Sieg, W. / Tapp, C., "Introduction to the Undated Draft" [= David Hilbert, 'Consistency Proofs for Fragments of Arithmetic' (Undated Draft)]. In: David Hilbert's Lectures on the Foundations of Arithmetic and Logic 1917-1933, hrsg. v. W. Ewald / W. Sieg, Berlin: Springer 2013, 378-395.

Introduction to the Undated Draft

This manuscript is in Hilbert's own hand. On eleven large pages Hilbert examines a new technique for proving the consistency of quantifier-free fragments of number theory. He clearly wrote the manuscript between the end of the Summer Semester 1920 and the beginning of the Winter Semester 1921/22. Indeed, there are excellent reasons to think that it was written before talks he gave in Copenhagen (15 and 17 March 1921) and Hamburg (in the early summer of 1921). The publication 'Neubegründung der Mathematik' (*Hilbert 1922b*) was based on these talks, and the results reported here are intermediate between those from the notes for the lectures held in the Summer Semester of 1920 ('Probleme der mathematischen Logik', *Hilbert 1920**, reproduced in this Chapter) and those from the 'Neubegründung'. Thus, the manuscript is of systematic importance for understanding the development of proof-theoretic techniques. It is also historically significant, since it shows Hilbert pushing forward the novel investigations of his proof theory in crucial ways.

1. Background.

In the lecture notes for 'Probleme der mathematischen Logik', Hilbert reproved the consistency result presented to the 1904 International Congress of Mathematicians in Heidelberg; cf. *Hilbert 1905b* and section 1 of the Introduction to Chapter 3 below. Hilbert considered the Heidelberg proof as the first *direct* consistency argument. The considerations in the Summer Semester of 1920 take up those investigations for the first time since 1904; they are now (on p. 39) characterized as *proof-theoretic*. The basic and purely equational system in these lectures has the axioms

1. $1 = 1$,
2. $a = b \to (a + 1 = b + 1)$,
3. $(a + 1 = b + 1) \to a = b$,
4. $a = b \to (a = c \to b = c)$,
5. $a + 1 \neq 1$.

(See pp. 37 and 39 of the lecture notes, pp. 366 and 367 above.) *Modus ponens* is the sole inference rule, and a substitution rule allows the replacement of variables in axioms.[1] For the system of axioms 1.–4. Hilbert shows that any provable equation must be of the form $t = t$; thus the full system is consistent since $a + 1 = 1$ is not derivable from 1.–4. A crucial lemma in Hilbert's proof states that one cannot derive formulas with more than two '$\to$'-signs.

[1] Hilbert states explicitly that substitution is allowed in axioms only. However, discussing applications of the induction schema, he later talks about substituting a sign 'u' for the variable 'a' in theorems of the form $Z(a) \to F(a)$; see the lecture notes, pp. 44–45, above p. 370f.

At the very end of the notes, Hilbert indicates a way of strengthening the system and in this way making it more adequate for the formalization of mathematical practice. The system is to be expanded by the introduction of (*i*) a predicate letter 'Z', indicating the sort 'Zahlzeichen', with the accompanying axioms $Z(1)$ and $Z(a) \to Z(a+1)$; (*ii*) the induction rule in the form: for any formula $F(a)$, infer the formula $Z(a) \to F(a)$ from $F(1)$ and $F(a) \to F(a+1)$; and (*iii*) some additional equations for arithmetic operations. For example, the recursion equations for addition and multiplication are introduced:

6. $a + (b + 1) = (a + b) + 1$,
7. $a \cdot 1 = a$,
8. $a \cdot (b + 1) = a \cdot b + a$.[2]

As Hilbert points out (notes, p. 46, above p. 371), whenever axioms for a new sign or concept are introduced that modify the possibilities for giving proofs, the consistency of the resulting system has to be checked. Hilbert views the problem of establishing the consistency of these successively stronger formal frameworks as a significant and difficult task. Nevertheless, he points out that the 'guidelines for a complete proof of the consistency of number theory have thus been given' (p. 46 of the notes, p. 371 above).

Hilbert took on this difficult task during the break following the Summer Semester of 1920 and in the following Winter Semester. Hilbert reported in a postcard to Bernays of 22 October 1920 that his current work in proof theory was progressing quite well.[3] Moreover, he gave two talks in Göttingen on 21 and 22 February 1921 under the title 'Eine neue Grundlegung des Zahlbegriffs'. These considerations he presented, we assume, in two lectures in Copenhagen (entitled 'Axiomenlehre und Widerspruchsfreiheit') and then Hamburg, the contents of which form the basis of the paper 'Neubegründung der Mathematik' published in 1922. When commenting on Hilbert's first foundational paper of the 1920s, the Editors of his *Gesammelte Abhandlungen* describe it as reflecting a transition between two different stages in the development of proof theory.[4]

The manuscript considered here illuminates this transition in a way that is very significant for proof theory. Hilbert tackles the consistency problem for extensions of the sort described above, but he now includes also four new inference rules, and these new rules make it impossible to employ the

---

[2]The axioms 1 to 5 just formulate equality principles together with the characteristic properties of 1 and the successor operation, whereas the additional axioms express properties of the arithmetical operations and a form of the induction principle. This significant separation would have been clearer if Hilbert had written the successor operation, not in the form '$\ldots + 1$', but rather (as Dedekind did) as '$\varphi(\ldots)$'. Then there would have been a meaningful base equation for addition, namely, $a + 1 = \varphi(a)$.

[3]Hilbert says that his *Beweistheorie* 'beständig ganz gute Fortschritte macht'; see *Sieg 1999*, p. 35. For more on the context of the remark, and for the remainder of the contents of the postcard, see *Sieg 1999*, pp. 34–37.

[4]See *Hilbert 1935*, p. 168, n. 2. It seems plausible to assume that Bernays was responsible for the remark in this note. See p. 19, n. 33 above and also the further references given there.

proof-theoretic techniques used in the lecture notes for the Summer Semester of 1920, since there is no longer a fixed bound on the length of formulas occurring in proofs. An ingenious way of transforming proofs is introduced, but Hilbert's novel considerations also contain oversights and mistakes. In the 'Neubegründung', the results that go beyond the very basic theory treated in the notes for the Summer Semester of 1920 are stated without proofs. At the end of the paper, Hilbert writes:

> To conclude this first report, I would like to remark that P. Bernays has been of the greatest assistance to me in working out the ideas presented here; I owe it to his continued support that we now have throughout incontestable proofs. (*Hilbert 1922b*, 177.)[5]

In the lectures for the Winter Semester 1921/22, one finds different proofs for a reformulation of the theories. Four of the five inference rules are replaced by axioms, retaining *modus ponens* as the sole rule. (See Kneser's *Mitschrift* of the lectures for the Winter Semester of 1921/22, beginning with the report of the lecture for the 26 January 1922, below pp. 577ff.) In the official *Ausarbeitung* of the course for that semester, the consistency proof is dramatically refined and extended; for details, see the Introduction to Chapter 3.

