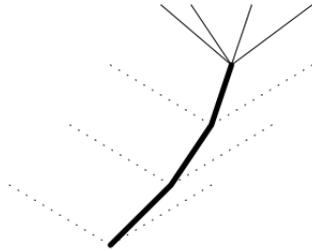


The Hilbert-Brouwer controversy resolved?

Per Martin-Löf

Since this lecture is called the Beth lecture, it is natural to ask oneself: what connection is there between Beth and intuitionism? and, to my mind, the most obvious answer is: the Beth models for intuitionistic propositional and predicate logic. I would like to remind you of a certain picture that we all use when thinking in terms of possible worlds, namely the picture in which we are always at a certain stage, considering a number of possible future alternatives, and then one of those alternatives is realized and we get to a new stage, at which we are faced with some other alternatives, one of those alternatives materializes, and then the whole process continues in the same way. That means that, at each moment, we are at the end of a certain path, which I will draw by a thick line, symbolizing that which has already materialized, and we are considering all possible future alternatives, one of those alternatives materializes, as a result of which the thick path gets extended, and then the pattern repeats itself:



We all think in terms of this picture: it underlies possible world semantics, and it is natural to ask: where does it really come from? I mean: who had this picture for the first time? Most of us have probably met it in connection with Kripke semantics (Kripke 1965), but actually Kripke semantics was predated by Beth semantics by a couple of years: Beth's papers were published in 1956 and 1959 (Beth 1956a, 1959), and Beth in turn drew on Brouwer here, because he took the nodes, or stages, to form a spread in Brouwer's sense

(Brouwer 1918B). So the picture that we standardly draw in possible world semantics is actually the picture of a spread, and maybe this is the most important, or at least most widespread, use that the notion of spread has been put to. Moreover, the thick path, which gets extended as soon as one of the future alternatives with which we are faced materializes, is precisely a choice sequence in Brouwer's sense, of which only a finite initial segment has been determined at each stage of its development. So there is this connection between Beth and intuitionism, which makes it not quite inappropriate that this lecture is called the Beth lecture, although this is a meeting celebrating the centenary of Brouwer's intuitionism.

Now the other thing one may wonder about is: why is this meeting held in France and not in the Netherlands as one would expect? And, again, this is not so strange as one might think, and several of the previous speakers have already addressed this question. It has to do with the fact that, as far as mathematics is concerned, Brouwer's intellectual lineage, which he referred to in many places, derives from Poincaré and proceeds via Borel to himself, and anyone who is familiar with Brouwer's work knows that this is not just a rationalization but an adequate account of how he was formed intellectually. There is a touching statement of this that he himself made in 1949, the last time, I suppose, that he was in Paris on a professional visit. Thanks to Van Dalen, this has now become available to anybody who wants to read it, in the second volume of his Brouwer biography, and so I will quote this passage here:

Addressing you here, I also feel an emotion of recollection. It is now exactly forty years ago, it was in December 1909 and January 1910, that I inhaled for the first time the scientific air of Paris. And I recall vividly and with a deep gratitude the encouraging benevolence, with which I, a young beginner, was received by the grand old men of that time, whose names had been tied to the grandiose evolution of the mathematical sciences which was taking place, and to whom, through their published courses and their monographs my generation owes the greater part of its knowledge and a considerable part of its inspiration. It was the five classical memoirs on analysis situs of Poincaré and the reflections on dimensionality of Poincaré that were to open the perspectives in which my thoughts on topology developed themselves, and it was the studies of Poincaré and Borel, in particular the manner in which the latter had introduced the notion of measure, that made me glimpse the direction in which I had to seek the primordial origin of mathematical thought. It is a very special sensation to find of these great old men of the past, after a so considerable lapse, some who survive full of vigour (van Dalen 2005, pp.848–849, translation modified).¹

¹ I am grateful to Dirk van Dalen for providing me with the French original of this passage of Brouwer's Paris address.

And I imagine that those who survived full of vigour in front of him were Borel, in the first place, and presumably also Hadamard, who was older than Borel but nevertheless full of vigour in 1949. So this was Brouwer's final tribute to his French ancestors, Poincaré and Borel.

