the notebook, there is a note "März 42 p. 243–58 Brouwer Unterbrechung durch Grundlagen," probably referring to *Over de grondslagen*. Finally, in *AH* 14 there are notes on Brouwer's non-Archimedean systems dated March 1942 which show parallelism with the

Gödel also read the philosophical parts very carefully and understood them well. In any case, these notes were made later than most of the material in *Max Phil* 3 and 4 as well as *Questions and Remarks*, and there are few remarks on intuitionism after mid-1941; therefore it is difficult to say how the reading of Brouwer's dissertation affected Gödel's view on intuitionism.

Gödel met Brouwer personally only once when the latter visited Princeton in 1953. In a letter to his mother dated 31st October, Gödel mentions a famous professor from Holland who has come for a visit. He writes that his lectures were not very popular with the audience ("u. mit Recht"), but that he had to invite him to his house, once for lunch and once for tea. He adds that Brouwer, who was over 70 at the time, was probably looking to earn money in the United States, as the pensions are relatively low in the Netherlands. It is fairly clear that no matter what Gödel thought about intuitionism, he did not much appreciate Brouwer's personal style, which was, admittedly, quite different from his own.

Gödel's early opinion of intuitionism was generally negative, and even in the 1941 lectures, he is relatively critical toward intuitionistic logic and the proof interpretation. Despite all of this, the tone of his personal notes is neutral or even positive. There is no trace of the critical arguments seen in the previous chapters. I should also add that "finitism" is very rarely mentioned in the early *Max Phil* notebooks or *Questions and remarks* and Hilbert's Programme, as far as I know, not even once. This is all very much in contrast with the 1930s lectures, where the discussion of constructive foundations is centred around the Programme.

In a remark written in *Max Phil* 3 around April 1941, Gödel characterises intuitionism as follows (p. 145):

> Remark (*Foundations*): The work towards building intuitionistic mathematics has the following characteristics:
>
> 1. The questions often dissolve already through clarification of the concepts (i.e., it all follows from the definitions).
> 2. The answer is mostly unambiguous (what is "correct").
> 3. There is a close relation to word-language and the language-intuition can be used with benefit.

---

notes on the dissertation.

4. One has the feeling that there is something "deep" behind there.
5. Many foundational (and philosophical) problems find their exact formulation and a solution "in a model."[20]

Although Gödel never wrote broadly on the topic, we find some fragmentary remarks in *Max Phil* and his other notes written at the same time. Points 1 and 3 represent two recurring themes in Gödel's notes. It is clear why Gödel wrote down point 5: even if he did not spend extensive time on the philosophical issues related to intuitionism, he appreciated its foundational value, as we can see from the vast amount of notes he wrote on intuitionistic mathematics in the latter half of 1941. As a concrete example, he believed that he could prove the independence of the Continuum Hypothesis using intuitionistic analysis. In what follows, I will discuss Gödel's philosophical ideas, and in the next section, I am going to give a summary of the mathematical notes.

Point 1 states that the mode of problem-solving characteristic to intuitionistic mathematics is to clarify concepts. The idea that finding the right concepts and making them clear enough leads to progress in mathematics is characteristic to Gödel's thought. E.g., in the often discussed appendix to "What is Cantor's continuum problem?" (Gödel, 1964), Gödel expresses the view that problems such as the Continuum Hypothesis can be solved if we make our concepts clear enough to come up with the "correct" additional axioms to ZFC. However, this view is already present in his philosophical notes of 1940 and 1941.

Towards the beginning of *Max Phil* 3, written in October or November of 1940, there is a long list of mathematical activities (*Tätigkeiten des Mathematikers*), the second of which is "Präzisieren (eines anschaulichen Begriffs)" (p. 28) – "Präzisieren" being an expression he would often be used interchangeably with "Klarmachen" in the context of making concepts more precise. In

[20] Bem. (*Grundl.*): Das Arbeiten in der Richtung des Aufbaus der intuit⟦uitionistischen⟧ Mathematik hat folgende Charakteristika:

a) Die Fragen lösen sich oft bereits durch Klarmachen der Begriffe (d.h. es folgt alles aus den *Df.*).
b) Die Antwort ist meist eindeutig (was das "Richtige" ist).
c) Es besteht eine nahe Verwandtschaft zur Wortsprache, und das Sprachgefühl kann mit Nutzen angewendet werden.
d) Man hat das Gefühl, dass etwas "Tiefes" dahintersteht.
e) Viele Grundlagen- (und philosophische) Probleme finden dabei ihre exakte Formulierung und Lösung "an einem Modell".

January 1941, when Gödel was starting to work on the functional interpretation, he has written (p. 43):

> Remark. The real progress in mathematics and philosophy consists in that one first has certain few vague concepts. By allowing them to somehow work against each other and by applying them to each other, they become sharp and create other vague concepts which will be handled in the same way, etc.[21]

We find several more remarks on the fruitfulness of concept-clarifying in mathematical work in the first half of *Max Phil* 3. Gödel suggests that the business of constructive mathematics is particularly often like this: we read, e.g., that for Gentzen's consistency proof "and related questions," it is sufficient for the proof to "make the concepts clear" (p. 29).[22] Likewise, on p. 70:

> Remark. The question, in which direction one can work toward in mathematics is the same as which questions can be solved by a mere "clarification of concepts" or which problems can be solved by mere natural reasoning (or by mathematical method applied *lege artis*).[23]

What kind of concepts does intuitionistic mathematics attempt to clarify? Gödel's understanding here looks, at first sight, somewhat different from Brouwer's thought. In the characterisation of intuitionism in *Max Phil* 3, point 3 relates intuitionism to natural language and language-intuition, intuition of natural language concepts. "Sprachgefühl" is a word that denotes something similar to what linguists call the "native speaker's intuition," except that Gödel's language-intuition does not refer only to syntax, but also to semantics. The aspect of making precise that is connected to intuitionism, then, is making concepts of *natural language* more precise.

In mid-1941 (*Max Phil* 4, p. 182), Gödel has written:

[21] Bem. Die tatsächliche Fortschritte in Mathematik und Philosophie bestehen darin, dass man zuerst einige wenige unscharfe Begriffe hat. Indem man diese irgendwie gegeneinander wirken lässt und aufeinander anwendet, werden sie scharf und erzeugen neue unscharfe Begriffe, welche in derselben Weise behandelt werden, etc.

[22] Bem. Für den Gentzen'schen Widerspruchfreiheitsbeweis und zusammenhängenden Fragen scheint es wirklich so zu sein, dass das "Klarmachen der Begriffe" für den Beweis hinreicht.

[23] Bem. Die Frage, in welche Richtung kann man in der Mathematik arbeiten is dieselbe, wie welche Fragen lassen sich bloss durch "klarmachen der Begriffe" lösen oder welche Probleme lassen sich bloss durch natürlichen Vernunft (oder durch math⟦ematische⟧ Methode *lege artis* angewendet) lösen.

> Remark (*foundations*): Intuitionism is somehow more akin to word-language than classical mathematics: cf. e.g. quantifiers by two different variables. Difference between "all" and "each." German words are imposed upon it: complete argument series, level, to give more information, etc. The development of intuitionistic mathematics leads to making precise certain words of the German language.[24]

If we read the passage literally, it is an obvious misdescription of Brouwer's intuitionism. Investigating language in order to find out something about mathematics is like trying to analyse, say, the formation of cyclones by looking more closely at a weather map: one identifies the (incomplete) representation with the phenomenon itself and studies the properties of the second-level object as opposed to the original object.

Nevertheless, if we replace linguistic concepts by mental concepts or constructions, the resulting view is perhaps closer to intuitionism. The idea that basic concepts of intuitionism are psychological also occurs in some of Gödel's notes. In notes on Bernays' article "Sur le platonisme dans les mathématiques" written some months later, Gödel writes "Intuitionismus (= Psychologism)," and in a remark from March 1942 (*Max Phil* 4, p. 258), around the time he was reading *Over de grondslagen*, states that

> The meaning relation has nothing to do with psychology (not even idealised) [except that perhaps in Brouwer meanings of mathematical statements are psychological things, just as in psychology.][25]

In *Max Phil* 3 (p. 149), written in April 1941, Gödel has written down a goal to "put the psychological concepts in order," and listed reasons why this would be a fruitful project. The first reads, "Anwendung für die Grundlagen (Int⟦uitionismus⟧ ist eine schematisierte Psych⟦ologie⟧)."

---

[24] Bem. (*Grundl.*): Der Intuit⟦uitionismus⟧ ist irgendwie mit der Wortsprache verwandter als die klassische Mathematik: vgl. z.B. Quantoren durch zwei verschiedenen Variablen. Unterschied ⟦zwischen⟧ "alle" und "jedes". Es drängen sich deutsche Wörter auf: vollständige Argument-Reihe, Niveau, mehr Informat⟦ion⟧ geben, etc. Die Entwicklung der int⟦uitionistischen⟧ Mathematik führt dazu, dass gewisse Worte der deutschen Sprache zu präzisieren sind.

[25] Die Bedeutungrelation hat nichts mit Psych⟦ologie⟧ (auch nicht idealis⟦iert⟧) zu tun [ausser vielleicht bei Brouwer die Bedeutungen der mat⟦hematischen⟧ Sätze psych⟦ologische⟧ Dinge sind ebenso wie in der Psychol⟦ogie⟧.]

Clarification of concepts as a method of problem-solving makes sense particularly in the context of "psychological" intuitionism, whose objects are mental constructions. In *Questions and Remarks*, Gödel writes:

> 29. There are perhaps statements (logic and set theory) of which nothing is left over when one leaves out the human thought-language. (Neither might there be objective facts left over in actual mathematics.)
>
> I.e., the former would be analytic in the true sense of the word.
>
> Criterion: in order to solve the problem, it is enough to make the concepts clear.[26]

Gödel seems to believe that the concepts that intuitionism clarifies are intuitive or "everyday" concepts that are closer to the human experience than the ones classical mathematics is occupied with. In the previous section, we saw that classical logic had no way of expressing statements that seemed intuitive, e.g., the fact that some problems are undecidable. What the clarification of these intuitive concepts amounts to and whether it involves formalisation – something that Gödel had placed importance in the 1930s but grown increasingly displeased with in the 1940s – is not made clear in Gödel's notes.

The idea of concept clarification in intuitionism is connected, to some degree, to the functional interpretation. In the Princeton Lectures, Gödel stated as his goal to make the definitions of intuitionistic connectives more precise in order to make their constructivity transparent. This, too, is problem-solving by clarifying concepts. One could go as far as to say that the functional interpretation makes precise the concept of proof in Heyting Arithmetic. Van Atten and Kennedy (2009, 494) write that seen this way, the intuitionist can accept Gödel's grounds for the functional interpretation, because "if one considers the Proof Interpretation, not in general, but limited to a particular topic, it is only to be expected that it can be turned into something more specific; the computable functionals are an example of just this for the case of arithmetic." They do note, though, that Gödel made it clear in the *Dialectica* article that he does not claim that the functional interpretation of HA "reproduce[s] the

[26] 29. Es gibt vielleicht gewisse Sätze (Logik und Mengenlehre) von denen *nichts* übrig bleibt, wenn man die menschliche Denksprache weglässt. (Weder vielleicht in der wirklichen Mathematik objektive Sachverhalte übrig bleiben).

D.h. die erste wären im wahren Sinne des Wortes analytisch.