In the second section we present Hilbert's considerations for the basic system in detail, bringing to the fore the genuine proof-theoretic insights and pointing out the oversights. The difficulties for overcoming these are discussed at the beginning of the third section of this Introduction; we give an argument for Hilbert's claims that involves a more complex ordering of proofs, but in contrast to the notes for the lectures in the Winter Semester 1921/22 retains Hilbert's local transformations of proofs. In the fourth section, the correctness result is used, as is done by Hilbert, to prove the consistency of the basic system and extend it to stronger theories. We end with some methodological and historical remarks.

2. Correctness proof, attempted.

The basic system of the manuscript consists of axioms for equality, successor (addition of 1), and the Z-symbol:

1.) $1 = 1$,
2.) $a = b \rightarrow (c = b \rightarrow a = c)$,
3.) $a = b \rightarrow a + 1 = b + 1$,
4.) $Z(1)$,
5.) $Z(a) \rightarrow Z(a + 1)$,
6.) $a = b \rightarrow (Z(a) \rightarrow Z(b))$.

---

[5] The original German reads:

> Zum Schluß dieser ersten Mitteilung möchte ich noch bemerken, daß mich bei der Durchführung und Ausarbeitung der hier dargelegten Ideen P. Bernays aufs wesentlichste unterstützt hat; seiner fortgesetzten Hilfe verdanke ich es, daß jetzt die einwandfreien Beweise durchweg vorliegen.

The version reprinted in Hilbert's *Gesammelte Abhandlungen* omits the second part of the remark. See *Hilbert 1935*, also 177.

The system has substitution rules that allow the replacement of individual variables in axioms by arbitrary terms and of propositional variables in inference rules by arbitrary formulas. There are no fewer than five inference rules for $\to$. Though they are not explicitly stated by Hilbert, they can be reconstructed from their use:

$$(\to 1)\frac{B}{A \to B} \qquad (\to 2)\frac{A \quad A \to B}{B} \qquad (\to 3)\frac{A \to B \quad B \to C}{A \to C}$$

$$(\to 4)\frac{A \to (B \to C)}{B \to (A \to C)} \qquad (\to 5)\frac{A \to (B \to C)}{(A \to B) \to (A \to C)}\,.$$

Hilbert does not explicitly define 'proof'; but the definition he intends is clear, both from the discussion in this manuscript and the presentations in the roughly contemporaneous lecture notes for the Summer Semester of 1920, and the 'Neubegründung' paper. A *proof* is a (finite) sequence of inferences such that each premise of an inference is either (an instance of) an axiom or the conclusion of an earlier inference.

Hilbert's basic system includes neither a sign for negation nor negative atomic predicates, thus consistency in the classical sense is a trivial result. Instead, Hilbert proves that an equation $k = l$ can be derived from 1.)–6.) only if k and l are the same sign or refer to the same sign.[6] The argument considers a 'hypothetically given concrete proof' of $k = l$, where 'k' and 'l' refer to different signs. Viewing the latter as a property E of proofs and using a principle of the form 'if there is a proof with property E, then there is a (relatively) shortest such proof', Hilbert assumes that the given proof cannot be shortened further.[7] This obviously presupposes a measure of 'length' for proofs. From Hilbert's indirect argument, it is clear that the number of occurrences of the conditional sign '$\to$' is taken as that measure. We now give a reconstruction of Hilbert's argument, and then discuss the problems with it.

Theorem. An equation $k = l$ can be proved only if 'k' and 'l' refer to the same number sign.

Proof. To obtain a contradiction, assume that we are given a proof of an equation of the form $k = l$, where the proof cannot be shortened and where 'k' and 'l' do not refer to the same sign, i. e., we have a (relatively) shortest proof of a false equation $k = l$. We distinguish cases as to how $k = l$ has been obtained.

[6]Those signs are not necessarily *Zahlzeichen* (number signs). 'Referring to a *Zahlzeichen*' means 'being shorthand for a *Zahlzeichen*'. For example, '2' is shorthand for the *Zahlzeichen* '$1 + 1$' and thus '2' *bedeutet* ('refers to' or 'means') $1 + 1$. For reasons of economy, we will use 'refer to' instead of 'are or refer to'.

[7]For this principle and its motivation, see the lecture notes for the Summer Semester of 1920, p. 38 (above, p. 366), and also *Bernays 1922a*, 18–19. We use the expression '(relatively) shortest proof' as shorthand for 'proof that cannot be shortened further' (with respect to a set of shortening rules, rules which are not explicitly stated by Hilbert).

First, we note that $k = l$ cannot come from Axiom 1.), as 'k' and 'l' would then refer to the same sign, contradicting our assumption. Thus, the proof must have length greater than 0, and the equation $k = l$ must be the conclusion of an inference rule. The only possible rule is ($\rightarrow$2), *modus ponens*, since all other rules have at least one $\rightarrow$-sign in their conclusion. So the last inference in the given proof must be of the form

$$ {\scriptstyle(\rightarrow 2)}\frac{\begin{array}{c}U\\ U \rightarrow k = l\end{array}}{k = l} \,. $$

U and $U \rightarrow k = l$ must have proofs already, and we distinguish now cases as to how $U \rightarrow k = l$ has been established.

1. $U \rightarrow k = l$ is (an instance of) an axiom. Only Axiom 3.) is of the right form. Thus U is an equation $k' = l'$ and k is $k' + 1$, l is $l' + 1$. 'k'' and 'l'' must refer to different signs; otherwise, '$k' + 1$' and '$l' + 1$' would also refer to the same sign, which was excluded. Thus, the proof of U establishes $k' = l'$, 'k'' and 'l'' refer to different signs, and that proof is shorter than the one given for $k = l$, contradicting our assumption.