Another thing that has been alluded to by several of the previous speakers is the fact that Brouwer did not introduce the terms intuitionism and formalism until 1911 in his review of a book on elementary mathematics by Manoury (Brouwer 1911A). There the terms appear for the first time, but what he had in mind then with intuitionism was the French school, Poincaré and Borel in the first place, and when, in his inaugural lecture of 1912, he wanted to refer to his own viewpoint, he called that neo-intuitionism (Brouwer 1912A). In fact, it was not until the twenties that he took the shrewd step of calling his own conception intuitionism tout court, qualifying his predecessors instead as either the pre-intuitionists or the old-intuitionists, by Heyting also called the semi-intuitionists. So this much about the appropriateness of holding this meeting in France.

Now to the proper subject of my talk, the Hilbert-Brouwer controversy. What that controversy was about I take it that we all know, but I would like to recall the crucial steps in this controversy, which started exactly in 1900 and stretched over a period of almost thirty years, until 1928. That year I take to be the natural stopping point as far as the involvement of Hilbert and Brouwer themselves is concerned. In 1900, it was in the *Mathematical Problems* paper (Hilbert 1900b) that Hilbert stated for the first time what we now call the second Hilbert problem, the problem of the consistency of the arithmetical axioms, as he called it, which means that really this story begins with the previous paper of his *Über den Zahlbegriff* (Hilbert 1900a), published in the same year, in which he introduced for the first time the modern axiom system for the real numbers. Now it is in his detailed statement of the second problem that he conceives of the idea that one should prove the consistency of the axioms, not in the natural way by constructing a model, as one had always done in geometry, but by a study of the proof figures, viewed as purely combinatorial sign configurations. So this is the birthplace of Hilbert's program, and then he gave a somewhat more detailed sketch of it in his Heidelberg lecture of 1904, *Über die Grundlagen der Logik und der Arithmetik* (Hilbert 1905).

This was all that was available to Brouwer when he wrote his thesis in 1907, which we are here to commemorate (Brouwer 1907), and it is remarkable that he was able to give such a penetrating criticism of Hilbert's program at this early stage, when one remembers that the only items that were available to him were the Heidelberg lecture of 1904 and the *Mathematical Problems* paper from 1900. And the most important point for the future development, and also for my talk, was no doubt his criticism already there of Hilbert's solvability axiom, or solvability conviction, if you prefer, the conviction that every mathematical problem can be solved positively or negatively. The importance he attached to this point is clear from the fact that he put it last in his series

of theses, saying outright that this conviction of Hilbert's is unfounded, which is to say that he thought that there was no evidence at all for this conviction. The following year, 1908, this was turned into the criticism of the law of excluded middle by simply identifying the law of excluded middle, in its constructive interpretation, with Hilbert's solvability axiom. Then, in 1912, he gave his inaugural lecture *Intuitionism and Formalism* (Brouwer 1912A): so now the terms intuitionism and formalism have come in, but it did not contain so much new actually: it was only an inaugural lecture after all.

A new phase in this controversy began in 1921 with the publication of Weyl's paper *Über die neue Grundlagenkrise der Mathematik* (Weyl 1921), and that is what really fired it and made it so bitter. (And it seems clear that it had to do with the fact that Weyl was after all Hilbert's doctoral student: he took his doctor's degree with Hilbert, and I do not know, but presumably Hilbert thought of him as the best of his doctoral students over the years.) So Weyl had been converted to intuitionism and wrote this paper, which is very rhetorical, and where had he learned this kind of rhetoric? From Hilbert, no doubt. Hilbert was a master in writing rhetorical papers: if you think of his *Axiomatisches Denken* (Hilbert 1918), for instance, it is a very beautiful piece of rhetoric, and now he suddenly had this quite rhetorical paper directed against him, and that fired him quite clearly, mathematically as well as rhetorically, and his response was to give lectures at various universities, all of them on essentially the same topic. The most well-known is the one that he gave in Hamburg in 1921: so that was immediately after the Weyl paper, and that was printed in 1922, called *Neubegründung der Mathematik* (Hilbert 1922). There we find several of the memorable quotations that we have all read, and that were his answer to Weyl's rhetoric. After all, he wanted to make clear that he was not only the mathematical master but also the rhetorical master, and I do not know what one should say about this rhetoric that came into the purely scientific debate as an additional element: it has its negative sides, we probably all agree, but it is not entirely negative, producing as it did some of the more quotable sentences that we have in this area, and if we are not now letting us be seduced by this rhetoric, the quotations can with profit be used on various occasions to bring the purely scientific issues out in sharper relief. The following year, there appeared a quite similar paper of Hilbert's, *Die logischen Grundlagen der Mathematik* (Hilbert 1923), and then, in 1925, *Über das Unendliche* (Hilbert 1926), again with a lot of rhetoric in it and, as we now know, also mathematical mistakes. The final Hilbert paper that I want to refer to, containing as it does some of Hilbert's most outrageous sentences about Brouwer and intuitionism, is *Grundlagen der Mathematik* (Hilbert 1928), with whose publication we are already in 1928.