Kriterium: zur Lösung des Problems genügt es, sich die Begriffe klarzumachen.

meaning of the logical particles introduced by Brouwer and Heyting" (Gödel, 1958, 249–251).[27]

In the end, it seems that Gödel wanted to, at least, leave open the question about the intuitionistic meaning of the functional interpretation. In a letter draft to Dirk van Dalen dated 13th October, 1969, not included in the *Collected Works*, Gödel writes:

> My relationship with Intuitionism consists primarily in some theorems I proved about certain parts of intuitionistic mathematics in particular that publication in *Dialectica* 12. The question as to whether this paper is important for the *foundations* of Intuitionism I must leave for Intuitionists to answer. I did not write the paper from this point of view and some supplementation would be necessary in order to clarify its relevance to the foundations of Intuitionism.

To make things more complicated, there is a 17-year gap between the Yale and Princeton lectures and the *Dialectica* article of 1958, and the contexts of the two are somewhat different: unlike in Princeton and Yale, in the 1958 article Gödel frames the discussion around the Hilbert Programme and not intuitionistic logic. In the light of what has been seen so far, it is not far-fetched at all to say that Gödel saw the functional interpretation as a contribution to Brouwer's intuitionism, as well. I already mentioned in **section 3.2.1** that in *PLI*, Gödel explicitly aimed at explaining the meaning of intuitionistic connectives in a more precise way.

When talking about his earlier works, Gödel tended to downplay the extent to which he considered philosophical questions. He might not have published many philosophical works, but he thought of foundational questions in his philosophical notebooks and even earlier, e.g., in two notebooks titled *Quantenmechanik*, written in 1935, where he not only discusses quantum mechanics but philosophical questions related to physics and mathematics.[28] Certainly his philosophical standpoint in the 1960s and 1970s is more mature, but it is not unexpected that one's views should become more refined in 20–30 years' time.

[27] Page numbering in Gödel (1958) refers to the English translation in Gödel (1990).

[28] The *Quantenmechanik* books have now been edited for publication by Tim Lethen and Oliver Passon (2020).

Gödel's notes in *Max Phil*, as well as *Questions and Remarks*, paint a somewhat different picture of his views on intuitionism than the lectures. They show that Gödel did think about the nature of intuitionism, and he saw many positives in it. For him, intuitionistic mathematics was a tool that is useful in solving some problems and a viewpoint that helps to see some issues more clearly: something that has both heuristic and philosophical value.

Above, I have tried to explain the sort of philosophical value that Gödel believed intuitionism and constructive mathematics has. In the last section, I will briefly discuss the heuristic value. Whereas the philosophical remarks on intuitionism in Gödel's notebooks are concise, Gödel's studies in intuitionistic logic and mathematics comprise hundreds of pages of often fragmentary notes in the *Arbeitshefte*. Transcribing, translating, and analysing the material in detail will be a lengthy process that goes beyond my present aim, but in what follows, I hope to give a rough idea of Gödel's mathematical goals in 1941. Moreover, I will briefly discuss the *Dialectica* article of 1958 in relation to what has been said so far.

4.3 The functional interpretation after Princeton and Yale

Several questions concerning the functional interpretation were left open in the previous chapter, and Gödel himself did not return to them until much later. The lectures on intuitionism and the functional interpretation were not widely known until Gödel finally presented the functional interpretation in public in a paper titled "Über ein bisher noch nicht benützte Erweiterung des finiten Standpunktes" (Gödel, 1958). As the title shows, Gödel's main focus is not intuitionistic logic; instead, the functional system is presented as an extension to Hilbert's finitary standpoint.

As for the worries around the constructivity of the interpretation, it appears, at first glance, that Gödel does not so much provide answers, but rather refines his views to avoid giving them. He no longer states that the functional interpretation is superior to Gentzen's approach. The existence property, likewise, is not discussed. The "strict criteria of constructivity" are not mentioned at all, the question about the computability proof is dropped, and identity of all types of functions is, as the cancelled passage in the Princeton Lectures suggests, assumed as primitive.

To which extent does all of this express a genuine change in philosophical viewpoint, as opposed to a way to get around questions that Gödel could not answer? In what follows, I will first discuss Gödel's investigations in intuitionistic mathematics. The time between spring and autumn of 1941 was taken up by intensive work on the topic, but quite soon after, Gödel gave up his work on intuitionistic mathematics, and by 1943, after several failed endeavours, logic and mathematics altogether. The disappointments that Gödel encountered were, I believe, a partial reason for why he did not return to the topic for almost two decades, and for his changed viewpoint of 1958. However, considering the development of Gödel's philosophical views in the early 1940s – which might be themselves motivated by the frustration caused by failed mathematical goals – it is fair to say that there was a genuine shift in Gödel's thinking soon after the Princeton Lectures.

4.3.1 Computability proof and mathematical goals

Arbeitsheft 9 (p. 2–3) contains a list, parts of which were already quoted in **section 3.2**, titled "Vorl: 1941 Sommer." This list turns out to be a plan for the Princeton lecture course.[29] Nearly every item in the list of 11 points appears in the lectures, apart from point 9 "Impossibility of the [consistency] proof in a smaller system"[30] and the last point of the list, "Computability of all functionals in system **S**."[31] As explained in the previous chapter, the latter problem would become a vital issue in evaluating the functional system as a more *constructive*, as opposed to just more precise, alternative to the proof interpretation. In the Princeton Lectures, three alternatives were mentioned: the unsatisfactory inductive proof, no proof, and a proof that uses transfinite ordinals.

It is not entirely straightforward to work out the relationship between the three alternatives and the motivation behind each one of them. The inductive proof was the earliest one, which Gödel soon realised was unsatisfying. We know that he was writing in haste, and it might be that he wrote the part on no proof as a last resort, unable to come up with a proof, and then cancelled

[29] "Sommer" here refers to the Spring Term at the Institute for Advanced Study, which took place between February and May. It is common in Germany and Austria to divide the academic year into *Winter- und Sommersemester*, where the Winter Term would begin around October and the Summer Term around April.

[30] 9′. Unmöglichkeit des Beweises im einem kleineren System.

[31] 11. Berechenbarkeit aller Funktionen in **S**. ⟦**S** here is an alternative name for what Gödel would later call system Σ.⟧

it as he had another idea that he believed could be carried through. However, I do not think this is the case, as it appears from Gödel's notes that he was aware of the proof using transfinite ordinals before the Yale lecture and before introducing the functional system in the Princeton Lectures.

Gödel's remark in the Yale lecture that "it is contained in the notion of a procedure that it can always be carried through" (Gödel, 1941, 196) can be interpreted as a statement that a proof is not *necessary*; however, he does mention that a proof for this is "pretty complicated." This suggests that Gödel is not referring to the simple inductive proof. Curiously, this passage in the Yale lecture is a later addition on a blank sheet stuck between two spiral notebook pages. It appears that Gödel was, at this point, genuinely of two minds about the proof.

It would be interesting to know whether the passage in the Yale lecture was written before or after the Princeton notes. Given the dates in the *Arbeitshefte*, as well as what Gödel writes about the computability proof, it is plausible to conjecture that most of the Yale notes were written in the beginning of 1941. As for the Princeton notes, we know that the semester took place between February and May and, based on the handful of mistakes which are corrected later in the lectures, it also seems that Gödel was preparing the notes one lecture at a time shortly before the lecture. It seems, then, that Gödel wrote the passages about the computability proof in *PLI* around the same time that he made the addition to the Yale lecture.

It is highly likely that Gödel did not come up with the desired proof by transfinite induction, as no proof is presented in the 1941 lectures or the 1958 publication. However, it seems that he did at least try. In the *Arbeitshefte*, in particular, there is a large amount of material related to the functional interpretation and other studies in intuitionistic mathematics. To give an idea of the extent of Gödel's notes, we may look at the index to the *Arbeitshefte* (030016) prepared by Gödel himself.[32] Under the topic "Interpretation der intuitionistischen Logik," there are in total 96 notebook pages, and under "Widerspruchsfreiheit $\neg(p)(p \vee \neg p)$" another 43 pages. These notes are, for the most part, directly related to the Princeton course and contain early sketches of the results in the lecture notes. Under "Zusammenhang zwischen ε_0 und System der endlichen Typen," related to the computability proof, we find another 50 pages of notes. As for the extensions of the functional system, the pages listed under

[32] The English translation of this list can be found in full in Dawson and Dawson (2005).

"Berechenbare Funktionen höheren Typs" are 81 in total. Another item related to intuitionistic mathematics is "Unabhängigkeit Kontinuumshypothese nach Brouwer'scher Methode," under which 66 pages of notes are listed.

Gödel worked on the topic with a manic intensity. The first notes on the functional system are in *Arbeitsheft* 7, dated January 1941. The last notes on the Brouwerian method for Continuum Hypothesis are in *Heft* 13, where the date that appears 20 pages after Gödel's last notes is "Beginn 42." During that time, he wrote over 300 pages of notes on the functional interpretation and intuitionistic mathematics only, and this excludes the Yale and Princeton lecture notes. During 1941, he also wrote 150–200 shorter philosophical remarks on various topics in *Max Phil* as well as 100 pages of mathematical and logical results in *Resultate Grundlagen*. Some of his loose notes, such as the 45 pages of *Questions and Remarks*, seem to be dated around the same time, as well. It is no wonder that he took a break from mathematical work in 1943.

The notes on the computability proof can be found, for the most part, in *Hefte* 8 and 10. Towards the end of *Heft* 8, we find the date 1./II.41; *Heft* 10 contains no dates. Given that *Heft* 12 is dated 19./IX.1941, the *Hefte* 9–11 have likely been written in the spring and summer of 1941.

It seems, then, that the first proof that Gödel had was the inductive proof in *RG*, but that shortly afterwards, he started experimenting with an alternative proof. Given the dates of *Heft 8*, it appears that he must have known about the existence of such a proof before writing the Yale lecture and the above passages in *PLI*. But as the no proof option is still mentioned in the Princeton Lectures and hinted at in the Yale lecture, it appears that he did take it seriously. I will discuss this briefly in the next section.

Finally, a quick note on Gödel's other mathematical aims. As seen from the topics listed in the index to the *Arbeitshefte*, he did study intuitionistic mathematics beyond the functional interpretation. Besides the computability proof, some other themes that arise in the *Arbeitshefte* are the extension of the functional system to cover analysis and using intuitionistic methods to prove the independence of the Axiom of Choice and eventually the Continuum Hypothesis. In the Bulletin of the Institute for Advanced Study for the academic year 1941–1942 it is stated that "[in] 1941 Dr. Gödel lectured on some results concerning intuitionistic logic, and in 1941–1942 he will continue his researches on this subject and its connection with the continuum problem."

Considering the dates of the *Hefte*, it seems that Gödel had quit working on the computability proof already in the autumn of 1941: *Heft* 12 has the date 19./IX.1941, and the last entries on the proof are in *Heft 10*. It seems that in September, he was already working on set theory, as seen from topics in *AH* and *RG*. *Heft* 13, which contains notes on Brouwerian analysis, has been written around the beginning of 1942. By March of 1942, a date that can be found in *Heft* 14, we find no more remarks on any of these topics. Notably, although some results on intuitionistic logic are contained in *RG* 3 and 4, the topics in *AH* listed above do not have corresponding entries in *RG*, suggesting that they did not lead to finished results. No publications came out of these investigations.

4.3.2 Between 1941 and 1958

In the end, Gödel chose not to publish any of the lectures of 1933, 1938, or 1941. In a letter to Frederick Sawyer written around 1974, Gödel writes that the reason for this was that his interests had shifted, and that "there was not too much interest in Hilbert's Program at that time" (Gödel, 2003b, 210). Indeed, both are true, although it also seems that Gödel became frustrated with the lack of progress in the goals related to the functional interpretation and intuitionistic mathematics in general.