2. $U \rightarrow k = l$ is the conclusion of ($\rightarrow$1). This inference is therefore of the form

$$ {\scriptstyle(\rightarrow 1)}\frac{k = l}{U \rightarrow k = l} \,. $$

Hence, $k = l$ is established already by a proof shorter than the given one; contradiction.

3. $U \rightarrow k = l$ is the conclusion of ($\rightarrow$2):

$$ {\scriptstyle(\rightarrow 2)}\frac{\begin{array}{c}V\\ V \rightarrow (U \rightarrow k = l)\end{array}}{U \rightarrow k = l} $$

where V and $V \rightarrow (U \rightarrow k = l)$ have already been established. This case is considered further after we have dealt with cases 4. and 5.

4. $U \rightarrow k = l$ is the conclusion of ($\rightarrow$3). In this case, Hilbert transforms the given proof by replacing the inferences

$$ {\scriptstyle(\rightarrow 3)}\frac{\begin{array}{c}U \rightarrow V\\ V \rightarrow k = l\end{array}}{U \rightarrow k = l} \qquad {\scriptstyle(\rightarrow 2)}\frac{\begin{array}{c}U\\ U \rightarrow k = l\end{array}}{k = l} $$

with

$$ {\scriptstyle(\rightarrow 2)}\frac{\begin{array}{c}U\\ U \rightarrow V\end{array}}{V} \qquad {\scriptstyle(\rightarrow 2)}\frac{\begin{array}{c}V\\ V \rightarrow k = l\end{array}}{k = l} \,. $$

He argues that the overall number of '$\rightarrow$'-signs has thus been reduced by 2, so that the replacement yields a shorter proof of $k = l$, contradicting the assumption.

5. $U \rightarrow k = l$ is the conclusion of ($\rightarrow$4) or ($\rightarrow$5), but this cannot be, as the conclusions of these rules are of a different form.

We now continue the analysis of case 3., and distinguish sub-cases as to how the formula $V \rightarrow (U \rightarrow k = l)$ could be established.

3.1. $V \to (U \to k = l)$ is (an instance of) an axiom, in which case only Axiom 2.) comes into question, which means we would have $k = m \to (l = m \to k = l)$. But now '$k$' or '$l$' cannot both refer to the same sign as 'm' does, for otherwise (by the transitivity and symmetry of the metatheoretical relation) 'k' and 'l' would also refer to the same sign. Whichever of 'k' and 'l' fails to refer to the same sign as 'm', in either case, we would have a shorter proof of a false equation, namely the proof of V or of U. This would contradict our assumption.

3.2. $V \to (U \to k = l)$ is the conclusion of ($\to$1). Then one can obtain a shorter proof of $k = l$ by dropping the central inference in the following three

$$(\to 1)\frac{U \to k = l}{V \to (U \to k = l)} \qquad (\to 2)\frac{V \quad V \to (U \to k = l)}{U \to k = l} \qquad (\to 2)\frac{U \quad U \to k = l}{k = l}\,.$$

That again contradicts our assumption.

3.3. $V \to (U \to k = l)$ is the conclusion of ($\to$2). This inference has to have $W \to (V \to (U \to k = l))$ as a premise. Hilbert argues that one cannot prove such a formula: it has three '*vorgeschaltete* [superposed]') '$\to$'-signs[8], but no axiom has more than two and no inference rule except ($\to$1) increases the number of 'superposed' '$\to$'-signs. Consequently, such a formula can be proved only by an application of ($\to$1) that is superfluous, since it could be 'cancelled' by a subsequent application of ($\to$2), thus contradicting our assumption.

3.4. $V \to (U \to k = l)$ is the conclusion of ($\to$3). Then the proof contains these inferences:

$$(\to 3)\frac{V \to W \quad W \to (U \to k = l)}{V \to (U \to k = l)} \qquad (\to 2)\frac{V \quad V \to (U \to k = l)}{U \to k = l} \qquad (\to 2)\frac{U \quad U \to k = l}{k = l}\,.$$

It can be shortened by using the following inferences:

$$(\to 2)\frac{V \quad V \to W}{W} \qquad (\to 2)\frac{W \quad W \to (U \to k = l)}{U \to k = l} \qquad (\to 2)\frac{U \quad U \to k = l}{k = l}\,.$$

That contradicts our assumption.

3.5. $V \to (U \to k = l)$ is the conclusion of ($\to$4). Then the original proof can be shortened by replacing

$$(\to 4)\frac{U \to (V \to k = l)}{V \to (U \to k = l)} \qquad (\to 2)\frac{V \quad V \to (U \to k = l)}{U \to k = l} \qquad (\to 2)\frac{U \quad U \to k = l}{k = l}$$

with

$$(\to 2)\frac{U \quad U \to (V \to k = l)}{V \to k = l} \qquad (\to 2)\frac{V \quad V \to k = l}{k = l}\,.$$

Again, we have a contradiction to our assumption.

[8] An occurrence $\to_1$ of the '$\to$'-sign is called '*vorgeschaltet*'/'superposed' to an occurrence $\to_2$ if they are found in a context like '$\ldots \to_1 (\ldots \to_2 \ldots)$'.

3.6. $V \to (U \to k = l)$ is the conclusion of $(\to 5)$. Then the original proof can be shortened by replacing

$${\scriptstyle (\to 5)}\frac{U \to (V_1 \to k = l)}{(U \to V_1) \to (U \to k = l)} \qquad {\scriptstyle (\to 2)}\frac{\begin{array}{c} U \to V_1 \\ (U \to V_1) \to (U \to k = l) \end{array}}{U \to k = l} \qquad {\scriptstyle (\to 2)}\frac{\begin{array}{c} U \\ U \to k = l \end{array}}{k = l}$$

with

$${\scriptstyle (\to 2)}\frac{\begin{array}{c} U \\ U \to (V_1 \to k = l) \end{array}}{V_1 \to k = l} \qquad {\scriptstyle (\to 2)}\frac{\begin{array}{c} U \\ U \to V_1 \end{array}}{V_1} \qquad {\scriptstyle (\to 2)}\frac{\begin{array}{c} V_1 \\ V_1 \to k = l \end{array}}{k = l}\ .$$

This again contradicts our assumption.

Thus, it seems that all cases have been shown to be impossible, therefore ruling out the existence of a provable equation $k = l$ where 'k' and 'l' refer to different signs. 'Q.E.D.'