Now Brouwer, who is always said to have been such a difficult person, was actually relatively quiet, pursuing his mathematics, during these years. His basic papers on intuitionism, his reconstruction papers, date after all from the years 1918–1927, and I have only been able to locate one paper, albeit a very well-known one, *Über die Bedeutung des Satzes vom ausgeschlossenen Dritten*

in der Mathematik, insbesondere in der Funktionentheorie (Brouwer 1924N), which contributes to this rhetorical fight in any clear way. It is that paper which contains the often quoted half of a sentence:

eine durch keinen Widerspruch zu hemmende unrichtige Theorie ist darum nicht weniger unrichtig, so wie eine durch kein reprimierendes Gericht zu hemmende verbrecherische Politik darum nicht weniger verbrecherisch ist (Brouwer 1924N, p.3).²

Of course, that was quite in parity with Hilbert as far as rhetoric is concerned, but it seems to be the single place where he took to this kind of language, with the exception of his final contribution to the controversy with Hilbert, which was his 1928 paper *Intuitionistische Betrachtungen über den Formalismus* (Brouwer 1928A2), a paper which I take it that we have all studied in one way or the other. So this is the controversy as it has come down to us through their own writings, and Brouwer himself characterized the situation, already towards the end of his inaugural lecture of 1912 in the following way:

So far my exposition of the fundamental issue, which divides the mathematical world. There are eminent scholars on both sides and the chance of reaching an agreement within a finite period is practically excluded (Brouwer 1913C, p.96).

which means that he thought that it was almost excluded that this controversy would not go on forever. Now my own attitude towards this is that, when there is very serious disagreement, to the extent that controversy is a better word, and you cannot make up your mind, what is the natural attitude? Well, keep quiet and just wait and see what happens: maybe eventually new information will accumulate that will make us look at the controversy in a different way. And, indeed, this is to my mind what has happened: although it did not take an endless time, it took a very long time, and I would say that it is only around ten years ago, in the nineties that is to say, that we have come to get a proper understanding of the scientific content of this controversy, an understanding which makes it not unreasonable to speak of a settlement of the controversy, so that it finally ceases to be a problem for us.

So now I will tell two different stories that have eventually provided the information that makes the Hilbert-Brouwer controversy look different now from what it looked like in 1928, and the first of these is the double-negation interpretation. Naturally, the mathematical community was at the time very much bothered by this controversy, and especially those who were young enough to have their formative years in the twenties, they eagerly wanted to learn: what was it all about, and could something be done about it? One of them was Kolmogorov, and one can see in his paper *On the law of the excluded middle*

² an incorrect theory, even if it cannot be inhibited by any contradiction that would refute it, is none the less incorrect, just as a criminal policy is none the less criminal even if it cannot be inhibited by any court that would curb it (van Heijenoort 1967, p.336).

from 1925 (Kolmogorov 1925) a very strong wish to reach some kind of objective verdict in this controversy, but not only in that paper: there is also a much less known paper that was published in a popular scientific journal in 1929, *Contemporary debates on the nature of mathematics* (Kolmogorov 1929), which testifies to the same strong wish. And, if you read the *Résumé* which you find at the end of his 1925 paper, it is clear that he thought actually at the time that he had been able to settle this controversy. And how did he settle it? Well, he settled it by forgetting about Brouwer's distaste for language and logic: he sat down with the Hilbert axioms for propositional and predicate logic, removing those that did not look all right on a constructive interpretation and retaining the other ones, except for the crucial law *ex falso quodlibet*, which it remained for Heyting to add to the system in 1930, and then he simply interpreted all of classical propositional and predicate logic by means of constructive propositional and predicate logic, utilizing what we now call the double-negation interpretation. And, indeed, that does give a constructive interpretation of classical logic, and he thought that, having achieved that, the problem of giving a constructive interpretation of classical mathematics in its entirety was essentially solved. This was the mistake, because mathematics cannot be built on propositional and predicate logic alone: there has to be some amount of set theory also, and two questions then arise, first: what should be the laws of this set theory? and, second: does this set theory with classical logic, which is to say classical mathematics, allow the double-negation interpretation to go through? These questions were not addressed in Kolmogorov's 1925 paper, so he was overoptimistic at the time.