According to Kreisel (1987, 104), Gödel "dropped the project [of the functional interpretation] after he learnt of recursive realizability that Kleene found soon afterwards." The realisability interpretation (Kleene, 1945) uses general recursive functions to define "realisers," i.e., numbers which validate a formula, for each statement of intuitionistic arithmetic. Then, e.g., $\exists x A(x)$ is realised by (the Gödel number of) an ordered pair $(\ulcorner x \urcorner, a)$ if a realises $A(x)$; $A \supset B$ is realised by $\ulcorner \varphi \urcorner$, where φ is a recursive function, if whenever a realises A, $\varphi(a)$ realises B. Kleene's interpretation proves that HA has both the existence and disjunction properties, something that Gödel, as we saw, was not able to prove by the functional interpretation. Alongside the challenges posed by the computability proof, the failure of the proof of the existence property which, according to Gödel, defines constructivity for a logic, could also have contributed to Gödel's frustration.

Gödel's functional interpretation did not become known to the public until Kreisel gave a talk titled "Gödel's interpretation of Heyting's Arithmetic" at Cornell University in 1957. The month-long Cornell conference was a remarkable event with an impressive list of participants. Feferman and Feferman

(2004, 222) underline the importance of the conference to the field: "for the first time, logicians of every stripe came to grips with what their colleagues had been up to; the opportunity to have face-to-face contact and lengthy discussions with individuals who had previously been only disembodied names was exciting, and the general feeling of exhilaration was enhanced by a combination of appreciation and competition." The crowd divided into fractions of prominent logicians - such as Tarski, Church, Kleene, Rosser, and Quine – and their students, many of whom would become famous logicians themselves.

Kreisel was a "wild card," who, in Gödel's absence, presented his functional interpretation of arithmetic, a result that was entirely unexpected for the audience (Feferman and Feferman, 2004, 225–226). His own contribution was extending Gödel's functional system to analysis (Feferman, 1993, 33–34). In the same year, Kreisel also lectured on the functional interpretation of analysis in Amsterdam (Kreisel, 1959).

Kreisel's interest in functional interpretations seems to have stirred Gödel's own, or perhaps he felt like he needed to publish the result that Kreisel had developed further.[33] In 1958, Gödel's paper "Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes" was published in the journal *Dialectica*. There are some formal differences to the 1941 presentation. Gödel does not prove the results on the existence property in Σ or the consistency of the negation of PEM; as the sole application, we get the consistency proof for classical arithmetic. There is also a mention of the extension of the functional system to transfinite types. Most importantly, Gödel interprets functionals intensionally: instead of a defined notion of equality for higher types, equality is now primitive for all types. It follows that, as the cancelled passage in *PLI* suggests, a computability proof is no longer required. The notion of computability has to be considered as "immediately intelligible" (Gödel, 1958, 245).[34]

In a footnote, Gödel admits that the intensional interpretation leaves an element of vagueness similar to the proof interpretation (Gödel, 1958, 245, footnote 5):

> One may doubt whether we have a sufficiently clear idea of the content of this notion, but not that the axioms given below hold

[33] On the last unnumbered pages of *RG*, clearly written much later than most of the notebook series, we find Gödel's notes on Kreisel's proof of consistency of analysis by the functional interpretation. However, the analysis of Gödel's notes and the comparison with Kreisel's proof is beyond the scope of my thesis.

[34] Page numbering refers to the English translation in Gödel (1990).

> for it. The same apparently paradoxical situation also obtains for the notion, basic to intuitionistic logic, of a proof that is informally understood to be correct. As the considerations presented below and the intuitionistically interpreted theory of recursive functions and functionals show, these two notions are, within certain limits, interchangeable as primitive notions. If the notion of computable function is not to implicitly contain the notion of proof, we must see to it that it is immediately apparent from the chain of definitions that the operations can be performed, as is the case for all functions in the system **T** specified below.

But "immediately intelligible" and "immediately apparent" are still just as informal as the notion of an intuitionistic proof. In the next footnote, Gödel refers to Turing's informal concept of computability, writing that if the notion of a computable function were not already intelligible, Turing's definition would make no sense at all (ibid., footnote 6). This does not yet prove that Gödel's notion is *more constructive* than the intuitionistic notion of proof. Gödel did not, however, consider this issue further until he started to work on the 1970s reprint of the article.

A question left open in the previous section is whether Gödel seriously considered the no proof option in 1941. Van Atten (2015, 202–204) suggests that Gödel's mathematical struggles in 1941 led him to reconsider the extensional interpretation of functionals and the need for a formal computability proof. This was the start of a "shift towards the intensional" and the view adopted in the *Dialectica* article. Van Atten quotes a passage in *Max Phil* 4 (p. 198–199), written in August or September of 1941:

> Perhaps the reason why no progress is made in mathematics (and there are so many unsolved problems), is that one confines oneself to ext⟦ensions⟧ – thence also the feeling of disappointment in the case of many theories, e.g., propositional logic and formalisation altogether.[35]

There is a similar tone in the remark Gödel made to Kreisel in 1955, in connection with the functional interpretation, about "Aussichtslosigkeit, that is,

[35] Bem. (*Grundl*) Vielleicht kommt man in der Mat⟦hematik⟧ deswegen nicht weiter (und gibt es so viele ungelöste Probl.) weil man sich auf ext. beschränkt – daher auch das Gefühl der Enttäuschung bei manchen Theorien z.B. dem Aussagenkalkül und der Formalisierung überhaupt. ⟦The English translation is Van Atten's.⟧

hopelessness of doing anything decisive in foundations by means of mathematical logic" (Kreisel, 1987, 107).

According to Van Atten, the disappointment expressed in the passage from *Max Phil* ("e.g., propositional logic") could refer to Gödel's struggles with the interpretation of intuitionistic connectives. There were undoubtedly other sources of mathematical discouragement, as well; the idea of a consistency proof of analysis and using Brouwerian analysis to prove the independence of the Continuum Hypothesis were silently buried sometime in 1942. Indeed, at the end of the above note, Gödel has written "vgl. p. 215," where we find a list of "Fälle wo in der Mathematik vielleicht ein Rückgriff auf Intensionales nötig." On top of the list stands "Continuum problem and the Axiom of Choice." However, we find no mention of the functional system here.[36]

Kreisel also suggests that Gödel had a "total shift of emphasis" between 1941 and 1958 (Kreisel, 1987, 104). Feferman, on the other hand, comments that Gödel must have misremembered because one can see no such shift, "though the 1941 lecture mentions a few applications that are not contained in the 1958 Dialectica article" (Feferman, 1993, 33). He refers to Gödel's remark in a letter written on 16th May 1968 to Paul Bernays, where Gödel states that in the 1958 article "I placed no particular value on the philosophical matters; rather, I was mainly concerned with the mathematical result, whereas now it is the other way around" (Gödel, 2003a, 261).

I do not believe that Feferman is entirely correct here. In the light of what we have seen, it seems rather that Gödel misremembered, or at least exaggerated, in the second case. The remark quoted by Feferman is a footnote to a sentence which reads: "As I took up the manuscript in order to have it typed, I found in the philosophical introduction (i.e., the first $3\frac{1}{2}$ pages in *Dialectica*) much even in the original text [that was] presented in such an unsatisfying and fragmentary way that I considered numerous supplementary remarks and changes as absolutely necessary." This is true: there are plenty of problems left open both in the Princeton Lectures and the *Dialectica* paper. And not all of those problems were given answers in the 1972 version of the article, either. (As mentioned, Gödel could not understand why the validity

[36] A note in *Max Phil* 4 from around the same time states: "Das Auswahlaxiom ist auch ein Beispiel für ein Ax⟦iom⟧, welches dann evident ist, wenn man überhaupt nur gewisse Begriffe als sinnvoll anerkennt. In diesem Fall: unendliche gesetzlose Extens⟦ionen⟧" (p. 201). That is, because the extensional interpretation the choice function immediately validates the Axiom of Choice, in order to show the consistency of its negation, one needs to find an intensional interpretation.

of $A \supset (A\&A)$ requires the assumption of the existence of characteristic functions.) It is correct to say that he placed more importance on the philosophical issues in 1972, but this is not to say that the functional interpretation of 1941 and 1958 was simply a formal exercise.

One thing to keep in mind is that even later on, when Gödel accepted taking functionals as primitive, he was still interested in finding a formal proof. We know that he was not happy with William Howard's proof (Howard, 1970) which used transfinite ordinals, and when Howard stayed at IAS in 1972–1973, Gödel kept pressing him to find a simpler proof. Howard recalls:[37]

> In my final meeting with Gödel, in the summer of 1973, Gödel's last piece of advice to me was:
>
> "Now, about that problem. You should give it to a graduate student."
>
> I was taken aback. Was he telling me: "You cannot do something that a graduate student should be able to do." My reply was:
>
> "Well, I don't know. The problem is harder than it looks."
>
> He might have been teasing me a bit; but I think, mainly,
>
> (a) he remained fixed in his belief that, when looked at in the right way, the problem ought to have a simple solution.
>
> (b) it would be highly desirable to have such a solution.

Even though Gödel seems to argue in 1958 that the notion of a computable finite-typed functional is best left as primitive and requires no formal justification, he still believed that such a proof would have value, and he appears to have placed plenty of importance on it.

All of this underlines the fact that it is oversimplification to say that in 1941 or 1958, Gödel's only goals were mathematical, and in the 1970s, all of them were philosophical. Gödel was well known for his tendency to examine a variety of views, some of which he did not himself accept, just to see what they contain. Moreover, he was rarely dogmatic about a viewpoint, believing in a foundational pluralism where different viewpoints touch different aspects of mathematics. It could be that even though Gödel saw the informal concept of a computable functional as more fundamental than the formal one, he could

[37] Email message to Jan von Plato, 8th June, 2007. Reproduced with the permission of William Howard.

have still believed that a proof might be mathematically useful. After all, he had cancelled the no proof passage in the Princeton Lectures because he believed he could come up with a proof. Perhaps he would still subscribe to the general maxim suggested in the previous sections: if an informal notion can be made mathematically precise, then it should be made so.

Gödel's view in the 1940s is, in general, more philosophically nuanced – one cannot find strong statements like those in the 1930s on Platonism and intuitionistic logic – and starting to lean towards the pluralism that characterises his philosophical works. Interestingly, this view is more present in the notes written immediately after the Yale and the Princeton lectures than in the lecture notes themselves. It is as though the challenges had him reconsider his earlier views and made him more tolerant of those of others as well. However, we can admit that there was a genuine change in Gödel's views and, at the same time, admit that what led him to turn towards the intensional was his disappointment after several failed mathematical projects. The change did not make these mathematical ambitions meaningless, but rather, it offered another viewpoint to foundational work.

It is precisely on the philosophical front where the differences between the 1941 and 1958 versions of the functional interpretation are the clearest. There is a similar shift of emphasis between the 1930s lectures, which are framed around the Extended Hilbert Programme, and the Princeton and Yale lectures, which are presented as an investigation of intuitionistic logic. In his introduction to the *Dialectica* article, Troelstra notes that, according to Kreisel, the original goal was to obtain the existence property for HA, whereas in 1958, "Gödel presents his results as a contribution to a liberalized version of Hilbert's program" (Troelstra, 1990a, 219).

Between 1933 and 1958, Gödel has, in a sense, completed the full circle. In **Chapter 2**, I characterised Gödel's early viewpoint as a kind of formalism: in the "Present situation" lecture of 1933, he called for a reformation of Hilbert's programme of justifying classical mathematics constructively. He saw intuitionistic logic and its hidden vague elements as a poor choice for a foundational framework. In 1938, he suggested that a system based on finite-typed functionals could be used instead of Gentzen's transfinite induction to obtain a more constructive consistency proof. In 1941, however, Gödel presents the functional interpretation as a contribution to intuitionistic logic as opposed to the Hilbert Programme. His notes written around the same time make it clear that he thought little about the latter, whereas he studied intuitionistic mathematics intensely in 1941–1942. Finally, in 1958, we find ourselves again in the context of the Extended Programme. The last change is illustrated by the comparison of the title of the Yale lecture, "In what sense is intuitionistic logic constructive?", and that of the *Dialectica* paper, "On a hitherto unutilized extension of the finitary standpoint."