The proof so presented has three kinds of genuine difficulties:

(1) The transformed configuration is *not* necessarily shorter than the original proof; indeed, in cases 4., 3.4., and 3.6 the transformed proof may in fact be longer. In these cases, V, W, or V_1 respectively occur more often in the 'shortened' than in the original proof. If these formulas contain a sufficiently large number of '$\to$'-signs, the transformed proofs may contain more of these signs than the original ones and, consequently, be longer.

(2) In cases 4., 3.4., 3.5., and 3.6, the transformed configuration is *not* necessarily itself a proof. Consider the case where U is the conclusion of an inference. If it is moved to a position earlier in the proof, its premise(s) may lose their justification. If, on the other hand, the whole proof of U is moved in order to guarantee that the resulting configuration is a proof, the proof which results may be much longer then the original.

(3) The iterated $(\to 2)$-applications in case 3.3 are *not* resolved satisfactorily. *Prima facie* longer chains of $(\to 2)$-applications can occur that have, for example, a $(\to 4)$-clause at their beginning before coming to a $(\to 1)$-clause. Therefore, neither the length of those chains can be determined in advance, nor the kind of $(\to 1)$-clause which comes first, nor which of the later *modus ponens* applications could be used to construct the desired contradiction. It seems that another inductive argument is necessary to confirm the claim that formulas with more than two 'superposed' '$\to$'-signs cannot be derived in Hilbert's calculus.

Hilbert must have considered the proof-theoretic transformations as strictly local ones, affecting only the 'end-piece' of the derivation. This is not correct, as we have seen, and the problem is addressed in the Winter Semester of 1921/22 in two stages, as discussed in the Introduction to Chapter 3. Here we just mention the second stage as recorded in Bernays's *Ausarbeitung*: the linear proofs are transformed first into proof trees (or '*Beweisfäden*') such that every formula (except the endformula) has exactly one successor. This is reflected in Hilbert's next foundational paper 'Die logischen Grundlagen der Mathematik' (*Hilbert 1923a*), a paper based on Hilbert's talk at the 1922 meeting of the *Gesellschaft Deutscher Naturforscher und Ärzte* in Leipzig and received for publication by the *Mathematische Annalen* on 29 September 1922.

If one adopts a tree representation of proofs, Hilbert's basic proof-theoretic idea can be preserved, though not without adding new and interestingly complex considerations. This is carried out in the next section.

3. Correctness proof, repaired.

The suggestion of using proof-trees instead of linear proofs seems to make the first problem mentioned above, the danger of transforming proofs into possibly longer ones, even worse. Consider, for example, case 3.6. above. Using the tree representation we have to consider a proof with these final inferences:

$$ {\scriptstyle(\to 2)}\dfrac{U \qquad {\scriptstyle(\to 2)}\dfrac{U \to V_1 \qquad {\scriptstyle(\to 5)}\dfrac{U \to (V_1 \to k = l)}{(U \to V_1) \to (U \to k = l)}}{U \to k = l}}{k = l} \; . $$

Hilbert's transformation turns this configuration into

$$ {\scriptstyle(\to 2)}\dfrac{{\scriptstyle(\to 2)}\dfrac{U \quad U \to V_1}{V_1} \qquad {\scriptstyle(\to 2)}\dfrac{U \quad U \to (V_1 \to k = l)}{V_1 \to k = l}}{k = l} \; . $$

This duplicates *the entire proof of U*, which might contain all kinds of inferences and formula complexities. The conclusion to be drawn is therefore: The number of '→'-signs occurring in a proof is not a suitable measure for the 'length' of proofs, and the same holds for every norm which simply counts the complexities of formulas or the numbers of signs used.

Consider now case 1. in Hilbert's proof, when a (→2)-clause has Axiom 3.) as its main premise; i. e., the final inference is of the form:

$$ {\scriptstyle(\to 2)}\dfrac{k' = l' \quad k' = l' \to k = l}{k = l} \; . $$

Hilbert's technique of shortening requires that in this case the subproof of $k' = l'$ be shorter than the whole proof; i. e., an application of *modus ponens* must raise the norm. But then, one faces a problem in case 4., when the last *modus ponens* is preceded by a (→3)-inference:

$$ {\scriptstyle(\to 2)}\dfrac{U \qquad {\scriptstyle(\to 3)}\dfrac{U \to V \quad V \to k = l}{U \to k = l}}{k = l} \; . $$

Here Hilbert's transformation leads to

$$ {\scriptstyle(\to 2)}\dfrac{{\scriptstyle(\to 2)}\dfrac{U \quad U \to V}{V} \qquad V \to k = l}{k = l} \; . $$

If the proof of U has the highest norm among the indicated subproofs, then the 'shortened' version is longer than the original one.

Finally, consider the case of inference $(\to 4)$ followed by *modus ponens* $(\to 2)$, i. e., case 3.5. The proof consisting of all the partial proofs but leaving aside the $(\to 4)$-inference must be shorter than the original one. So our norm must count the $(\to 4)$-occurrences.

We conclude that a norm is required which counts applications of $(\to 2)$ and $(\to 4)$, weighs $(\to 3)$ more heavily than $(\to 2)$, and is able to deal with the special situation of case 3.6., when a whole subproof is duplicated. We propose now a norm that satisfies these requirements and allows us to 'repair' Hilbert's proof, i. e., to prove the correctness assertion by making use of his proof transformations.

Call a part of a proof an $(n, 2)$-context if it consists of the inference $(\to n)$ and an application of *modus ponens* whose main premise is $(\to n)$'s conclusion:

$$\scriptstyle(\to 2)\displaystyle\frac{A \qquad \scriptstyle(\to n)\displaystyle\frac{\dots \quad (\dots)}{A \to B}}{B}\ .$$

A part of a proof is called an $(n, 2, 2)$-context, if it consists of the inference $(\to n)$ and two applications of *modus ponens* as follows:

$$\scriptstyle(\to 2)\displaystyle\frac{B \qquad \scriptstyle(\to 2)\displaystyle\frac{A \qquad \scriptstyle(\to n)\displaystyle\frac{\dots \quad (\dots)}{A \to (B \to C)}}{B \to C}}{C}\ .$$

For a given proof let $n_{5,2,2}$ denote the number of its $(5, 2, 2)$-contexts, n_3 the number of its $(\to 3)$-inferences, n_4 the number of its $(\to 4)$-inferences, and n_2 the number of the $(\to 2)$-inferences. We associate with a given proof the pair

$$\big(n_{5,2,2}, 2n_3 + n_4 + n_2\big)$$

as its *length.* A proof d_1 is *shorter than* d_2 in case its length is lexicographically before that of d_2. With this notion of length we go back to prove Hilbert's claim. First, we establish the following lemma via Hilbert's local operations.