Then the double-negation interpretation was independently rediscovered, as we all know, by Gödel and Gentzen in 1933 (Gödel 1933a), but, more importantly, they extended it from pure logic to first-order arithmetic, and that immediately gave a constructive interpretation, and therefore a constructive consistency proof, of first-order arithmetic, but the question remained: could it be extended beyond first-order to second- or higher-order arithmetic? On this point, a crucial role was played by Spector's last paper, published posthumously in 1962 (Spector 1962), since he died the year before. Spector's paper contained a consistency proof of analysis by an extension of principles formulated in current intuitionistic mathematics, and the extension was by the principle of so-called bar recursion of higher type. The paper gave new hope that maybe one could after all give a constructive interpretation of all of classical second-order arithmetic, which would yield a positive solution to Hilbert's second problem. Now Spector's principle of bar recursion of higher type was a quite complicated scheme of axioms, and so the question immediately arose: do we at all have an interpretation that validates them? The natural interpretation turns out to be by means of the so-called continuous functionals of finite type, and if you start thinking about what abstractions you need in order to prove that these axioms are validated under the continuous functional interpretation, the conclusion is: we need all of classical second-order arith-

metic, strengthened by the axiom of dependent choices, which means that no reduction whatsoever has been obtained.³

This led to Kreisel's proposal, in 1968, of accepting instead intuitionistic second-order arithmetic, which he showed to have the very good property that you immediately interpret classical second-order arithmetic in it by means of the old double-negation interpretation (Kreisel 1968). So, if intuitionistic second-order arithmetic is constructively acceptable, then we do indeed have a solution to the second Hilbert problem, because Hilbert's second problem was at an early stage associated precisely with classical second-order arithmetic. So optimism at this time.

Then this was further extended by Myhill in 1971, with a correction in 1974, from second-order logic to full simple type theory (Myhill 1971, 1974). So now we had intuitionistic simple type theory over arithmetic at the bottom, and again the double-negation interpretation extended: classical higher-order arithmetic is easily translatable into the corresponding intuitionistic theory by means of the double-negation interpretation, so if this intuitionistic theory is comprehensible, or acceptable, then we have indeed, among other things, a consistency proof for a full-scale system of classical mathematics. And then Friedman (Friedman 1973) carried it even a bit further, to Zermelo-Fraenkel set theory, by modifying the axioms of Zermelo-Fraenkel set theory so that they became formally intuitionistic, by which I mean that the logic is intuitionistic, and he proved that classical Zermelo-Fraenkel set theory is interpretable in this intuitionistic, or formally intuitionistic, version of Zermelo-Fraenkel, thereby bringing this line of development to its natural conclusion. It looked very promising: maybe we just had to accept these impredicative intuitionistic systems as comprehensible.

This is what I did myself during a year and a half, from 1970 to 1971, as a result of Girard's extension of the proof of normalization to second-order logic by means of his so-called candidates of reducibility in 1970 (Girard 1971). That had been a long-standing open problem, to prove normalization for second-order logic, and, indeed, if you allow yourself to use second-order abstractions on the metalevel, you could do that, as Girard showed. I was naturally impressed by this and leaned towards simply accepting those principles: at least we had learned how to use them to prove this particular combinatorial result on normalization. But this optimism did not last very long in my case, because at the same time approximately, in 1969, the Curry-Howard correspondence, or isomorphism, between propositions and sets had been discovered (Howard 1980). I will use the term set, but you should understand it in the sense of individual domain, or quantificational domain: in first-order predicate logic, you always quantify over a domain, and Curry-Howard is a correspondence, or isomorphism, between propositions and sets in that sense. The Curry-Howard correspondence was to me from the beginning the natural

³ I am grateful to Thierry Coquand for criticism which has made me let the text differ at this point from the talk as I gave it at the conference.