However, the way that Gödel approaches the Hilbert Programme in 1958 is quite different from the 1930s and his way of extending it is certainly not formalistic in the sense that the functional interpretation itself now contains informal and abstract components. Indeed, by this time, Gödel has given up the finitistic ideals. There are no criteria for strict constructivity such as those of the earlier lectures, and the surveyability criterion is officially dropped. Gödel states that the "specifically finitistic element" on the intuitive evidence or concreteness of objects has to be given up if one wishes to extend the Hilbert Programme (Gödel, 1990, 245). All extensions of the finitary viewpoint contain abstract, as opposed to concrete, elements. Moreover, informality and intensionality in the notion of a computable function are accepted.

It appears that the roots of this change lie indeed in the Yale and Princeton lectures of 1941, when it became clear that the justification of the functional interpretation as more constructive than the other alternatives was not straightforward. If the functional interpretation is to show that Heyting Arithmetic is constructive, then the "vast" notion of a functional needs to be shown at least better defined than that of an intuitionistic proof. Gödel never managed to obtain the computability proof that he wanted. Mathematical challenges and disappointment such as this eventually drew him from mathematics into philosophy.

Gödel did live to see the birth of a new field of research from his interpretation, starting from Kreisel and then Clifford Spector, who met Kreisel at the Cornell University summer school in 1957, was greatly inspired by his work, and later came to use Gödel's method to prove the consistency of analysis. Kreisel, who edited the paper for publication after Spector's untimely death in 1961, remarks that Spector "valued highly his discussions with P. Bernays and K. Gödel on the subject of the present paper" (Spector, 1962, 1, footnote 1), although Gödel notes in his postscript to the article that Spector had already obtained the main result before these conversations (Spector, 1962, 27). Around 1970, many new variants of the functional interpretation emerged (e.g., Shoenfield, 1967; Parsons, 1970; Diller and Nahm, 1974).

Despite the challenges, Gödel always considered the *Dialectica* paper as a significant accomplishment. In a sketch of a bibliography prepared for the seventh edition of "The consistency of the continuum hypothesis," he listed the 1958 article among his most important publications (040359.5). Under another list titled "Was ich publizieren könnte," found in the same folder, Gödel has written "Zirkelfreie Interpretation der Dial⟦ectica⟧ Arbeit." Perhaps he appreciated it, not in spite of, but precisely because of the challenges that had helped him to reshape and redefine his views over a span of two decades.

Chapter 5

Conclusion

The aim of my study was to investigate Gödel's three lectures on intuitionism and the constructive foundations of mathematics, published in 1995 in his *Collected Works*, against the background of his unpublished papers and personal notes. This study divided quite naturally into two distinct phases: the formalist outlook of the 1930s lectures and the study of intuitionistic logic in the Yale and Princeton lectures of 1941. Moreover, I discussed changes in Gödel's views in 1941, when he encountered several problems in the functional interpretation and was driven to expand his foundational views.

In addition to the posthumously published materials, my primary sources included the notes made for Gödel's first public presentation of his results on intuitionistic logic in Hans Hahn's logic seminar of Winter Term 1931–1932, the notes for the 1933 and 1941 lectures as well as a lecture course given in Princeton in 1941, related materials in the series *Arbeitshefte* and *Resultate Grundlagen*, remarks on intuitionism in the early *Maximen und Philosophie* notebooks, and certain sets of loose notes such as the *Questions and Remarks* from 1940–1941. The material that was written in shorthand – i.e., everything except the US lectures – was first reviewed carefully to isolate the passages related to intuitionism and constructive mathematics, which were then transcribed and translated.

5.1 Summary of the findings

In his first logical works of 1931–1932, Gödel already showed an interest in the interconnections between classical and intuitionistic logic, which he saw mostly from a negative perspective. His early readings consisted primarily of intuitionistic logic, and although he had seen Brouwer lecture in Vienna in 1928, it is likely that he had not read any of Brouwer's works at this point. The notes that Gödel made in preparation for Hans Hahn's logic seminar in late 1931 show that, like Hahn and Karl Menger, who were both likely influences behind his thoughts, he believed that intuitionistic logic is less different from classical logic than commonly thought. He based this belief on the fact that classical concepts can be interpreted intuitionistically. These ideas shaped his early understanding of, and his relatively negative approach to, intuitionism and its logic.

Gödel's first public criticism of intuitionism appears in the 1932 talk on the negative translation between intuitionistic and classical arithmetic. His interpretation of the result was that "the system of intuitionistic arithmetic and number theory is only apparently narrower than the classical one, and in truth contains it, albeit with a somewhat deviant interpretation" (Gödel, 1933c, 295). This is the only perspective to intuitionism that Gödel took in the 1930s.

The framework of the 1930s foundational lectures is the post-incompleteness Hilbert Programme and its possible extensions. Gödel's critique of intuitionistic proofs as vague and Platonism as built on questionable metaphysical foundations represents a kind of formalism not too far away from Hilbert's views. This view is especially prominent in the draft version of the 1933 lecture, "The present situation in the foundations of mathematics." Gödel places plenty of importance to a consistency proof for classical arithmetic, and eventually analysis, by means that are ideally constructive. What ideally constructive means for him is expressed in the following "strict criteria" of constructivity:

1. All primitive relations should be decidable and all primitive functions calculable.

2. Existential quantifiers should be used only in the sense of an abbreviation of an instance and negated universal quantifiers only as an abbreviation of a counterexample.

3. The objects of the theory should be surveyable.

These criteria, Gödel states, define Hilbert's finite mathematics. He argues that neither intuitionistic arithmetic nor Gentzen's consistency proof by transfinite induction satisfies the criteria. In 1938, he claims that there is a way to extend Hilbertian finitism without compromising on any of the criteria: a system based on finite-typed functionals. However, Gödel does not give any details on this system or how the consistency proof should be carried out.

It is evident that intuitionistic logic does not satisfy criterion 2, although one feels compelled to ask why this matters, given that the intuitionistic quantifiers are not interdefinable: one does not get existential claims out of negated universal statements. This was a point misunderstood by many in the 1920s, even Weyl claiming that negated universal statements are not intuitionistically meaningful. However, Gödel knew intuitionistic logic and its properties well. The reason why he disregards the issue of the independence of the two quantifiers is that, according to him, the negative translation shows that they correspond precisely to *classical* existence statements. Therefore the concept of absurdity does not do the job it is supposed to.

Another critical point was the vagueness or absoluteness in the intuitionistic concept of proof, which Gödel said already in 1931 to denote a proof by any means imaginable, independent of any formal system. This is entirely correct. However, because the totality of those proofs is certainly not surveyable, intuitionistic logic in its proof interpretation does not satisfy criterion 3 either.

Gödel's criteria of constructivity as well as his criticism of intuitionism or intuitionistic logic – two concepts that he used more or less interchangeably in his early works – is entirely in line with Hilbertian formalism. Gödel not only believed that classical mathematics must be proven consistent to justify its Platonistic metaphysics, but he believed that formalisation is a necessary component of this justification and that meaning should play little role in it, at least outside the finitary part of mathematics. In general, as his interpretation of the negative translation shows, Gödel was not particularly sensitive to issues of meaning. Accordingly, concepts such as "vague" or "surveyable" in connection with intuitionistic proofs are not discussed in depth. In the 1930s, "vague" meant both formally imprecise and unsurveyable; in the 1941 lectures, he used the word only in the first sense.

The framework in which Gödel presents the functional interpretation in 1941 is entirely different from the formalistic standpoint of the 1930s. The title of the Yale lecture was "In what sense is intuitionistic logic constructive?", which summarises Gödel's agenda in the Princeton and the Yale lectures. The Princeton notes and early sketches in the *Arbeitshefte* make it clear that the functional interpretation of Heyting Arithmetic is presented as, first of all, a precise interpretation of intuitionistic connectives (as opposed to the vague proof interpretation), and secondly, as a proof of its constructivity, i.e., a proof that it has the existence property. The strict criteria of constructivity now include only 1 and 2, showing a clear departure from Hilbertian finitism. A system that satisfies condition 2 has the existence property by default, making its constructivity wholly transparent. Gödel also mentions the relative consistency proof of classical arithmetic, although this is presented as a corollary rather than the main result.

Strictly speaking, the functional interpretation does not quite fulfil Gödel's promises. First of all, whereas it is certainly more precise than the proof interpretation in the sense of being formal, not informal, it is not a faithful interpretation of Heyting Arithmetic, because it satisfies Markov's Principle and the Independence of Premise, which are not intuitionistically acceptable. Moreover, it does not prove the existence property for Heyting Arithmetic but only for its extension to higher types, HA^{ω}. It is not even certain whether it is more constructive than Heyting Arithmetic in the first place, as it is not apparent that higher-type functionals can be shown computable *in a strictly constructive manner*. Gödel's first computability proof, written down in *Resultate Grundlagen*, uses ordinary induction, although he soon noticed that this undermines the whole point of the functional interpretation as an interpretation of Heyting Arithmetic.

At the time of the Princeton Lectures, Gödel considered two other options for the proof: either take computability as primitive or use transfinite induction to show that every functional can be reduced to a number by a sequence of substitutions. In the end, he rejected the first option and pursued the second. Despite several attempts recorded in the mathematical workbooks, he could not come up with a satisfactory proof by the end of the lecture course, and soon after, he gave up trying. For the next year or so, Gödel studied intuitionistic mathematics intensely, his apparent purpose being, first of all, to extend the functional interpretation to analysis, and most importantly, to prove the independence of the Continuum Hypothesis. Neither of these efforts seems to have given any definitive results. After a dry year of failed attempts, Gödel gave up and shifted his focus to philosophy.

Most of Gödel's philosophically inclined remarks on intuitionism date from this year of struggles, from late spring of 1941 to summer of 1942. We see from his notes that he was now open to consider other aspects of intuitionism than its classical interpretations. He came to appreciate the greater expressive power of intuitionistic logic, a trade-off for less deductive power. Gödel also acknowledged the distinction between negative and positive interpretations of intuitionism, the first arising from the classical viewpoint and the second from intuitionism itself. In his early works, he was solely focused on the first sense, almost as if he failed to see the second sense at all. The notes of 1941, discussed in **Chapter 4**, make it clear that he soon came to recognise the fruitfulness of the second standpoint, although he never worked seriously in the philosophy of intuitionism.

5.2 Philosophical undercurrents

Several themes run through Gödel's early works on intuitionism from the first logical works to the 1941 lectures and notes, two of them particularly prominent. First is Gödel's desire to reinterpret intuitionism in a variety of classical frameworks and his interest in the relationship between classical and intuitionistic logic. The second is his aim of clarifying or making precise concepts of constructive mathematics in order to arrive at *results* concerning constructive mathematics. The first theme is already present in the early writings, and still in the 1940s, Gödel believed that the study of intuitionistic logic from the classical viewpoint is valuable. Around early 1941, he writes in the *Questions and Remarks*:

> 56. A fruitful standpoint for grounding intuitionistic mathematics: reduce the problems to ones that make sense also in classical mathematics.[1]

There is a connection between the themes, as Gödel was attempting to make intuitionistic logic more precise by classical means with the modal interpretation and the Σ-translation. Later on, he acknowledged that also non-classical and informal interpretations could be fruitful, or in some cases, even necessary.