Lemma. A proof containing any $(1, 2)$-, $(3, 2)$-, $(4, 2, 2)$-, or $(5, 2, 2)$-contexts can be shortened.

Proof. Consider the following cases:

1. There are $(5, 2, 2)$-contexts in the proof (i. e., $n_{5,2,2} \neq 0$). Then the proof can be shortened by replacing an uppermost such context

$$\scriptstyle(\to 2)\displaystyle\frac{A \qquad \scriptstyle(\to 2)\displaystyle\frac{A \to B \qquad \scriptstyle(\to 5)\displaystyle\frac{A \to (B \to C)}{(A \to B) \to (A \to C)}}{A \to C}}{C}$$

by

$$\scriptstyle(\to 2)\displaystyle\frac{\scriptstyle(\to 2)\displaystyle\frac{A \quad A \to B}{B} \qquad \scriptstyle(\to 2)\displaystyle\frac{A \quad A \to (B \to C)}{B \to C}}{C}\ .$$

n_2 is increased by at least one and the number of inferences stemming from the proof of A is doubled. But the transformed proof is shorter, as we chose an uppermost $(5, 2, 2)$-context and, consequently, $n_{5,2,2}$ is lowered by one.

2. There are $(1, 2)$-contexts in the proof. Consider an arbitrary one:

$$(\to 2)\frac{A \quad (\to 1)\dfrac{B}{A \to B}}{B}.$$

The subproof for the lower B can now be replaced by the subproof for the upper B, reducing n_2 by one. Since all other numbers stay the same or are also reduced, we have a shorter proof.

3. There are $(3, 2)$-contexts in the proof. Consider an arbitrary one:

$$(\to 2)\frac{A \quad (\to 3)\dfrac{A \to C \quad C \to B}{A \to B}}{B}.$$

By Hilbert's technique this can be transformed into

$$(\to 2)\frac{(\to 2)\dfrac{A \quad A \to C}{C} \quad C \to B}{B}.$$

This transformation increases n_2 by one and reduces n_3 by one. Since the latter counts twice in the length, and since all other numbers stay the same, we have arrived at a shorter proof.

4. There are $(4, 2, 2)$-contexts in the proof. Consider an arbitrary one:

$$(\to 2)\frac{A \quad (\to 2)\dfrac{C \quad (\to 4)\dfrac{A \to (C \to B)}{C \to (A \to B)}}{A \to B}}{B}.$$

It can be transformed into

$$(\to 2)\frac{C \quad (\to 2)\dfrac{A \quad A \to (C \to B)}{C \to B}}{B}$$

which is a shorter proof, since all subproofs occur already in the original proof and the number of $(\to 4)$-inferences is reduced by one. Q.E.D.

Now, we turn to the proof of Hilbert's theorem. The lemma allows us to prove Hilbert's claim, and the proof we give parallels the attempted proof.

Theorem. An equation $k = l$ can be proved only if 'k' and 'l' refer to the same number sign.

Proof. We proceed indirectly and assume that we have a proof of an equation $k = l$, where the proof cannot be shortened and where $k = l$ is a false equation.

We distinguish cases as to how the end-formula $k = l$ has been obtained. First, we note again that $k = l$ cannot come from Axiom 1.), since $k = l$ is a false equation. Thus, $k = l$ must be the conclusion of a $(\to 2)$-inference:

$$(\to 2)\frac{U \quad U \to k = l}{k = l}$$

and we distinguish subcases as to how $U \to k = l$ has been established.

1. $U \to k = l$ is an axiom. Only Axiom 3.) is of the right form, and we obtain a shorter proof of a false equation exactly as in case 1. of Hilbert's proof, contradicting our assumption.

2. $U \to k = l$ is the conclusion of a ($\to$1)- or ($\to$3)-inference. Then the proof contains a $(1,2)$- or $(3,2)$-context and can be shortened according to the Lemma, in contradiction to our assumption.

3. $U \to k = l$ can be neither the conclusion a ($\to$4)-inference nor of a ($\to$5)-inference, as those inferences have conclusions of a different form.

4. $U \to k = l$ is the conclusion of a ($\to$2)-inference. Then the proof ends in

$$\cfrac{U \qquad \cfrac{V \quad V \to (U \to k = l)}{U \to k = l}\,{\scriptstyle(\to 2)}}{k = l}\,{\scriptstyle(\to 2)}\ .$$

Now we distinguish subcases as to how $V \to (U \to k = l)$ has been obtained.

4.1. $V \to (U \to k = l)$ is an axiom. It can only be Axiom 2.) and, in analogy to the 'old' case 3.1., we get a proof that has one or two ($\to$2)-applications less than the original one, contradicting our assumption.

4.2. $V \to (U \to k = l)$ is the conclusion of a ($\to$1)-, ($\to$3)-, ($\to$4)- or ($\to$5)-inference. Then there are $(1,2)$-, $(3,2)$-, $(4,2,2)$-, or $(5,2,2)$-contexts, and the proof can be shortened according to the Lemma, contradicting our assumption.

4.3. $V \to (U \to k = l)$ is the conclusion of a ($\to$2)-inference. Then the proof ends in:

$$\cfrac{U \qquad \cfrac{V \qquad \cfrac{W \quad W \to (V \to (U \to k = l))}{V \to (U \to k = l)}\,{\scriptstyle(\to 2)}}{U \to k = l}\,{\scriptstyle(\to 2)}}{k = l}\,{\scriptstyle(\to 2)}\ .$$

Consider the uppermost ($\to$2)-application in the uninterrupted chain of ($\to$2)-applications in the proofs of the main premises. Its main premise cannot be an axiom since all the axioms have a different form. And, of course, it cannot be the conclusion of a ($\to$2)-inference, since we chose the uppermost ($\to$2)-inference in the chain. Thus it can only be the conclusion of a ($\to$1)-, ($\to$3)-, ($\to$4)- or ($\to$5)-inference. But then there are $(1,2)$-, $(3,2)$-, $(4,2,2)$-, or $(5,2,2)$-contexts and the proof can be shortened according to the Lemma, contradicting our assumption. Q.E.D.