completion of the Brouwer-Heyting-Kolmogorov interpretation of intuitionistic first-order predicate logic, just drawing the full consequences, so to say, of the Brouwer-Heyting-Kolmogorov interpretation. Now the paradox that Girard discovered in 1971 (Girard 1972), in the first version of constructive type theory, then called intuitionistic type theory, showed that Curry-Howard is incompatible with impredicativity: if you combine the Curry-Howard identification of propositions and sets with impredicativity, that is, the possibility of quantifying over propositions, over classes, over relations, and so on, you end up with a contradiction. So one of them has to go: either Curry-Howard has to go or there is some problem with impredicativity, with which there had been problems from the very beginning: when the notion itself was introduced by Russell in 1906 (Russell 1906), it was precisely because it was a problematic notion. So it seems fair to say that Girard's paradox put an end to the optimism about impredicative intuitionism in the 1960s, and maybe the form in which it convinced most people was the form that was given to it in 1986, which is after the calculus of constructions had come into being, independently and at the same time by Coquand in France and by Hooke and Howe as well as Mitchell in the United States, who proved that, if you combine impredicative second-order existential quantification with the strong rules of existential elimination, which are clearly valid under the Brouwer-Heyting-Kolmogorov interpretation, you already arrive at an inconsistency (Coquand 1986; Hook and Howe 1986; Mitchell and Harper 1988). So, hopeful as impredicative intuitionism had seemed for some time, during the approximate decade from Spector's paper to Girard's paradox, it nevertheless ended in disillusion.

My second story is about ordinal proof theory, more particularly, about the proof-theoretic scale, which is the result of a long development in German proof theory, originally aiming at fulfilling Hilbert's program, which is to say producing finally the dreamt of constructive consistency proof of classical analysis. The development of this particular branch of Hilbert's proof theory begins as we all know with Gentzen's proof of the consistency of first-order arithmetic by means of purely combinatorial reasoning extended by transfinite induction up to ε_0 , an ordinal about whose constructive accessibility nobody has any doubt. Now this ordinal measures such a natural level of abstractions that there are many systems which lie on this level, not only classical but also intuitionistic first-order arithmetic, Gödel's theory T, Heyting arithmetic extended to all finite types, HA^ω , as it is usually called, possibly even with the constructive axiom of choice that can be formulated in that system, and also constructive type theory without any well-orderings and universes is a theory on this level. The year in which it was first reached, by Gentzen, is 1936 (Gentzen 1936).

Then, after the Second World War, this line of investigations was continued by Lorenzen and Schütte (Schütte 1951) in the fifties. The next natural stopping point is arithmetical analysis, whose proof-theoretic ordinal was determined to be $\varepsilon_{\varepsilon_0}$, and, if you begin iterating the process any finite number

of times, you will get up to the first critical epsilon number. This is the ordinal of full ramified analysis, which is to say arithmetic with ramified second-order quantification of all finite levels. Then, beyond that, there is an ordinal of some significance, which is usually denoted by $\phi_{\varepsilon_0}(0)$. It is the so-called Veblen notation that is used, and this ordinal was determined by Friedman in the late sixties (Friedman 1968) to be the ordinal of classical analysis, which is to say second-order arithmetic, with no comprehension axiom but instead with the so-called Σ_1^1 axiom of choice, which is to say the axiom of choice limited to formulas of Σ_1^1 form. And, again, there is a corresponding version of type theory: you just need to add to constructive type theory the axioms for one universe, and you get a system of precisely this proof-theoretical strength.

Then we get to the very well-known ordinal I_0 , which was identified by Schütte and Feferman in 1963–64 as the ordinal of full predicative analysis, that is, predicative analysis with not only finite but also transfinite ramification levels (Feferman 1964; Schütte 1965). And, again, this corresponds to a natural system of constructive type theory, namely constructive type theory with not just one universe but a whole sequence of successive universes U_0, U_1, \dots indexed by the sequence of natural numbers.

Then, beyond this, we reach a very significant level of abstractions: in terms of ordinals, it is the Howard ordinal $\phi_{\varepsilon_{\Omega+1}}(0)$, whose notation involves the so-called Bachmann extension of the Veblen hierarchy. This is a very significant level, because it is the level where, in addition to ordinary mathematical induction and recursion, you have transfinite induction and recursion, so it is precisely the kind of abstractions that Brouwer learned from Borel: if you remember the quotation that I gave from his lecture at the Sorbonne in 1949, he said that he was particularly indebted to Borel's definition of measure, and how did that go? Well, it was by transfinite recursion on the build-up of the Borel set, and that is a transfinite recursion on countable well-founded trees, precisely the kind of reasoning that he himself came to use in connection with the bar theorem. So this is roughly the proof-theoretic strength of the methods used by Brouwer himself in his writings, since he did not go to any higher number classes, and the theory which has this strength is usually referred to as ID_1 . The Howard ordinal is also the ordinal of Kripke-Platek set theory, and it is the ordinal of Aczel's constructive version of Zermelo-Fraenkel set theory, which, although it has this comparatively limited proof-theoretic strength, is nevertheless such that, when you add the law of excluded middle to it, you get full classical Zermelo-Fraenkel set theory.

