papers/hilbert/paper.logic
created 08-17-2001
revised last 08-28-2001

*Hilbert's New Problem**

Ruediger Thiele

Karl-Sudhoff-Institut fur Geschichte der Medizin und Naturwissenschaften
Universitaet Leipzig

and

Larry Wos

Mathematics and Computer Science Division
Argonne National Laboratory
Argonne, IL 60439

Abstract

Throughout the twentieth century, the worlds of logic and mathematics were well aware of Hilbert's twenty-three problems and the challenge they offered. Although not known until very recently, there existed yet one more challenge offered by Hilbert, his twenty-fourth problem. This problem focuses on finding simpler proofs, on the criteria for measuring simplicity, and on the ''development of a theory of the method of proof in mathematics in general''. Of the three themes of Hilbert's twenty-fourth problem, the first two are central to this article. We visit various areas of logic, showing that some of the studies of the masters are indeed strongly connected to this newly discovered problem. We also demonstrate that the use of an automated reasoning program (specifically, W. McCune's OTTER) enables one to address this challenging problem. We offer questions that remain unanswered.

Keywords: Hilbert's twenty-fourth problem, proof simplification, automated reasoning

1. A Significant Find

Hidden in Hilbert's many notebooks in Goettingen, Germany, a treasure was found. Specifically (in Notebook Cod. ms. Hilbert 600:3) the treasure is a problem—Hilbert's twenty-fourth—not presented at his 1900 Paris lecture nor in the printed versions of that lecture [Thiele2001]. (In Section 3, we focus directly on Hilbert himself and on his writings.)

This newly discovered problem focuses on finding simpler proofs, on the criteria for measuring simplicity, and on the ''development of a theory of the method of proof in mathematics in general''. Of the three themes, this article addresses the first two. Also featured here is the strong connection to this problem of some of the research of logicians that include C. A. Meredith, A. Prior, and I. Thomas. We then demonstrate the value of addressing criteria of Hilbert's twenty-fourth problem with heavy reliance on W. McCune's automated reasoning program OTTER [McCune1994]. All of the proofs we find are Hilbert-style axiomatic, employing one or more specific inference rules; never do we rely on induction or metaargument. We might, for example, rely solely on condensed detachment, a rule employed by many logicians. On the other hand, if the problem emphasizes equality, we employ

*This work was supported in part by the Mathematical, Information, and Computational Sciences Division subprogram of the Office of Advanced Scientific Computing Research, U.S. Department of Energy, under Contract W-31-109-Eng-38, and in part by the German Academy Leopoldina, Halle, under Contract LPD 1996.

paramodulation, a rule that enables the program to treat equality as ''understood''.

To encourage further research focusing on this intriguing Hilbert problem, we offer questions that remain unanswered. We also cite various successes that address Hilbert's concern for the discovery of simpler proofs.

2. Criteria of Proof Simplicity

Compared with the properties of proofs, more familiar to many are the properties of axiom systems. For example, of substantial concern are the number of members in an axiom system, the total number of symbols, the maximum number of distinct variables, and independence. Indeed, J. Lukasiewicz spent three years of research culminating in his discovery of his 23-letter single axiom (the following) for two-valued sentential (or propositional) calculus [Lukasiewicz1970].

CCCpqCCCNrNstrCuCCrpCsp

Discoveries of this type naturally led to questions about the existence of a shorter single axiom for two-valued logic. C. A. Meredith later did in fact find a *simpler* single axiom [Meredith1953], the following.

CCCCCpqCNrNsrtCCtpCsp

Not only is the Meredith axiom shorter, 21 symbols versus 23, but it also is simpler in that it relies on five distinct variables rather than on six.

In the context of Hilbert's twenty-fourth problem, for proofs, two related criteria merit examination, proof length (the number of deduced steps) and variable richness (the maximum number of distinct variables that occur in a deduced step). In particular, all things being equal, the shorter the proof, the simpler the proof. Similarly, the less variable richness, the simpler the proof. We thus have two criteria of proof simplification pertinent to the new Hilbert problem.

For an example of the logician's interest in proof refinement with respect to length, one need only begin with Meredith. In the early 1950s, he published a 30-step condensed detachment proof that deduces a 3-basis for *C*5 from his single axiom for that area of logic [Lemmon1957]. The area *C*5 is the implicational fragment of *S*5, itself a modal logic. A few years later, (apparently with some satisfaction) Meredith published a second proof of this theorem, one of length 23 [Meredith1964].

With Prior, Meredith later published a 33-step proof that abridges the Lukasiewicz proof for his shortest single axiom for the implicational fragment of two-valued logic [Meredith1963]. With the program OTTER—and addressing the new Hilbert problem, although its existence had not yet been declared—we found a further abridgment, a proof of length 32.