4. Consistency proof and extensions.

Hilbert extends this result to an axiom system enlarged by two axioms:

1.′) $a = a$,

3.′) $a + 1 = b + 1 \to a = b$.

The considerations for extending the original proof are immediate. For the next step, Hilbert adds the axiom

7.) $a + 1 \neq 1$.

Explicit contradictions are now expressible and the correctness proof is extended to a consistency proof proper. Assume the system 1.)–7.) is inconsistent, i. e., proves an equation $k = l$ and the corresponding inequality $k \neq l$. Since inequalities can only come from Axiom 7.), they must be of the form $t + 1 \neq 1$. The equation is thus of the form $t + 1 = 1$ and must be provable from 1.)–6.). But that is impossible by the correctness theorem, for the terms '$t + 1$' and '1' refer to different number signs. Hence, the system 1.)–7.) is consistent.

In perfect accord with his general method of proceeding from simple to complex cases, Hilbert extends this consistency result step-by-step to stronger systems. The pattern is this: A proof of an inconsistency in the extended system is shown to be transformable into a proof that does not use the new axioms; since the smaller system is consistent, the extended one must be.

Consider Hilbert's extension of the axiom system given above by

8.) $a \neq b \rightarrow a + 1 \neq b + 1$,
8.′) $a + 1 \neq b + 1 \rightarrow a \neq b$,
9.) $a = b \rightarrow (a \neq c \rightarrow b \neq c)$.

If the axioms 8.)–9.) are used in a proof of an inequality $k \neq l$ which contradicts a provable equation $k = l$, then another inequality $k' \neq l'$ must occur earlier in the proof, a fact which follows from the conditional form of the axioms 8.)–9.). Hence the subproof of $k' \neq l'$ uses one application of 8.)–9.) fewer than does the original proof. Correspondingly, the proof of $k = l$ has to be extended (by steps in the basic system) in order to prove $k' = l'$ and thus to reobtain a contradiction.[9] This procedure can be carried out until a proof of a contradiction in the system without 8.), 8.′), and 9.) has been obtained; since this was shown to be impossible, the extended system cannot prove a contradiction.

The consistency proof for the next axiom considered by Hilbert, Axiom 10.), is of special interest, for Hilbert here introduces a completely new method, namely the systematic substitution of terms. The axiom reads:

10.) $1 + (a + 1) = (1 + a) + 1$.

This allows for the proof of equations of the form $k = l$, where 'k' and 'l' do not refer, *prima facie*, to the same number sign. Thus, the core of the original

---

[9] If an application of 8.) has to be eliminated, k is $k' + 1$ and l is $l' + 1$. So one can use an application of Axiom 3.′), that is, $k' + 1 = l' + 1 \rightarrow k' = l'$, and *modus ponens*. In case of 8.′), use Axiom 3.), that is, $k = l \rightarrow k + 1 = l + 1$, and *modus ponens*, since $k + 1$ is k' and $l + 1$ is l'.

Axiom 9.) is a little more involved. Let an instance of this be $m = k \rightarrow (m \neq l \rightarrow k \neq l)$, and assume $m = k$ and $m \neq l$ proved previously. Then the proof of $m \neq l$ is the natural candidate for the 'shorter' proof in the sense of using one application of 8.)–9.) fewer than the original proof of a contradiction. So k' is m and l' is k. But how then to conclude the contradiction? The proof of $m = k$ could be extended by an application of Axiom 2.): $m = k \rightarrow (l = k \rightarrow m = l)$, yielding $l = k \rightarrow m = l$. But in order to infer $m = l$ from that, one would need $l = k$, while in the assumption of the original contradiction one only has $k = l$. Here Hilbert uses the symmetry of equality, that is, the principle (1) below (p. 391), and therewith Axiom 1.′).

correctness proof, that $k = l$ can be derived only if 'k' and 'l' refer to the same number sign, has to be amended. The left-hand side of the equation 10.) involves terms of the form $1 + (t + 1)$. If one could transform the given proof of a contradiction into a proof which does not contain terms of this form, then 10.) is not applied in this proof. Consequently, the transformed proof would establish a contradiction from 1.)–9.) alone, which was shown to be impossible.

How can the terms $1 + (t + 1)$ be removed from a given proof of a contradiction? Hilbert's idea is to replace them either by $(1 + t) + 1$, making 10.) a version of Axiom 1.$'$), or by an arbitrary smaller term, say, 1. Both replacements preserve the proof as a proof, provided that they are executed carefully.

Consider a list of all terms of the form $1 + (t + 1)$ which occur (maybe as sub-terms) in the proof. Then take a t for which $1 + (t + 1)$ occurs in the proof, but not $1 + ((t + 1) + 1)$, and distinguish two cases.

First case: 10.) occurs for $a : t$.[10] Then replace in the whole proof all occurrences of $1 + (t + 1)$ by $(1 + t) + 1$. Terms of the form $1 + (t + 1)$ can occur in the proof only in two ways: as (parts of) terms which are substituted for the variables in the axioms, *or* as the left-hand side of Axiom 10.) when t is substituted for a. Under the replacement of $1 + (t + 1)$ by $(1 + t) + 1$, every instance of an axiom remains a (different) instance of the same axiom, except that instances of Axiom 10.) for $a{:}t$ are transformed into $(1+t)+1 = (1+t)+1$, i. e., into an instance of 1.$'$). So, after the replacement, the proof remains a (slightly different) proof of a (possibly different) contradiction. But it contains no terms of the form $1 + (t + 1)$ anymore. Instead, it contains more terms of the form $1 + t$, but this form was already present in the original proof as the right-hand side of the instance of Axiom 10.) for $a{:}t$.

Second case: 10.) does not occur for $a{:}t$. Then simply replace all occurrences of $1 + (t + 1)$ by 1. The proof remains a proof, but it has one type fewer of $1 + (t + 1)$-terms.

In each case, one has a new proof of a (possibly new) contradiction which contains one type of $1+(t+1)$-terms fewer than the original proof. By iterating this procedure, one eliminates all terms of the form $1 + (t + 1)$ from the proof and obtains a proof of a contradiction from 1.)–9.) alone, which is impossible according to the earlier consistency result.

5. Historical and methodological remarks.

In the introductory remarks, it was emphasized that this manuscript is intermediate between the lectures from the Summer Semester of 1920 (and thus also the ideas from *Hilbert 1905b*) and the essay 'Neubegründung der Mathematik', published in 1922.