Then, if you start iterating the inductive definitions, so as to have not only the second number class but also the third and the higher number classes, finitely high, you get to an ordinal which I can only write down in a notation such as $\theta(\Omega_\omega)$, where θ is technically called a collapsing function in this kind of proof theory. The only thing you need to know about it is that it is a function that takes an ordinal of any size and collapses it into a countable ordinal, and that, for the particular large ordinals, like Ω_ω , that we are going to collapse, the countable ordinals obtained by the collapsing procedure will be given by

notation systems which are defined in a purely combinatorial manner. The ordinal $\theta(\Omega_\omega)$ measures the proof-theoretic strength of the theory of finitely iterated generalized inductive definitions, and it is also the ordinal determined by Takeuti in 1967 for classical analysis with the Π_1^1 comprehension axiom and the schema of mathematical induction limited to Π_1^1 formulas (Takeuti 1967). Viewed type-theoretically, it is the ordinal of constructive type theory, without universes, but now with the general well-ordering operation W . On the other hand, if you take instead classical analysis, still with the Π_1^1 comprehension axiom, but now without the restriction on the induction scheme, you get to a slightly larger ordinal here, which happens to be $\theta(\Omega_\omega \cdot \varepsilon_0)$, but this is a rather unnatural ordinal, as you see, much less natural than $\theta(\Omega_\omega)$. On the other hand, if you not only add ordinary mathematical induction for formulas of arbitrary complexity but strengthen that from ordinary induction to bar induction, then you get even further here to an ordinal which is the collapse of the first epsilon number beyond Ω_ω , $\theta(\varepsilon_{\Omega_\omega+1})$, again a relatively technical and unnatural ordinal in comparison with $\theta(\Omega_\omega)$.

The next really important step further was taken in 1982 by Pohlers and Jaeger (Pohlers and Jäger 1983), who determined the proof-theoretic strength of classical analysis with the Δ_2^1 comprehension axiom plus the scheme of bar induction for formulas of arbitrary complexity to be $\theta(\varepsilon_{I+1})$. Now what makes this particularly interesting is that, for the first time, a large cardinal appears in the ordinal notation: it is the collapse of the first epsilon number beyond the first inaccessible, which is what I denote by I here. At the earlier stages, we have had not so wildly big cardinals, or initial ordinals, coming in, but now, remarkably enough, an inaccessible cardinal makes its appearance in the ordinal notation, and this in a piece of constructive proof theory. The ordinal $\theta(\varepsilon_{I+1})$ is also the proof-theoretic ordinal of Aczel's constructive set theory, when it is strengthened by the axiom that he calls the regular extension axiom. To reach this level with constructive type theory, you need to have both the well-ordering operation and a universe which is closed under the well-ordering operation. As shown by Setzer in his thesis of 1993 (Setzer 1993), the resulting system of type theory has the proof-theoretic ordinal $\theta(\Omega_{I+\omega})$, an ordinal which is even slightly bigger than $\theta(\varepsilon_{I+1})$.

We are now already in the nineties, when one has ventured to extend these results even further, and the only way in which one has succeeded in doing that has been by adding new large cardinal axioms, although they have been given constructive meaning. There was first the step taken by Rathjen in 1990–91 of replacing the inaccessible that we have already seen by a Mahlo cardinal instead (Rathjen 1990), thereby reaching $\theta(\varepsilon_{M+1})$, and proving this to be the ordinal of Kripke-Platek set theory, when strengthened by an axiom which forces the universe to be Mahlo (Rathjen 1991). Analogously, in 1996, Setzer added the axioms for a Mahlo universe to type theory and determined the proof-theoretic ordinal of the resulting system of type theory to be $\theta(\Omega_{M+\omega})$ (Setzer 1996). So this is the ordinal strength of constructive type theory with the well-ordering operation, as before, but now adding axioms for, not the

usual kind of universe, but a so-called Mahlo universe. Actually, one yet further step has been taken, both by Rathjen in terms of Kripke-Platek set theory and by Setzer in terms of constructive type theory, replacing the Mahlo cardinal by a weakly compact cardinal instead, but I have to leave this last step out.