In the introduction to this thesis, I noted that Gödel's three lectures on constructive foundations were first seen as surprising, given his later statements about being a lifelong mathematical Platonist. It could be replied that he did tend towards pluralism, exploring a wide variety of different philosophical

[1] 56. Fruchtbarer Gesichtspunkt für Begründung der intuit⟦ionistischen⟧ Mathematik: die Probleme auf solche reduzieren, welche auch in gewöhnlicher Mathematik einen Sinn haben.

and mathematical views, in nearly all of which he saw some value. Kreisel (1980, 209) recalls:

> [In] his publications Gödel used traditional terminology, for example, about *conflicting* views of 'realist' or 'idealist' philosophies. In conversation, at least with me, he was ready to treat them more like different *branches* of the subject, the former concentrating on the things considered, the latter on processes of acquiring knowledge about these objects or about the processes.

Kreisel also refers to Gödel's later statements that his focus on the realist concepts rejected by the constructivists helped him solve the completeness question and come up with the incompleteness results. This paints a picture of Gödel moving between different traditions, seeking out fruitful concepts for analysis and application.

However, this is Gödel several decades after the early works. Arguably, there is less variance in his early works, which are mostly mathematical in nature. Perhaps the surprising element is that the only foundational considerations we find in the 1930s lectures concern mainly epistemic questions, and there are no remarks whatsoever related to mathematical realism. Moreover, Gödel is sympathetic towards Hilbert's goals: in the early formulations of the incompleteness theorem, he even states that it does not apply to Hilbert's finitistic mathematics. It could be argued that he said this because he did not want to cross Hilbert. However, Hilbert was no longer mathematically active in 1931, and in late 1933, when Gödel supported Hilbert's formalism to give an epistemic justification to classical mathematics "hopelessly entangled" with Platonism, the incompleteness results were universally accepted. The first positive remark that I have been able to find in my research is from ca. April 1941, when Gödel writes towards the end of *Max Phil* 3 (p. 145) that impredicative definitions as well as non-instantiated existential statements are a "proof for Platonism." Apart from this one line, I have not found any discussion of mathematical Platonism in Gödel's notes of 1941 or early 1942.

Whereas Gödel's self-proclaimed mathematical Platonism might be well-hidden in his early works, the idea of concept-clarification as a path to solving problems is an aspect of Plato's thought that Gödel certainly shared. In Plato's dialogues, the source of our confusion often lies in concepts that are poorly defined; they are either ambiguous, deficient, or contain unfounded and irrational beliefs. By revealing the faulty thinking that underlies a certain definition, one can start to build a more accurate and precise definition.

For Plato, this way of obtaining knowledge is coupled with conceptual realism. Through the clarification of concepts, one can reach the underlying idea that is the true extension of the concept, if that idea exists. However, concept-clarifying as a method of problem-solving is not incompatible with the intuitionistic viewpoint, either. As Sundholm and Van Atten (2008, 71) put it, Platonism and Brouwerian intuitionism share an ontologically descriptive epistemology: the Platonist describes ideas, the intuitionist describes mental constructions. A fundamental problem for Platonism is the connection between ideas and our inquiry of them. Plato admitted that this is a problem that one simply has to live with: perhaps some of our concepts are non-concepts in the sense that they end up misdescribing an idea. A fundamental problem for Brouwer, on the other hand, is the connection between mental constructions and our linguistic description of them. Brouwer did not see this as a problem, though, but only as a necessary evil: language was never made for communication, so one should expect some level of indeterminacy on the linguistic front.

As I said in **Chapter 2**, I am not claiming that Gödel was against Platonism *as a metaphysics*. What is clear, though, is that metaphysics did not much interest Gödel in the 1930s or even 1941: his foundational studies were all epistemically motivated, and he saw the project of an epistemically responsible treatment of mathematics as highly important. He simply did not occupy himself with the question about the objects of mathematical inquiry. He certainly did not share Brouwer's pessimism about language – few modern intuitionists do, at least to the degree that Brouwer did – but he believed that we could make our intuitive concepts clear enough to solve problems about them. This is a belief Gödel maintained even after he became explicit about his mathematical realism, and it eventually led him to seek an answer to the question of intuition of essences in Husserl.

Gödel's first attempts to understand intuitionistic logic involved treating it in the framework of classical logic. In the case of the negative translation, he seemed almost disappointed that the task turned out so easy. Of course, it would be expected that classical statements can be expressed in intuitionistic logic, which is capable of handling finer distinctions. However, Gödel turned the translation upside down, in a sense, and claimed that the fact that negated universal statements correspond to classical existence statements proves that the concept of absurdity is not as helpful as it claims to be. But this is misleading, because classical existence, from the intuitionistic point of view, is an ambiguous concept: one cannot distinguish between its two senses that are expressed in intuitionistic logic by $\exists x$ and $\neg\forall x\neg$. By the negative translation, then, some existential claims are weakened because any existential statement will be turned into a negated universal statement, whether an instance of it could be proven or not.

Another formal endeavour to interpret intuitionistic logic is the modal translation. In fact, this is the form of a "proof interpretation" of intuitionistic logic that Gödel discussed in the 1938 Zilsel lecture. In the Hahn notes discussed in **Chapter 2**, he notes that this is a more natural interpretation of intuitionistic logic in a classical framework that retains the meaning of intuitionistic connectives. Although the negative translation is not discussed in the notes, the modal interpretation is arguably less deviant than the negative translation. The modal interpretation is still not entirely natural: the axiom that defines the system **S4**, $Bp \supset BBp$ is not necessarily meaningful from the intuitionistic viewpoint, because the statements "it is provable that it is provable" or "it is assertible that it is assertible" are themselves no longer assertions (Heyting, 1931, 113). There are additional issues with quantifiers, where the semantics of quantified **S4** become questionable: the completeness proof needs to assume Markov's Principle, which is not intuitionistically acceptable.

Whether Gödel put much importance on these interpretations as showing something about the nature of intuitionism is not certain. They could be seen only as formal results, although Gödel's remarks in the Hahn notes, as well as the critique of intuitionism in the first presentation of the negative translation, suggest that this is incorrect. What can be said, though, is that Gödel did not see intuitionism as an attractive solution for constructive foundations because of its vague elements and the doubts that arise from the negative translation. It should also be kept in mind that he never agreed with Brouwer's goal of restricting mathematics only to its constructive part.

As mentioned, the 1941 version of the functional interpretation is presented as an interpretation of Heyting Arithmetic and as an attempt to make it more precise by replacing the vague concept of proof by the sharp concept of a functional. This, too, is precision through formalisation, which does not necessarily fit into the Brouwerian idea of logic-free mathematics.

Does it matter that all of these interpretations are unnatural or unintended in the sense that they are not in line with the philosophical foundation for intuitionism? What the natural proof interpretation does is to give meaning, not interpret, in the sense of giving a mathematical model of some kind. Sundholm (1983, 159) writes that in the case of the latter type of interpretations,

> it makes perfect sense to inquire as to the truth of the mathematical proposition
>
> $$A \leftrightarrow A^*$$
>
> There is no attempt in such interpretations to explain the meaning of the propositions but instead one uses the propositions. If we were to ask the same question for the Heyting-Kolmogoroff explanation the result is a piece of nonsense like
>
> $$A \leftrightarrow \text{ the explanation of } A$$
>
> This is nonsense because the propositional connective $\leftrightarrow$ needs to be filled with propositions and the right-hand side is not a proposition (in use), but a meaning-explanation of the proposition.

Here A^* denotes the interpretation of A by functionals, modal logic, realisability, etc.

Perhaps one is asking a wrong question, then, in considerations about the adequacy of such interpretations. In several places, especially in the early works, Gödel states that the deviancy of a translation is of no issue, as long as one has a model (i.e., an interpretation in the latter sense). However, he also uses such interpretations to criticise intuitionistic logic for possibly having classical elements, although such models cannot show this. There is a confusion between the formal and the meaningful, and between making precise in the sense of formalisation and in the sense of *clarifying* concepts or finding the right concept to match a certain idea or construction.

On the other hand, Gödel's notes of 1941, discussed in **section 4.2** above, suggest other ways to approach questions of meaning. Tieszen (1995) identifies three directions in the philosophical basis of intuitionistic mathematics. The first, exemplified by Heyting and Martin-Löf, stresses the introspective aspect of intuition in mathematical construction. The mathematician's mind is the source of the different constructions of mathematical objects, proofs, and so on: these are identified with the acts of intuition, judgement, and construction. The second direction, adopted to Troelstra and Van Dalen, drops the purely introspective aspect as practically inaccessible and uses Kreisel's idea of "informal rigour" instead. Informal rigour is applied when we carefully analyse our intuitive concepts to arrive at mathematically valuable conclusions. The third view, represented by Dummett, adopts the doctrine of "meaning is use," adding both a linguistic and a social element to the foundations of intuitionism.

Gödel's remarks on the relationship between intuitionism and the analysis of concepts of natural language fit well into the second direction. Kreisel characterises informal rigour as the "old-fashioned idea" that "one obtains rules and definitions by analyzing intuitive notions and putting down their properties" (Kreisel, 1967, 138). The goal of informal rigour is to "make this analysis as precise as possible [...] in particular to eliminate doubtful properties of the intuitive notions when drawing conclusions about them" as well as to extend the analysis to solve mathematical problems (ibid., 138–139). This idea is present in Gödel's later work, a shift that, according to Van Atten (2015, 204), "constitutes a remarkable *rapprochement* with intuitionism." Kreisel tells that Gödel encouraged him to develop the idea of informal rigour, although he warned him that mathematicians would probably not be entirely fond of the idea (ibid., 205).

In his 1940s notes, Gödel does not use the word "informal" – *informell* would not, in any case, be the proper term here – or *intuitiv*, but he instead talks about *Wortsprache*, natural language, or *natürliche Vernunft*, natural reasoning. These terms seem to refer to both informality, in the sense of referring to concepts of natural language and psychology, and *non-formality*, in the sense of going beyond formal systems.

The idea of non-formality is tightly connected with Gödel's idea of *absolute* notions. As with Turing's notion of computability, Gödel believed that the notions of definability and provability could be given such absolute characterisations (see esp. Gödel, 1946).[2] Turing's notion of computability does not merely happen to coincide with all the previous definitions of computability (Gandy, 1988), but, as Kennedy (2014, 117) notes, it is also "informal yet entirely convincing," and Gödel appreciated particularly this formalism-independent as-

[2] Gödel also discusses absolute concepts in several places in both *Max Phil* 3–4 and *Questions and Remarks*, both written as early as 1940–1941, as well as the *Arbeitshefte* in 1941–1942.

pect of it. "Informal," as used in the *Dialectica* paper, is the English translation of the German word *inhaltlich*, which could also be translated as contentful. Now we get a distinction, one that Gödel overlooked for a long time, between the formal and the meaningful.

The concept of informal rigour as an interpretation of intuitionism is also present in Gödel's 1941 notes, and he saw this way of analysing concepts as valuable in its own right. It seems, then, that we can indeed trace the beginning of the shift, mentioned by Van Atten, back to the time of the Princeton Lectures, when Gödel started to reconsider his previous views.

Gödel also believed, though, that true intuitionism contains an irreducibly psychological element, and that an analysis of psychological concepts could be useful in solving foundational questions about intuitionism (see **section 4.2**). As mentioned, in several places in the early *Max Phil* notebooks and elsewhere, he identifies intuitionism as psychologism and meanings as psychological in Brouwer's work. This is in line with what he recounts in 1975 to Sue Toledo (2011, 206):

> Intuitionism perfectly meaningful.
>
> In class. math. hunt for axioms using extra-mathematical ideas.
>
> But axioms are about mathematical objects.
>
> In intuitionism isn't. Statements involve extra-math. element. Namely, the mind of the mathematician & his ego.
> ratl limits
>
> Statement[s] of int[uitionism] are psycholog. statements, but not of empirical psy[chology] – essential a priori psychology/not formal

But, he adds, "where classical mathematics seems to have found its primitive elements, in intuitionistic math. working with ideas that haven't been analyzed (e.g. concept of proof)" (ibid., 207). "Haven't been analysed" suggests that Gödel still believed that they could be analysed, apparently through an analysis of the related psychological concepts. In this sense, he also agreed with the intrinsic reading, a viewpoint not necessarily incompatible with the second: where the source of our intuitions might be, in Gödel's terms, psychological, the role of informal rigour is to "unfold and clarify our knowledge of these cognitive processes or structures" (Tieszen, 1995, 588–589).