---

[10] By '$a{:}t$', Hilbert indicates the substitution of the term t for the variable a.

The crucial observation in the 1905/1920 argument is that formulas occurring in derivations have bounded complexity.[11] Once that bound is obtained, the argument for correctness proceeds by first assuming that we have a relatively shortest proof of a false equation. Then one distinguishes cases as to how that equation can have been established, so obtaining a contradiction in all cases. The case of longer and longer (→2)-chains is ruled out by the complexity bound. Other cases, in which one can construct a shorter proof from the given one, are ruled out since a relatively shortest proof is being considered. This principle, assuming the existence of a relatively shortest proof, functions as a metatheoretical induction principle, and appears also in the correctness proof given above.[12] In 'Neubegründung', however, Hilbert abandons it in favour of a more perspicuous principle, a form of the least number principle.[13] This principle allowed Hilbert to present in the 'Neubegründung' a smoother version of the basic consistency proof from the Summer Semester of 1920.

At the end of the 1920 lectures, a programme is sketched for expanding the basic system by adding arithmetical axioms. Hilbert also realizes that 'one has to expand the forms of inference so that they suffice for proving the arithmetic theorems' (p. 45 of the notes, above p. 371). Both kinds of expansions are carried out in the manuscript. The procedures for handling arithmetic axioms are intricate and subtle, but they do not raise the fundamental issue raised by the new forms of inference.[14]

In the manuscript, the rules (→1)–(→5) for the conditional are introduced. Hilbert remarks that he could have restricted the calculus to the rules (→1) and (→2) and that in this way the consistency proof would have been simplified (p. II). The crucial addition to *modus ponens* is therefore (→1). It is required to bring to bear the new arithmetic axioms in the proof of equations, and it is because of this rule that the complexity of formulas in derivations is no longer bounded.

The addition of the other rules is seemingly motivated by the desire to prove longer chains of conditionals of equations, such as

(1) $a = b \to b = a$,

(2) $a = b \to (b = c \to a = c)$,

---

[11] See the Summer Semester lectures, pp. 40–43 (above, pp. 367ff.)

[12] For this principle, see p. 38 of the notes for the Summer Semester of 1920, above p. 366.

[13] See *Hilbert 1922b*, pp. 171–172.

[14] Later Hilbert abandoned rules for propositional logic in favour of axioms. For instance, in Kneser's *Mitschrift* of the lectures given in the Winter Semester of 1921/22, he says (p. 20):

> We need inferences other than [*modus ponens*], for example
>
> $$(\to)\frac{\begin{array}{c}\gamma \to \psi \\ \psi \to \chi\end{array}}{\gamma \to \chi}\,.$$
>
> But we only want this [i. e., *modus ponens*], and we add the others in the form of logical axioms, five in number.

For the original German, see p. 581 below.

(3) $a = b \to (b = c \to (c = d \to a = d))$.[15]
Since the formal system has no general substitution rule and no general equality axiom (schema), these chains may have been intended to make formal proofs more perspicuous. In any case, Hilbert proposed to prove (2) and (3) using the logical theorem: if $S \to (U \to V)$ and $S \to (V \to W)$ are provable, so is $S \to (U \to W)$.[16] The proofs of this logical theorem, and then of (2) and (3), make use of *all five* $\to$-rules. Hence, the above chains of equations may well have motivated Hilbert to introduce the rules; he even mentions that he could have taken the theorems (1)–(3) as axioms instead of adding the rules ($\to$3)–($\to$5).

As was mentioned (above, p. 379), the Editors of Volume 3 of Hilbert's *Gesammelte Abhandlungen* assert that the 'Neubegründung' reflects a transition from a first to a second stage in the development of proof theory. The first stage, represented on pp. 168–174 of the paper, concerns the treatment of the basic system and was discussed above. The second stage, represented on pp. 174ff., is described by the Editors as dealing with 'the intended formalism' of 'Neubegründung'.[17] Hilbert returns here to a formalism with only two inference rules, namely *modus ponens* and substitution, but with many logical and arithmetical axioms, including an induction schema.[18] The logical

[15]It is interesting to note that Hilbert explicitly called these sentences, including (1), 'examples of provable theorems' (see p. II of the manuscript, p. 399 below). Thus, he obviously considered the reflexivity of equality, necessary for the proof of (1), as an axiom (Axiom 1.′).

[16]Here is a proof of the logical theorem: From $S \to (U \to V)$ infer by ($\to$4) $U \to (S \to V)$, and from $S \to (V \to W)$ infer by ($\to$5) $(S \to V) \to (S \to W)$. Applying ($\to$3) to the two conclusions gives $U \to (S \to W)$; this finally yields $S \to (U \to W)$ by ($\to$4).

We can now prove (1)–(3) as follows.

Proof of (1): $b = b \to (a = b \to b = a)$ is an instance of Axiom 2.). *Modus ponens* applied to this and $b = b$, which is an instance of Axiom 1.′), yields (1).

Proof of (2): From (1) $b = c \to c = b$ infer by ($\to$1) $a = b \to (b = c \to c = b)$. Applying the logical theorem to this and to $a = b \to (c = b \to a = c)$, which is a version of Axiom 2.), leads to (2).

Proof of (3): Apply the logical theorem to $a = b \to (b = c \to a = c)$, which is a version of Axiom 2.), and to $a = b \to (a = c \to (c = d \to a = d))$, which is a ($\to$1)-weakening of $a = c \to (c = d \to a = d)$. This gives (3).

[17]See the reprinting of *Hilbert 1922b* in *Hilbert 1935*, p. 168, n. 2.