The crucial point now is that what we see in front of us at this stage is some kind of abyss, or chasm, which we do not seem able to pass, and that the classical type- as well as set-theoretic systems, second- and higher-order arithmetic, Zermelo's set theory and Zermelo-Fraenkel set theory, they are all on the other side of this abyss, and it is not only when you go to full second-order arithmetic but already when you go to the next form of the comprehension axiom, which is to say Π_2^1 comprehension, that you find yourself on the other side of the abyss. On this level of proof-theoretic strength, progress has been possible to make, again by Rathjen, only by beginning to postulate certain large cardinals with the right properties, so that you get notation systems which are big enough, but this is a complete change in the methodology of constructive proof theory, which brings it rather into contact with current set theory: you simply begin to postulate instead of building things up by means of evident principles, and, as it seems, that is the only way that is available if we want to proceed into this region, that is, to reach something like classical second-order arithmetic, which was Hilbert's original problem, or the corresponding intuitionistic theory, Heyting arithmetic of second-order.

So this is the picture that has emerged as the result of persistent work by the German proof-theoretic school during a period of more than seventy years. The original aim was to obtain a constructive consistency proof for classical analysis, which early on came to be identified with full second-order arithmetic, but we have now so much information that we know that this is out of our reach, and why? Well, if this is to be a constructive consistency proof, it will have to use constructively acceptable principles, and we know by now what are the strongest constructively acceptable principles that are available to us at the moment. It is of course never fixed at any absolutely precise level: as soon as you fix it, you can go beyond it by some kind of reflection principle, but basically we have at present exhausted the principles for which we can claim evidence, and this is a completely new situation. If you go back to the late sixties, for instance, when Takeuti did his fundamental work, it was still wide open how far you could go in this kind of proof theory. So we know by now that we cannot pass this abyss unless we are able to think up some brand-new strong constructive principles, and there is no sign whatsoever of that at the moment.

This means in particular that we have to take a different attitude to the systems on the other side of the abyss, like second-order arithmetic, and above all to the law of excluded middle. You see, some of the systems on this side of the abyss, namely Aczel's system CZF of constructive set theory on the one hand and constructive type theory on the other hand, are such that, when you add the law of excluded middle to them, you get full classical mathematics, which we know that we cannot interpret constructively unless a miracle inter-

venes. This is what I propose to call the second failure of the Hilbert program. The first failure of the Hilbert program was the one which was discovered by Gödel and of which we are now all aware, but that gave rise to the revised, or modified, Hilbert program, whose characteristic is that we no longer allow merely combinatorial methods in the consistency proof but arbitrarily strong constructive methods. But even this revised, or modified, Hilbert program has come to an end in the nineties, or has failed in the nineties, so it is the second failure of the original Hilbert program, which I cannot interpret in any other way than that we have to give up the dream of being able to establish the consistency of classical mathematics by constructive means.

We find ourselves in a new situation: it is a new picture that has emerged here, if you compare it with the optimistic picture that we had in the sixties, and to my mind the natural conclusion is that we have to look at the law of excluded middle in a different way, the way set theorists look at large cardinal and determinacy axioms rather. So we have constructive mathematics, and one non-evident principle that we can add to it is the law of excluded middle. Then some kind of explosion takes place: we pass from a fully understood system, which is on this side of the abyss, to full classical mathematics, which is on the other side of the abyss, and it is in the nature of things that we cannot be certain of the consistency of the classical system that we arrive at, because we transcend the principles that have evidence. On the other hand, we have inductive evidence of the consistency of the classical system in the sense that the classical set-theoretical principles have been investigated with respect to their consequences during such a long time and in such detail without there being any sign of an inconsistency arising from them. But, since we do go beyond what is intuitively evident when we adjoin the law of excluded middle, we are not certain that it cannot give rise to a contradiction in the way we are on the previous level, which is to say before the abyss. We are rather in the situation that Woodin has characterized by the sentence, or half of a sentence:

just as those who study large cardinals must admit the possibility that the notions are not consistent (Woodin 1998, p.330).