To conclude, whereas Gödel's works mainly concentrated on unnatural interpretations, he came to understand the difference between the two and had some ideas as to the natural interpretation and the philosophical foundation of intuitionism. The goal in both of these endeavours was to clarify unclear intuitive concepts and make them precise, something that, in Gödel's opinion, underlies intuitionistic mathematics and a large part of mathematical inquiry in general. However, they involve different senses of making precise: unintended interpretations formalise intuitionistic principles in other known frameworks, whereas intended interpretations attempt to explain the meaning of those principles. It was probably Gödel whose works gave birth to a field of research on the first front. In modern proof theory and philosophy of mathematics, both types of interpretations are studied, and both are fruitful in examining different aspects of intuitionism and its logic.

5.3 Limitations of the study

The most obvious limitations of my study are related to the data used, i.e., Gödel's shorthand notes. It is always challenging to work with previously unexamined material, and the effort of transcription and the vastness of the notes limited the amount of material I was able to go through. Especially in the case of less formal notes, such as *Max Phil*, I was forced to choose my material based on keywords to figure out which passages were possibly relevant before starting the transcription work. I also had to restrict myself to notes written in 1941 or early 1942, as this was the period most relevant to my study, leaving out most of the philosophical notebooks.

The initial scoping was very thorough, and I repeated the process several times during the writing of this thesis, but it cannot be ruled out that something has been missed. Moreover, entire sets of notes could have been left out, as there are thousands of pages of loose notes in addition to the notebook series, only a part of which I examined. I chose to exclude most from the beginning because these materials were more fragmentary and difficult to date.

Another challenge emerged from the fact that the material in the *Arbeitshefte* was far more extensive than I had expected, and much of it required mathematical understanding outside my field of expertise. The notes on the computability proof by transfinite ordinals, to give an example, turned out more difficult to interpret than I initially thought, and I found even more notes related to the proof as I was writing the dissertation. I soon quit trying to reconstruct the notes, because there were too many, the style was fragmentary, and because I felt that my formal skills would not be adequate for this task.

Another example is Gödel's idea of using Brouwerian analysis to build a counterexample for the Continuum Hypothesis, of which there were dozens of pages of materials. As I have only rudimentary expertise in set theory, I chose not to study these materials as a part of my dissertation, even though they could certainly tell us more about Gödel's motivations and goals for exploring intuitionism. I do not believe that this devalues any of the conclusions I have drawn in this thesis, although a study of the notes that I have excluded might elucidate aspects of Gödel's thought that I have not noticed. My primary aim, however, was to analyse Gödel's views on intuitionism in the three lectures in light of his notes, and I believe my sources were adequate for this task.

5.4 Suggestions for future research

There is still plenty of work to be done with Gödel's papers, only a fragment of which has been published so far. Likewise, there is a number of questions left open by my study, which could provide possible directions for further research.

As mentioned, although it is likely that Gödel did not come up with a computability proof for functionals in Σ by the method he described in the Princeton course, the notes in the *Arbeitshefte* could, if carefully reconstructed, provide us more information about his approach, and how it compares to the later computability proofs. A good place to start would be *Arbeitshefte* 8–10. In *AH* 8, we find the following description (p. 30–31), likely written in early spring of 1941:

> Programme goals:
>
> 1. Define a well-ordering between all expressions primitive-recursively and prove that the order of an expression decreases with every reduction.
> 2. Assign an ordinal number for each expression (from a certain field of constructed ordinal numbers) and so that the symbol is assigned primitive-recursively and ⟦add.: primitive recursive⟧ constructive proof that the ordinal number decreases with reduction.

3. For expressions without **L** through the ordinary computation operations, estimate how many reduction steps are needed and then generalise to **L**.[3]

L refers to the scheme that Gödel uses to denote primitive recursion, defined by $\mathbf{L}(n,k,x) = \underbrace{k\ldots k}_{n \text{ times}}(x)$. It is clear that the "programme" describes the proof by transfinite ordinals referred to in *PLI*.

Underneath there is another list:

Method

A. Direct method by examining the expressions of other types and then continuing (where eventually the next well-ordering will be defined by reduction on the previous one) ⟦add.: and a general theorem for expressions of the next type⟧.

B. By assigning branching schemes (*Verzweigungsschemata*) (resp. ordinal numbers) to the expressions and in such a way that

1. To the expressions of higher type functions of higher type higher types of branching schemes resp. ordinal numbers
2. ⟦To the expressions of higher type⟧ also branching schemes (ordinal numbers) and to the operation of "application" a certain function of ordinal numbers (branching schemes) ⟦add.: Probably the ordinal number of the branching scheme which one obtains, when by a complete argument series⟧

I. Assignment of the exactly correct ordinal number (branching scheme)

II. Assignment of an ordinal number $\geq$ the correct one

---

[3] Programme Ziele:

1. Eine Wohlordnung zwischen allen Ausdrücke in primitiv rekursiver Weise definieren und beweisen, dass die Ordnung eines Ausdrucks durch jede Reduktion verkleinert wird.
2. Jeder Ausdruck eine Ordinalzahl zuordnen (aus einem gewissen Feld von konstruierten Ordinalzahlen), und zwar so, dass das Symbol in primitiv rekursiver Weise zugeordnet wird und ⟦add.: prim.⟧ konstruktiver Beweis, dass die Ordinalzahl verkleinert wird bei Reduktion.
3. Für Ausdrücke ohne **L** mittels der gewöhnlichen Rechnenoperationen abschätzen, wie viele Reduktionsschritte nötig sind und dann verallgemeinern auf **L**.

C. By finitising the transfinite proof of existence of a reduction via Herbrand's consistency proof[4]

The methods **A**, **B**, and **C** correspond to steps 1, 2, and 3. Apparently, each expression can be identified with a well-ordering of previous expressions; with each reduction, it should be found lower down the ordering. *Verzweigungsschemata* refers to Gentzen's 1935 proof of consistency, which Gödel originally criticised for assuming a form of the Bar Theorem (Kreisel, 1987, 173–175). This inference occurs in connection with the reduction of a universal statement $\forall xA$. In order for $\forall xA$ to be correct, any arbitrary instance $A[x/m]$ must be correct. Now assume that A itself is a quantified statement, say, $\forall yB$. Then, given an arbitrary n, $\forall yB[x/n]$ is correct whenever for an arbitrary n, $B[x/m][y/n]$ is correct. But because A can have any finite number of quantifiers in it, B can itself be a universal statement, and so on. Gentzen's lemma states that any path, defined by arbitrary choices for substitutions in this long reduction tree, will eventually terminate in a quantifier-free formula whose correctness can be checked.

Given that Gödel's early notes on the functional interpretation are titled "Gentzen," it would be interesting to know more about the relationship between Gödel's functional system and Gentzen's 1935 proof. There are also notes on transfinite functionals, which are probably related to Gödel's goal of proving the consistency of analysis by extending the functional interpretation.

Another direction, also mentioned above, would be to collect Gödel's notes on intuitionistic mathematics in connection with the attempted independence

---

[4] Methode

A. Direkte Methode durch Untersuchen von Ausdrücke anderer Typen und dann fortsetzen (wobei eventuell die nächste Wohlordnung durch Induktion nach der vorhergehenden definiert wird) ⟦add.: und ein allgemeines Theorem für Ausdrücke zum nächsten "Typus"⟧

B. Durch Zuordnung von Verzweigungsschemata (bzw. Ordinalzahlen) zu den Ausdrücke, und zwar

1. Zu den Ausdrücke höheren Typus Funktionen höheren Typus von Verzweigungsschemata bzw. Ordinalzahlen
2. — — auch Verzweigungsschemata (Ordinalzahlen) und zur Operation der "Anwendung" eine gewisse Funktion von Ordinalzahlen (Verzweigungsschemata) ⟦Wahrscheinlich die Ordinalzahl desjenigen Verzweigungsschemas welches man erhält, wenn durch eine vollständige Argumentreihe⟧

I. Zuordnung der genau richtigen Ordinalzahl (Verzweigungsschema)

II. Zuordnung einer Ordinalzahl $\geq$ der Richtigen

C. Durch Finitisieren des transfiniten Beweises für Existenz der Reduktion, mittels des Herbrandschen Widerspruchsfreiheitsbeweises.

proof for Continuum Hypothesis. Although here, too, it is highly likely that Gödel did not obtain a proof, his notes in the *Arbeitshefte* might provide an idea of the method he wanted to use. Filed in the loose notes (060103,060527), we find notes related to Cohen's proof of 1963 as well as Gödel's "own method." As mentioned, the Editorial Project transcribed some of the notes but could not reconstruct the proof (Dawson and Dawson, 2005, 150). Perhaps this riddle could be solved by a closer examination of the earlier notes.[5]

Finally, in my study, I analysed Gödel's notes from a relatively short period of time, most dating from 1931–1933 and 1940–1942.[6] Gödel's later notes for the Dialectica paper and its revised version have already been studied in some detail (see Van Atten, 2015), but one would expect to find more materials related to intuitionism and the functional interpretation elsewhere in the loose notes.

---

[5] Dawson and Dawson (2005, 155) write that

> The [Cohen folder] was of immediate interest to set theorists, since, together with some passages from Arbeitshefte 14-15, it appears to contain Gödel's notes on his reputed proofs of the independence of the axioms of constructibility and choice in the framework of type theory. Early on it was transcribed in full, and as noted in Feferman's article "The Gödel Editorial Project: A synopsis", the transcript was sent for examination to several eminent set theorists, including Robert Solovay and A. D. (Tony) Martin. But none of those who have studied the notes have been able to develop a clear picture of what Gödel was thinking.

However, it is not mentioned whether the attempted independence proof for CH was examined. Moreover, some of the earlier loose notes, such as *Questions and Remarks*, contain related notes.

[6] The reason for the gap is that one can find fewer notes from 1934–1939, when Gödel made an intense detour to physics in 1935–1936 (his notes on the topic are to be published in Lethen and Passon (2020)), and then, after a short break, focused mostly on set theory for some time.

Bibliography

van Atten, M. (2009). On the hypothetical judgement in the history of intuitionistic logic. In Glymour, C., Wei, W., and Westerståhl, D., editors, *Logic, Methodology, and Philosophy of Science XIII: Proceedings of the 2007 International Congress in Beijing*, pages 122–136. King's College Publications.

van Atten, M. (2015). Gödel and intuitionism. In *Essays on Gödel's Reception of Leibniz, Husserl, and Brouwer*, pages 189–234. Springer-Verlag.

van Atten, M. (2017). The development of intuitionistic logic. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, winter 2017 edition. `https://plato.stanford.edu/archives/win2017/entries/intuitionistic-logic-development/`.

van Atten, M. (2018). Predicativity and parametric polymorphism of Brouwerian implication. `https://halshs.archives-ouvertes.fr/halshs-01619110v2`.

van Atten, M. and Kennedy, J. (2009). Gödel's logic. In Gabbay, D., editor, *Logic from Russell to Church*, pages 449–509. Elsevier.

Avigad, J. and Feferman, S. (1998). Gödel's functional interpretation. In Buss, S. R., editor, *Handbook of Proof Theory*, pages 337–405. Elsevier.

Barzin, M. and Errera, A. (1927). Sur la logique de M. Brouwer. *Académie Royale de Belgique, Bulletin de la classe des sciences*, 13:56–71.