A quite different criterion of proof simplification concerns lemma avoidance. Somewhat reminiscent of seeking an independent set of axioms, one might seek a proof that does not rely on thought-to-be-indispensable lemmas. The avoidance of various lemmas can contribute in an important way to simplicity of proof. For example, a search of the literature in the context of proving the dependence of one of the five Lukasiewicz axioms for infinite-valued sentential calculus suggests that three lemmas are indeed crucial, numbered by Rose and Rosser 2.22, 3.5, and 3.51. With OTTER, we have found a far simpler proof, one avoiding the cited three lemmas. The proof has length 30, shorter than any offered by the literature. The cited proof also avoids the use of any double-negation terms. A double-negation term is a term of the form $N(N(t))$ for some term t, where the function N denotes negation. The proofs in the literature rely on double negation. (Note that we do not have in mind in any way the double-negation laws that occur, say, in the intuitionistic calculus; here we are concerned exclusively with a class of term.)

We thus come to another criterion of proof simplification that might indeed have appealed to Hilbert, namely, the avoidance of a class of term. Logicians are in fact sometimes concerned with such avoidance, seeking proofs in which no deduced formulas contains as a subformula a formula of the form $i(t,t)$ for some term t, where i denotes implication (sometimes denoted by C, as in the Lukasiewicz and Meredith cited single axioms).

Still in the context of criteria of simpler proof, we have the analogue of size that is studied for axiom systems—the total number of symbols in the system or basis. Specifically, as D. Ulrich has

suggested, one can seek proofs whose size is strictly less than that of any proof in hand, where size measures the total number of symbols present in the deduced steps of a proof. Somewhat related to size is the measure of simplicity that is concerned with formula complexity. The formula complexity of a proof is *k* if and only if (1) there exists a deduced step in the proof relying on exactly *k* symbols, and (2) no deduced step in the proof relies on strictly greater than *k* symbols. (Similarly, for an axiom system or basis, one can focus on its complexity, the longest member in symbol count.)

Clearly other measures of simplicity might merit study. We have studied those specifically cited and will give examples of success in Section 4. We note the perhaps obvious: When one addresses a chosen criterion of simplicity, one often must accept that other criteria may suffer. Quite likely, this fact explains in part why Hilbert did not include this twenty-fourth problem in his Paris lecture, for it is indeed difficult to make precise.

3. On Hilbert and His Writings

Much of contemporary mathematics is foreshadowed in Hilbert's twenty-three problems, given in the published version of his 1900 Paris talk to the International Mathematical Congress. Near the end of 1899, he was invited to give this talk, but wavered between responding to an 1897 Poincare talk and presenting unsolved problems. In fact, not until the spring of 1900 did Hilbert decide to present problems concerning the future of research in mathematics. In July, he surprised his friends Hurwitz and Minkowski with printed versions of his forthcoming talk, a talk in which but twelve of the twenty-three problems were actually offered. The preparation of so many significant problems is indeed impressive in that Hilbert was lecturing ten hours a week.

Among the undated entries in the Notebook is a statement saying that Hilbert had in mind to present a twenty-fourth problem in Paris concerning the simplicity of proofs, as expressed with the excerpt we highlight (trans. Thiele 2000).

> The twenty-fourth problem in my Paris lecture was to be: Criteria of simplicity, or proof of the greatest simplicity of certain proofs.
>
> Develop a theory of the method of proof in mathematics in general. Under a given set of conditions there can be but one simplest proof.

That problem does not appear in any of the published versions of his 1900 talk. However, the twenty-fourth problem is recorded in his ''Mathematisches Notizbuch'', preserved in the Niedersaechsische Staats- und Universitaets-bibliothek Goettingen, Handschriftenabteilung (Cod. ms. D. Hilbert 600).

In short, Hilbert asks for the *simplest proof of any theorem*. That the entry was made some time after the Paris talk is remarkable: there are almost no entries dating before the Paris talk in August 1900 that were used for the preparation of the talk. Moreover, although Hilbert omitted the twenty-fourth problem in the lecture and the subsequent paper, his Notebook entry indicates that in his logical and foundational research, Hilbert did not cancel the question of simplest proofs. Indeed, the problem continued to intrigue him. We find it or versions of it in later works. Then, in 1917, Hilbert gave a lecture in Zurich, Switzerland, in which he mentioned examples (geometric constructions and invariant theory, especially syzygies) he had added to the idea of simplest proof in the Notebook. We offer one final bit of evidence concerning the importance to Hilbert of the omitted problem. At the end of his life, looking back and revisiting his research, Hilbert made and inserted a short ''index'' into his notebooks. Among the few key words on this one-page index, we again find the twenty-fourth problem.