[18]In the manuscript there is one passage where Hilbert claims to have *proved* the principle of complete induction, at least in a special case. (He says: 'Hier ist zum ersten Mal das Prinzip d⟨er⟩ vollständigen Induktion wenigstens im besonderen Fall wirklich bewiesen.') In fact, what he proves is only that the axiom $1 + a = a + 1$ is consistent with the other axioms. We are not sure what exactly Hilbert means here, but it seems plausible that 'having proved' means something like 'having proved consistent'. This way of reading what Hilbert says is suggested by a remark in the lectures for the Summer Semester of 1920. There Hilbert presents a procedure for the elimination of instances of the induction rule; he claims that this procedure provides both a relative consistency proof for induction and 'in a certain sense' a proof of the induction rule. The latter means, Hilbert explains, that all equations derivable with induction can also be derived without. (For the relevant passage, see p. 45 of the lecture notes, p. 370 above.) Even with this understanding, it remains to be explained how $1 + a = a + 1$ can be called 'induction' in any sense. We conjecture that

axioms are:
1. $A \to (B \to A)$,
2. $(A \to (A \to B)) \to (A \to B)$,
3. $(A \to (B \to C)) \to (B \to (A \to C))$,
4. $(B \to C) \to ((A \to B) \to (A \to C))$.
1., 3., and 4. correspond to ($\to$1), ($\to$4), and ($\to$3), whereas the contraction axiom 2. does not strictly correspond to rule ($\to$5) being slightly stronger.[19] Hilbert claims that a consistency proof for this formalism has been obtained, but does not sketch such a proof in the paper. The Editors state that the result is correct only when the universal quantifier is removed from the language and induction is formally given by a rule (obviously for quantifier-free formulas). They also remark that the principle of definition by primitive recursion can be added and is actually required to make the system adequate for formalizing mathematical practice.[20]

Hilbert's investigations in this manuscript treat for the first time a fragment of arithmetic that incorporates propositional logic as given by the rules for the conditional. It is this 'inferential expansion' that requires new techniques. These techniques are not used in the later lectures for the Winter Semester of 1921/22 to prove the consistency result as we have formulated it here. In this sense, the manuscript is transitional. However, there is a deeper sense in which the manuscript and the 'Neubegründung' paper are transitional. The object-theory in this manuscript is logically weak (a fragment of intuitionistic logic), and that holds also for the 'intended formalism' of 'Neubegründung', where the axioms for negation are of a very restricted form, i. e., treat only of arithmetical inequality.[21] Indeed, Hilbert emphasizes the significance of this constructive aspect of the formalism. (See *Hilbert 1922b*, p. 173.) However, in Bernays's *Ausarbeitung* of Hilbert's lecture course for the Winter Semester of 1921/22 the full axioms for negation are used.[22] In this *Ausarbeitung*, proof theory is introduced as 'Hilbert's Proof Theory', and finitist mathematics is pursued for the first time, and the distinction between object- and meta-theory is sharpened to a distinction between an object the-

---

he speaks of induction 'in gewissem Sinne' in order to indicate that $1 + a = a + 1$ cannot be proved without induction.

[19]See *Hilbert 1922b*, 175. The same axioms are added in the lectures from the Winter Semester of 1921/22; see Bernays's *Ausarbeitung*, the second p. 3, where they are called 'Axiome der Folge' (below, p. 494, and Kneser's *Mitschrift* for these lectures, pp. 20–21; below, p. 582).

[20]See the reprint of *Hilbert 1922b* in *Hilbert 1935*, 176, nn. 1,2.

[21]In *Hilbert 1922b*, p. 175, Hilbert uses $a \neq a \to A$ (short for $a = a \to (a \neq a \to A)$) and $(a = b \to A) \to ((a \neq b \to A) \to A)$ as axioms for negation.

[22]The *Ausarbeitung* consists of two parts, the second of which can be seen as a supplement written by Bernays, and in this there are the following two axioms for negation: $A \to (\overline{A} \to B)$ and $(A \to B) \to ((\overline{A} \to B) \to B)$. See Bernays's *Ausarbeitung*, the second p. 3, p. 494 below. Kneser's *Mitschrift* for the 1921/22 lectures (p. 22; below, p. 582) lists for negation just the same arithmetical inequality axioms as are set out in n. 21 above. For further discussion, see the Introduction to Chapter 3, p. 422, especially n. 8, and also *Bernays 1935*, 203.

ory with full classical logic and a meta-theory in which only restricted finitist principles are used. These considerations, both the theoretical reflections and the technical work, are presented in Hilbert's Leipzig lecture 'Die logischen Grundlagen der Mathematik'. This paper presents, according to Bernays, the principled basis for proof theory, in which Hilbert's proof-theoretic programme is fully articulated for the first time:

> With the formulation of proof theory which we encounter in the Leipzig lecture, the basic form of its structure had been achieved.[23]

6. Note on the Text.

The manuscript is part of a folder in the Hilbert *Nachlaß*, namely *Cod. Ms. D. Hilbert 602*. The folder was entitled 'Über Axiome' for catalogue purposes, but this is neither Hilbert's title, nor an appropriate one; it is taken from the heading ('Axiome') of a list of axioms with which the first page happens to begin. The folder contains fourteen sheets; only eight of them constitute the document presented here, and this document has no official title. Two of the sheets (numbered 9 and 10 in the folder) are written on the back of galley proofs for papers published in the *Mathematische Annalen* only in 1931, and thus, if the rough dating of the material in the present manuscript is approximately accurate, clearly stem from over a decade later. The notes Hilbert has written on these sheets concern a sketch of what he calls a 'Beweis des ε-Axioms'. Sheet 9 is actually written on the back of a proof-sheet for the article eventually published as *Pietrkowski 1931*, and sheet 10 for the article eventually published as *Hamburger 1931*. It is interesting to note that these articles appeared in exactly the same volume of the *Annalen* as the paper *Cauer 1931*, one of the proof-sheets of which was used for the additional remarks inserted in the one copy we have of the *Ausarbeitung* of the 'Probleme der mathematischen Logik' lectures from 1920. (See the Description of the Text for the lectures 'Probleme der mathematischen Logik', above, p. 376.) It is hard to resist the conclusion that the same stack of scrap paper was used.

As is typical for Hilbert's manuscripts, the document is littered with deletions (often of whole passages), and additions, often, but not always, as replacement for deleted text. These additions are written in no systematic way, being sometimes interlineated, sometimes written in the margins or any other space available on the page, and frequently with no clear indication of where they belong, if anywhere.[24] Nevertheless, an attempt has been made by the Editors to reconstruct the text as a readable document representing the flow

---

[23] *Bernays 1935*, 204. In the original German, the passage reads:

> Mit der Gestaltung der Beweistheorie, die uns in dem Leipziger Vortrag entgegentritt, war die grundsätzliche Form ihrer Anlage erreicht.

For more details concerning these developments, see the Introduction to Chapter 3.

[24] To give a flavour of the appearance of Hilbert's document, we have included a photographic reproduction of Sheet 1; see p. 415.

of the central ideas, while at the same time presenting all the essential changes for the reader to assimilate and judge.

Wilfried Sieg and Christian Tapp