Becker, O. (1927). Mathematische Existenz: Zur Logik und Ontologie mathematischer Phänomene. *Jahrbuch für Philosophie und phänomenologische Forschung*, 8:441–809.

Bernays, P. (1922). Über Hilberts Gedanken zur Grundlegung der Arithmetik. *Jahresbericht der Deutschen Mathematiker Vereinigung*, 31:214–222.

Bernays, P. (1930). Die Philosophie der Mathematik und die Hilbertsche Beweistheorie. *Blätter für Deutsche Philosophie*, 4:326–367.

Bernays, P. (1935a). Quelques points essentielles de la métamathématique. *L'Enseignement Mathématique*, 34:70–94.

Bernays, P. (1935b). Sur le platonisme dans les mathématiques. *L'Enseignement Mathematique*, 34:52–69.

Bernays, P. (1941). Sur les questions méthodologiques actuelles de la théorie hilbertienne de la démonstration. In Gonseth, F., editor, *Les Entretiens de Zürich sur le Fondements et la Méthode des Sciences Mathématiques, 6-9 Décembre 1938*, pages 144–152. Leemann.

Borel, É. (1908). Les "paradoxes" de la théorie des ensembles. *Annales scientifiques de l'É.N.S., 3e série*, 25:443–448.

Brouwer, L. E. J. (1905). Leven, Kunst en Mystiek (Life, Art and Mysticism). In Brouwer 1975, 45–53.

Brouwer, L. E. J. (1907). *Over de Grondslagen der Wiskunde*. Maas & Van Suchtelen.

Brouwer, L. E. J. (1908). De onbetrouwbaarkeit der logische principes. *Tijdschrift voor wijsbegeerte*, 2:152–158.

Brouwer, L. E. J. (1913). Intuitionism and formalism. *Bulletin of the American Mathematical Society*, 20:81–96.

Brouwer, L. E. J. (1924a). Intuitionistische Zerlegung mathematischer Grundbegriffe. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 33:251–256.

Brouwer, L. E. J. (1924b). Über die Bedeutung des Satzes vom ausgeschlossenen Dritten in der Mathematik, insbesondere in der Funktionentheorie. *Journal für die reine und angewandte Mathematik*, 154:1–7.

Brouwer, L. E. J. (1928). Intuitionistische Betrachtungen über den Formalismus. *Koninklijke Nederlandse Akademie van Wetenschappen, Proceedings of the Section Sciences*, 31:374–379.

Brouwer, L. E. J. (1929). Mathematik, Wissenschaft und Sprache. *Monatshefte für Mathematik und Physik*, 36:153–164.

Brouwer, L. E. J. (1930). Die Struktur des Kontinuums. In Brouwer 1975, 429–440.

Brouwer, L. E. J. (1949). Consciousness, philosophy and mathematics. In Beth, E. W., Pos, H. J., and Hollak, H. J. A., editors, *Library of the Tenth International Congress in Philosophy, August 1948*, pages 1235–1249. North-Holland.

Brouwer, L. E. J. (1975). *Collected Works 1: Philosophy and foundations of mathematics.* Edited by A. Heyting. North-Holland.

Crocco, G. and Engelen, E.-M. (2016). Kurt Gödel's *Philosophical Remarks (Max Phil)*. In Crocco, G. and Engelen, E.-M., editors, *Kurt Gödel: Philosopher-Scientist*, pages 33–54. Presses Universitaires de Provence.

Crosilla, L. (2019). Set theory: Constructive and intuitionistic ZF. In Zalta, E. N., editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2019 edition. `https://plato.stanford.edu/archives/spr2019/entries/set-theory-constructive/`.

van Dalen, D. (1995). Hermann Weyl's intuitionistic mathematics. *Bulletin of Symbolic Logic*, 1(2):145–169.

van Dalen, D. (1999). The role of language and logic in Brouwer's work. In Orlovska, E., editor, *Logic at Work*, pages 3–14. Springer.

van Dalen, D. (2004). Kolmogorov and Brouwer on constructive implication and the Ex Falso rule. *Russian Mathematical Surveys*, 59:247–257.

van Dalen, D. (2005). *Mystic, Geometer, and Intuitionist: The Life of L. E. J. Brouwer, Volume 2: Hope and Disillusion*. Oxford University Press.

van Dalen, D. (2008). Another look at Brouwer's dissertation. In van Atten, M., Boldini, P., Bourdeau, M., and Heinzmann, G., editors, *One Hundred Years of Intuitionism (1907–2007): The Cerisy Conference*, pages 3–20. Birkhäuser Basel.

Davis, M. (2005). What did Gödel believe and when did he believe it? *The Bulletin of Symbolic Logic*, 11(2):194–206.

Dawson, J. W. (1997). *Logical Dilemmas: The life and work of Kurt Gödel*. A. K. Peters.

Dawson, J. W. and Dawson, C. A. (2005). Future tasks for Gödel scholars. *The Bulletin of Symbolic Logic*, 11:150–171.

Detlefsen, M. (1998). Constructive existence claims. In Schirn, M., editor, *The Philosophy of Mathematics Today*, pages 307–335. Clarendon Press.

Diller, J. and Nahm, W. (1974). Eine Variante zur Dialectica-Interpretation der Heyting-Arithmetik endlicher Typen. *Archiv für mathematische Logik und Grundlagenforschung*, 16:49–66.

Dummett, M. A. E. (2000). *Elements of Intuitionism*. Oxford University Press.

Feferman, A. B. and Feferman, S. (2004). *Alfred Tarski: Life and Logic*. Cambridge: Cambridge University Press.

Feferman, S. (1993). Gödel's Dialectica interpretation and its two-way stretch. In Gottlob, G., Leitsch, A., and Mundici, D., editors, *Computational Logic and Proof Theory. Third Kurt Gödel Colloquium, KGC'93 Brno, Czech Republic, August 24–27, 1993*, pages 23–40. Springer.

Feferman, S. (1995). Introductory note to *1933o. In Gödel 1995, 36–44.

Feferman, S. (2005). The Gödel Editorial Project: A synopsis. *Bulletin of Symbolic Logic*, 11(2):132–149.

Franchella, M. (1995). L .E. J. Brouwer: toward intuitionistic logic. *Historia Mathematica*, 22:304–322.

Gandy, R. (1988). The confluence of ideas in 1936. In *A Half-century Survey on The Universal Turing Machine*, pages 55–111. Oxford University Press.

Gebhard, C. (2012). *Sprechtempo im Sprachvergleich: eine Untersuchung phonologischer und kultureller Aspekte anhand von Nachrichtensendungen*. PhD thesis, Humboldt Universität zu Berlin, Philosophische Fakultät II.

Gentzen, G. (1936). Die Widerspruchsfreiheit der reinen Zahlentheorie. *Mathematische Annalen*, 112:493–565.

Glivenko, V. (1928). Sur la logique de M. Brouwer. *Académie Royale de Belgique, Bulletin*, 14:225–228.

Glivenko, V. (1929). Sur quelques points de la logique de M. Brouwer. *Bulletins de la classe des sciences*, 15(5):183–188.

Gödel, K. (1931). Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme, I. *Monatshefte für Mathematik und Physik*, 38:163–198.

Gödel, K. (1932). Zum intuitionistischen Aussagenkalkül. *Anzeiger der Akademie der Wissenschaften in Wien*, 69:65–66.

Gödel, K. (1933a). Eine interpretation des intuitionistischen aussagenkalküls. *Ergebnisse eines mathematischen Kolloquiums*, 4:39–40.

Gödel, K. (1933b). The present situation in the foundations of mathematics. Printed in Gödel 1995, 45–53.

Gödel, K. (1933c). Zur intuitionistischen arithmetik und zahlentheorie. *Ergebnisse eines mathematischen Kolloquiums*, 4:34–38.

Gödel, K. (1938). Vortrag bei Zilsel. Printed in Gödel 1995, 86–113.

Gödel, K. (1941). In what sense is intuitionistic logic constructive? Printed in Gödel 1995, 189–200.

Gödel, K. (1946). Remarks before the Princeton Bicentennial Conference on Problems in Mathematics. In Gödel 1990, 150–153.

Gödel, K. (1958). Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes. *Dialectica*, 12:280–287.

Gödel, K. (1964). What is Cantor's continuum problem? In Benacerraf, P. and Putnam, H., editors, *Philosophy of Mathematics: Selected Readings*, pages 258–273. Prentice-Hall.

Gödel, K. (1986). *Collected Works. Vol. I. Publications 1929–1936.* Edited by S. Feferman, J. W. Dawson, S. C. Kleene, G. H. Moore, R. M. Solovay, and J. van Heijenoort. Oxford University Press.

Gödel, K. (1990). *Collected Works. Vol. II. Publications 1938–1974.* Edited by S. Feferman, J. W. Dawson, W. Goldfarb, C. Parsons, and R. M. Solovay. Oxford University Press.

Gödel, K. (1995). *Collected Works. Vol. III. Unpublished Essays and Lectures.* Edited by S. Feferman, J. W. Dawson, W. Goldfarb, C. Parsons, and R. M. Solovay. Oxford University Press.

Gödel, K. (2003a). *Collected Works. Vol. IV. Correspondence A–G.* Edited by S. Feferman, J. W. Dawson, W. Goldfarb, C. Parsons, and W. Sieg. Oxford University Press.

Gödel, K. (2003b). *Collected Works. Vol. V. Correspondence H–Z.* Edited by S. Feferman, J. W. Dawson, W. Goldfarb, C. Parsons, and W. Sieg. Oxford University Press.

Gödel, K. (2019). *Kurt Gödel: Philosophische Notizbücher / Philosophical Notebooks: Philosophie I Maximen 0 / Philosophy I Maxims 0.* Edited by E. Engelen. De Gruyter.

Gödel, K. (forthcoming 2020). *The Princeton Lectures on Intuitionism.* Edited by M. Hämeen-Anttila and J. von Plato. Springer.

Golland, L. and Sigmund, K. (2000). Exact thought in a demented time: Karl Menger and his Viennese mathematical colloquium. *The Mathematical Intelligencer*, 22(1):34–45.

Hahn, H. (1933). Die Krise der Anschauung. In *Krise und Neuaufbau in den exakten Wissenschaften*, pages 41–64. Deuticke.

Hahn, H. (1934). Gibt es Unendliches? In *Alte Probleme – Neue Lösungen in den exakten Wissenschaften*, pages 93–116. Deuticke, Leipzig und Wien.

Hahn, H. (1980). *Empiricism, Logic and Mathematics: Philosophical Papers.* Edited by B.F. McGuinness. Springer.

Hämeen-Anttila, M. (2019). Nominalistic ordinals, recursion on higher types, and finitism. *Bulletin of Symbolic Logic*, 25(1):101–124.

van Heijenoort, J. (1967). *From Frege to Gödel*. Cambridge: Harvard University Press.

Hesseling, D. E. (2003). *Gnomes in the Fog: The Reception of Brouwer's Intuitionism in the 1920s*. Birkhäuser.

Heyting, A. (1925). *Intuitionistische axiomatiek der projectieve meetkunde*. Noordhoff.

Heyting, A. (1927). Zur intuitionistischen Axiomatik der projektiven Geometrie. *Mathematische Annalen*, 98:491–538.

Heyting, A. (1930a). Die formalen Regeln der intuitionistischen Logik. *Sitzungsberichte der Preussischen Akademie von Wissenschaften*, 16:42–56.

Heyting, A. (1930b). Die formalen Regeln der intuitionistischen Mathematik I. *Sitzungsberichte der Preussischen Akademie von Wissenschaften*, 16:57–71.