We thus see that, throughout his life, Hilbert regarded the problem of simplest proof as an important one. Further, that entry in the Notebook attests to the fact that his proof theory goes back as far as the turn of the last century—almost two decades earlier than historians have believed up to now. It is Hilbert's question whether one proof is simpler than another. For the first time in the history of mathematics, the properties of proof were made an object of mathematical investigation; see Section 2. Moreover, Hilbert supplied the initiative to start research in a new field, a field now called metamathematics.

A natural question asks why Hilbert did not present this problem in Paris or later elsewhere. We suspect there exist two principal reasons. First, in Paris he was under the pressure of time, and he did not intend to give a complete overview of open problems. Second, to formulate the intuitively obvious

question focusing on simplest proofs in precise mathematical concepts is not at all easy. Because of the difficulty of defining the concept, Hilbert probably postponed the ''declaration'' of the general problem of simplicity and, instead, did some research on simplicity in specific fields.

4. Successes in Addressing Hilbert's Twenty-Fourth Problem

At this point, we turn to successes in proof simplification of the type that (almost certainly) Hilbert would have found satisfying. The context of simplification varies, as does the area of logic from which the theorem was taken. Often within the cited success is an implied question that remains open.

Each of the results we cite was obtained by relying heavily on the automated reasoning program OTTER. We, therefore, occasionally comment on how that program provided invaluable assistance. We also sometimes briefly touch on the methodology that was employed to reach the desired objective.

We begin with two-valued sentential calculus, sometimes known as classical logic. Our target was Meredith's proof of his single axiom (given in Section 2), a proof that completes by deducing the Lukasiewicz three-axiom system. The Meredith proof consists essentially of 41 applications of condensed detachment (has length 41), relies on formulas in which seven distinct variables occur, and contains seventeen double-negation formulas. By using other proofs (that will be cited) to direct OTTER's search, we discovered a 38-step proof with variable richness seven and containing thirteen double-negation formulas. (In our study, we also relied upon OTTER's offering of *ancestor_subsumption*, a procedure that automatically compares derivation lengths to the same conclusion, preferring the strictly shorter.) The program's attack and resulting proof are counterintuitive in the sense that the most-difficult-to-prove Lukasiewicz axiom (sometimes known as suffixing) is proved second and used to prove a less-difficult-to-prove axiom. In contrast, Meredith's proof proceeds in the order one would expect, proving suffixing last.

To find the cited 38-step proof, two proofs were used to guide the search (the proofs were first reported in [Fitelson2001]). The first proof, one of length 50, was discovered when we sought a proof with variable richness six. (OTTER offers the parameter *max_distinct_vars* to which one assigns a value that places an upper bound on the number of variables a newly deduced conclusion can rely upon.) The second was a proof of length 51, found when we decided to avoid double negation. (OTTER offers *demodulation*, ordinarily used for canonicalization and simplification, which we use to purge newly deduced conclusions of an unwanted type.) Apparently crucial to our success was the treatment of the deduced steps of the two proofs as most attractive symbol patterns, where their variables were treated as indistinguishable (resonance). (OTTER offers the *resonance* strategy [Wos1995] for this purpose; a powerful alternative is R. Veroff's *hints* strategy [Veroff1996]).

Since the first use of condensed detachment on Meredith's single axiom as both major and minor premiss yields a formula relying on five distinct variables, the limiting case in the context of variable richness is five. Such limiting cases were apparently of interest to Hilbert. We did in fact find a 69-step proof with that richness.

In addition to the cited explicit means for proof simplification, OTTER offers various methodologies, many of which rely upon the use of one or more strategies. For example, for term simplification we have found quite effective a methodology (relying on demodulation) that instructs the program to block the use of one or more chosen steps of the proof in hand. For proof refinement in the context of length, in addition to constantly using ancestor subsumption, we employ a methodology similar to that for term simplification and, more recently, a strategy called *cramming* [Wos2001]. The latter attempts to force many proof steps of one member of a conjunction into proofs of the other members to produce a shorter proof of the whole, recognizing that the individual subproofs will likely be longer than need be if attacked separately. An unexpected (to us) phenomenon we have encountered is that, occasionally, a shorter proof is found such that all of its formulas are among those of the longer proof under study.

Still with Meredith at center stage, we next turn to $C5$. He proved his single axiom for this area of logic by deducing one of his bases, a 3-basis, with a proof of length 23, variable richness 7, and size 358. Our studies, which did not rely on knowledge of his success, yielded three new 23-step proofs, the best of which has variable richness 6 and size 358. Further studies strongly suggest that in fact no proof of length strictly less than 23 exists; however, the improvement in the context of variable richness

is satisfying. At this time, we note that neither OTTER itself nor our formulation of methodologies offers the researcher a direct attack on proof size.