So the law of excluded middle has to be ranged together with the large cardinal and determinacy axioms rather than with the axioms of which we have a constructive, or contentual, understanding. This is in line with the fact that, if we reformulate the Brouwer-Heyting-Kolmogorov interpretation of intuitionistic propositional and predicate logic in terms of games and winning strategies, as suggested by Ranta in his paper *Propositions as games as types* (Ranta 1988), the law of excluded middle appears as the simplest and most general axiom of determinacy, which says that every game is determinate.

Now some final words on the Hilbert-Brouwer controversy, so as to justify the choice of the title of my talk: The Hilbert-Brouwer controversy resolved? To my mind, it is indeed a kind of resolution that we have reached as a result of the new information that has accumulated during the period from the sixties

to the nineties, and that I have reported on. So let us look at one side first, the Hilbert side.

We know what was Hilbert's foundational program: to formalize mathematics, allowing classical logic in the formalization, and then justifying it proof-theoretically, on the metalevel, by means of a consistency proof, which was first required to be purely combinatorial, like the consistency statement itself, and then liberalized to be merely constructive, allowing of any means as long as they carry conviction. This is precisely what has failed, not only in the first, but also in the second, more liberal, case. That means that the two-storeyed building of mathematics that Hilbert had in mind, that simply has not worked out: it is not to be avoided that we have to go through mathematics as it exists and disentangle the constructive parts of it from the uses of the law of excluded middle and Zermelo's axiom of choice. This is precisely what Brouwer started and what is still going on: at the present time, actually, a lot has been achieved in this direction. So this much about the Hilbert side.

Then, if you turn to Brouwer's views, the formulations that come first to my mind at least are:

mathematics is a mental construction (Brouwer 1947, p.477),

or:

mathematical objects are mental constructions,

on the one hand, and:

mathematics is a languageless activity of the mind
(Brouwer 1952B, p.141),

on the other. Now these statements are very understandable psychologically if one remembers that he was opposed to Hilbertian formalism: he was opposed to the idea that the mathematical objects are mere sign configurations, combinatorial sign configurations, as Hilbert thought, the classical formulation being:

am Anfang ist das Zeichen (Hilbert 1922, p.163),⁴

alluding to both Goethe's Faust and the Gospel of St. John. And Brouwer's way of contradicting this as much as possible was to say that, on the contrary, mathematical objects have nothing at all to do with signs, or with language: they are mental constructions, or thought constructions, but that seems to me to be a very difficult position to maintain nowadays. I think we must admit that Brouwer's views on language were of a rather old-fashioned nature: it is clear that he thought of language in the traditional way as the expression of thought, which is to say that the thoughts had to be there first, before possibly being expressed in language, and that view of language has after all been replaced, in the course of the last century, by the view of language as a

⁴ in the beginning is the sign

means of communication. And it is not for nothing that this shift has taken place: if you take Brouwer's view that the mathematical objects are mental constructions, or thought constructions, then there arises immediately the problem: how do we get to know them? and not much self-reflection is needed to see how we do get to know them. After all, there is no mathematical zoo to which we can be taken in order to have them displayed to us without any linguistic mediation: we go to listen to some lectures or to read some books or articles, which means that we are dependent on language and symbols from the very start. So this view of Brouwer's is hardly credible any longer because of the developments in the philosophy of language during the last century, as Dummett has emphasized for the first time in his paper *The philosophical basis of intuitionistic logic* (Dummett 1975). But there is a kind of mediating position, namely that the mathematical objects are not just purely formal sign configurations: they are meaningful sign configurations, and that is what gives them properties which do not come from their combinatorial nature, but come from the meaning with which they are endowed, as beautifully stated by Gödel in the beginning of his *Dialectica* paper (Gödel 1958). It is this view of the mathematical objects, mediating as it is between Hilbert and Brouwer, which has turned out to be the most credible one to my mind: at least it is the one which underlies all of my own work from 1974 until the present day. So, even if there is a question mark in my title, a cautious question mark, it seems to me that the progress that has been made during the thirty-year period from the sixties to the nineties is such as to make it not unreasonable to speak of a resolution of the Hilbert-Brouwer controversy, part of this resolution being that we have had to change our attitude towards the law of excluded middle in a quite drastic way.

Acknowledgement

I am greatly indebted to Mark van Atten, not only for recording the talk as I gave it at the conference on tape, but also for subsequently providing me with the technical equipment necessary for its transcription.