Heyting, A. (1930c). Die formalen Regeln der intuitionistischen Mathematik II. *Sitzungsberichte der Preussischen Akademie von Wissenschaften*, 16:158–169.

Heyting, A. (1930d). Sur la logique intuitionniste. *Académie Royale de Belgique, Bulletin de la Classe des Sciences*, 16:957–963.

Heyting, A. (1931). Die intuitionistische Grundlegung der Mathematik. *Erkenntnis*, 2:106–115.

Heyting, A. (1934). *Mathematische Grundlagenforschung. Intuitionismus. Beweistheorie.* Springer-Verlag.

Heyting, A. (1936). De ontwikkeling van de intuitionistische wiskunde. *Euclides*, 13:129–144.

Heyting, A. (1978). History of the foundations of mathematics. *Nieuw Archief voor Wiskunde*, 26:1–21.

Hilbert, D. (1922). Neubegründung der Mathematik: Erste Mitteilung. *Abhandlungen aus dem Seminar der Hamburgischen Universität*, 1:157–177.

Hilbert, D. (1926). Über das Unendliche. *Mathematische Annalen*, 95:161–190.

Hilbert, D. (1927). Die Grundlagen der Mathematik. *Abhandlungen aus dem Seminar der Hamburgischen Universität*, 5:65–85.

Hilbert, D. (1930). *Grundlagen der Geometrie, 7te Auflage*. Teubner.

Hilbert, D. (2013). *David Hilbert's Lectures on the Foundations of Arithmetic and Logic 1917–1933.* Edited by W. Ewald and W. Sieg. Springer-Verlag.

Hilbert, D. and Ackermann, W. (1928). *Grundzüge der theoretischen Logik.* Springer-Verlag.

Howard, W. (1970). Assignment of ordinals to terms for primitive recursive functionals of finite type. In A. Kino, J. M. and Vesley, R., editors, *Intuitionism and Proof Theory: Proceedings of the Summer Conference at Buffalo N.Y. 1968*, volume 60 of *Studies in Logic and the Foundations of Mathematics*, pages 443–458. Elsevier.

Johansson, I. (1937). Der Minimalkalkül, ein reduzierter intuitionistischer Formalismus. *Compositio Mathematica*, 4:119–136.

Kennedy, J. (2014). Gödel's 1946 Princeton bicentennial lecture: an appreciation. In Kennedy, J., editor, *Interpreting Gödel: Critical Essays*, pages 109–130. Cambridge University Press.

Kleene, S. C. (1945). On the interpretation of intuitionistic number theory. *The Journal of Symbolic Logic*, 10:109–124.

Kohlenbach, U. (2008). *Applied Proof Theory: Proof Interpretations and their Use in Mathematics*. Springer-Verlag.

Kolmogorov, A. (1925). O principe tertium non datur. *Matematiceskij Sbornik*, 32:646–667. Reprinted in English as "On the principle of the excluded middle" in van Heijenoort 1967, 414–437.

Kolmogorov, A. (1932). Zur Deutung der intuitionistischen Logik. *Mathematische Zeitschrift*, 35:58–65.

Kreisel, G. (1959). Interpretation of analysis by means of constructive functionals of finite types. In Heyting, A., editor, *Constructivity in Mathematics*, pages 101–128. North-Holland.

Kreisel, G. (1967). Informal rigour and completeness proofs. In Lakatos, I., editor, *Problems in the Philosophy of Mathematics*, pages 138–157. North-Holland.

Kreisel, G. (1971). The collected papers of Gerhard Gentzen. *Journal of Philosophy*, 68(8):238–265.

Kreisel, G. (1980). Kurt Gödel, 28 April 1906 – 14 January 1978. *Biographical Memoirs of Fellows of the Royal Society*, 26:148–224.

Kreisel, G. (1987). Gödel's excursions into intuitionistic logic. In Weingartner, P. and Schmetterer, L., editors, *Gödel Remembered*, pages 67–179. Bibliopolis.

Lethen, T. and Passon, O. (forthcoming 2020). *Kurt Gödel's Remarks on the Foundations of Quantum Mechanics*.

Mancosu, P. (1998). *From Brouwer to Hilbert: The Debate on the Foundations of Mathematics in the 1920s*. Oxford University Press.

Mancosu, P. (2010). *The Adventure of Reason: Interplay Between Philosophy of Mathematics and Mathematical Logic, 1900–1940*. Oxford University Press.

Menger, K. (1930). Der Intuitionismus. *Blätter Für Deutsche Philosophie*, 4:311–325.

Menger, K. (1979). *Selected Papers in Logic and Foundations, Didactics, Economics*. D. Reidel.

Moschovakis, J. R. (2009). The logic of Brouwer and Heyting. In Gabbay, D., editor, *Logic from Russell to Church*, pages 77–125. Elsevier.

Myhill, J. (1973). Some properties of Intuitionistic Zermelo-Frankel set theory. In Mathias, A. R. D. and Rogers, H., editors, *Cambridge Summer School in Mathematical Logic*, pages 206–231. New York: Springer Verlag.

Negri, S. and von Plato, J. (2001). *Structural Proof Theory*. Cambridge University Press.

Parsons, C. (1970). On a number theoretic choice schema and its relation to induction. In Kino, A., Myhill, J., and Vesley, R., editors, *Intuitionism and Proof Theory: Proceedings of the Summer Conference at Buffalo N.Y. 1968*, volume 60 of *Studies in Logic and the Foundations of Mathematics*, pages 459–473. Elsevier.

Parsons, C. (1995). Platonism and mathematical intuition in Kurt Gödel's thought. *Bulletin of Symbolic Logic*, 1(1):44–74.

Petersen, U. (2003). $L^iD^Z_\lambda$ as a basis for PRA. *Archive for Mathematical Logic*, 42(7):665–694.

von Plato, J. (1999). Proof theory of full classical propositional logic. Manuscript.

von Plato, J. (2017a). *The Great Formal Machinery Works: Theories of Deduction and Computation at the Origins of the Digital Age*. Princeton University Press.

von Plato, J. (2017b). *Saved from the Cellar: Gerhard Gentzen's Shorthand Notes on Logic and Foundations of Mathematics*. Springer.

von Plato, J. (2019). In search of the sources of incompleteness. In *Proceedings of the International Congress of Mathematicians (ICM 2018)*, pages 4043–4060.

von Plato, J. (forthcoming 2020). *Can Mathematics Be Proven Consistent? Kurt Gödel's shorthand notes and lectures on incompleteness*.

Prawitz, D. (1965). *Natural Deduction: A Proof-Theoretical Study*. Dover Publications.

Shoenfield, J. R. (1967). *Mathematical Logic*. Addison-Wesley.

Sieg, W. and Parsons, C. (1995). Introductory note to *1938. In Gödel 1995, 62–85.

Sigmund, K. (2011). 'Docent Gödel will not lecture'. In Baaz, M. et al., editors, *Kurt Gödel and the Foundations of Mathematics*, pages 75–109. Cambridge.

Spector, C. (1962). Provably recursive functionals of analysis: a consistency proof of analysis by an extension of principles formulated in current intuitionistic mathematics. In Dekker, J., editor, *Recursive Function Theory, Symposia in Pure Mathematics*, volume 5 of *Proceedings of Symposia in Pure Mathematics*, pages 1–27. American Mathematical Society.

van Stigt, W. P. (1990). *Brouwer's Intuitionism*. North-Holland.

Sundholm, G. (1983). Constructions, proofs and the meaning of logical constants. *Journal of Philosophical Logic*, 12(2):151–172.

Sundholm, G. and van Atten, M. (2008). The proper explanation of intuitionistic logic: on Brouwer's demonstration of the Bar Theorem. In van Atten, M., Boldini, P., Bourdeau, M., and Heinzmann, G., editors, *One Hundred Years of Intuitionism (1907–2007): The Cerisy Conference*, pages 60–77. Birkhäuser Basel.

Szabo, M. (1969). *Collected Papers of Gerhard Gentzen*. Studies in Logic and the Foundations of Mathematics. North-Holland.

Tait, W. W. (2006a). Gödel's correspondence on proof theory and constructive mathematics. *Philosophia Mathematica*, 14(1):76–111.

Tait, W. W. (2006b). Gödel's interpretation of intuitionism. *Philosophia Mathematica*, 14(2):208–228.

Tait, W. W. (2010). Gödel on intuition and on Hilbert's finitism. In Gödel, K., Feferman, S., Parsons, C., and Simpson, S. G., editors, *Kurt Gödel: Essays for His Centennial*. Association for Symbolic Logic.

Thiel, C. (1988). Die Kontroverse um die intuitionistische Logik vor ihrer Axiomatisierung durch Heyting im Jahre 1930. *History and Philosophy of Logic*, 9(1):67–75.

Tieszen, R. (1995). What is the philosophical basis of intuitionistic mathematics? In Prawitz, D., Skyrms, B., and Westerståhl, D., editors, *Logic, Methodology and Philosophy of Science IX*, volume 134 of *Studies in Logic and the Foundations of Mathematics*, pages 579–594. Elsevier.

Toledo, S. (2011). Sue Toledo's notes of her conversations with Gödel in 1972–5. In Kennedy, J. and Kossak, R., editors, *Set Theory, Arithmetic, and Foundations of Mathematics: Theorems, Philosophies*, Lecture Notes in Logic, page 200–207. Cambridge University Press.

Troelstra, A. S. (1973). *Metamathematical Investigation of Intuitionistic Arithmetic and Analysis*. Springer.

Troelstra, A. S. (1986). Introductory note to 1933e. In Gödel 1986, 282–287.

Troelstra, A. S. (1990a). Introductory note to 1958 and 1972. In Gödel 1990, 217–239.

Troelstra, A. S. (1990b). On the early history of intuitionistic logic. In Petkov, P. P., editor, *Mathematical Logic*, pages 3–17. Plenum Press.

Troelstra, A. S. (1995). Introductory note to *1941. In Gödel 1995, 186–189.

Troelstra, A. S. (2011). History of constructivism in the 20th century. In Kennedy, J. and Kossak, R., editors, *Set Theory, Arithmetic, and Foundations of Mathematics: Theorems, Philosophies*, Lecture Notes in Logic, pages 150–179. Cambridge University Press.

Urquhart, A. (2016). Russell and Gödel. *The Bulletin of Symbolic Logic*, 22:504–520.

Vidal-Rosset, J. (2012). L'argument de Russell-Tennant. In Guay, A., editor, *Autour des* Principia Mathematica *de Russell et Whitehead, Textes réunis et présentés par Alexandre Guay*, Collection Histoire et philosophie des sciences, pages 139–167. Éditions Universitaires de Dijon.

Wang, H. (1987). *Reflections on Kurt Gödel*. MIT Press.

Wang, H. (1996). *A Logical Journey. From Gödel to Philosophy*. MIT Press.

Wavre, R. (1924). Y a-t-il une crise des mathématiques? *Revue de Métaphysique et de Morale*, 31:435–470.

Wavre, R. (1926a). Logique formelle et logique empirique. *Revue de Métaphysique et de Morale*, 33:65–75.

Wavre, R. (1926b). Sur le principe du tiers exclu. *Revue de Métaphysique et de Morale*, 33:425–430.

Weyl, H. (1921). Über die neue Grundlagenkrise der Mathematik. *Mathematische Zeitschrift*, 10:39–79.

Weyl, H. (1925). Die heutige Erkenntnislage in der Mathematik. *Symposion*, 1:1–32.

Weyl, H. (1926). *Philosophie der Mathematik und Naturwissenschaft*. Oldenbourg.

Weyl, H. (1944). David Hilbert and his mathematical work. *Bulletin of the American Mathematical Society*, 50:612–654.

Zach, R. (2003). The practice of finitism: Epsilon calculus and consistency proofs in Hilbert's Program. *Synthese*, 137:211–259.