For a most dramatic refinement in the context of the criterion of proof size, we draw from work on equivalential calculus. The first proof that showed the formula *XHN* to be a single axiom (by S. Winker) has size 8120. More than a decade later, OTTER (relying on various methodologies) produced a proof of size 528. The Winker proof has length 159, whereas the later proof has length 19. The Winker proof has formula complexity equal to 103 (a formula requiring 103 symbols to present), whereas the later proof has formula complexity equal to 35. (OTTER offers a parameter, max_weight, that one can employ to prevent the retention of a newly deduced formula or equation whose symbol count exceeds the value assigned to that parameter.)

One interested in addressing aspects of Hilbert's twenty-fourth problem might find it piquant to learn that OTTER already offered explicit means for proof simplification almost a decade before the problem was discovered. Specifically, such means that include direct approaches to controlling formula complexity, variable richness, and the like were not designed originally with Hilbert in mind or, more generally, with proof refinement in mind. Instead, their original purpose was that of controlling the search, in part by restricting it, and in part by directing it. (Such direct approaches, however, are far from adequate to control the search; indeed, rather early on in the automation of logical reasoning it was recognized that strategy was required if substantial assistance was to be provided for attacking open questions and deep problems.) Summing up, the researcher intent upon addressing this new Hilbert problem experiences a bit of more-than-good fortune in the form of parameters, options, and methodologies.

5. Broadening the Attack on Hilbert's Twenty-Fourth Problem

We intend to continue our assault on various aspects of the new Hilbert problem. In that regard, we welcome contributions by e-mail, wos@mcs.anl.gov. For example, if some researcher has a proof in hand and wishes to find a shorter proof, or a proof with less variable richness, or a proof that addresses some other criterion of refinement, we are eager to begin an attack on the problem. Collaborations might prove most profitable.

Our top choices for questions concerning proof length are currently two. In the context of classical logic, does there exist a proof of length 37 (applications of condensed detachment) or less that deduces the Lukasiewicz three-axiom system from Meredith's single axiom? Does there exist a proof of length 29 or less that shows that (what we call) the fifth of Lukasiewicz's five axioms for infinite-valued sentential calculus is dependent on the other four?

CCCpqCqpCqp (axiom 5)

We know of more that ten proofs of length 30 for this theorem. We note that some of the proofs that OTTER has discovered quite likely would have eluded the unaided researcher for a long, long time, perhaps because of formula complexity, perhaps because of their unintuitive nature, or because of a variety of other factors. We draw this conclusion in part on the nature of some of our results—for example, the finding of proofs avoiding the use of thought-to-be-indispensable lemmas and the avoidance in many, many cases of double negation—and in part on some of the impressive data produced by our experiments. As evidence, we note that more than 30 CPU-hours (three actual days on a 400 MHz computer) were required to find a 41-step proof of a theorem in the implicational fragment of *R-mingle* (*RM->*) and for that result, more than 35,000 formulas were considered for initiating applications of condensed detachment.

If one is not inclined to address any of the various aspects of Hilbert's twenty-fourth problem but instead is eager to attack questions requiring a *first proof*, the use of McCune's automated reasoning program OTTER will still serve one well. With recourse to its parameters, options, and the methodologies we have formulated, our colleagues have recently added substantially to the theorems of logic. For *C*5, they have found two 2-bases that are shorter in symbol count than previously known; and they have found new single axioms for that area of logic. For *C*4, they have found a single axiom, long sought by Meredith and others. Moreover, for *RM->*, our colleagues discovered a new and simpler 3-basis (of smaller size) that removes the axiomatic need for both contraction and the Parks axiom. These successes and others—each obtained with the use of OTTER—will shortly be submitted for publication.

We invite you to send us similar (or not-so-similar) open questions, or to use OTTER yourself as your personal reasoning assistant in finding simpler or new proofs.

References

[Fitelson2000] ''Axiomatic Proofs through Automated Reasoning'', B. Fitelson and L. Wos, *Bulletin of the Section of Logic*, 29, no. 3 (October 2000) 125-136

[Lemmon1957] Lemmon, E. J., Meredith, C. A., Meredith, D., Prior, A. N., and Thomas, I., ''Calculi of Pure Strict Implication'', Philosophy Department, Canterbury University College, Christchurch, New Zealand, 1957. i+22 pp. (mimeographed).

[Lukasiewicz1970] Lukasiewicz, J. *Selected Works*, edited by L. Borokowski, North Holland, Amsterdam, 1970.

[McCune1994] McCune, W., *OTTER 3.0 Reference Manual and Guide*, Tech. Report ANL-94/6, Argonne National Laboratory, Argonne, IL, 1994.

[Meredith1953] Meredith, C. A., ''Single Axioms for the Systems (C,N), (C,O), and (A,N) of the Two-Valued Propositional Calculus,'' *J. Computing Systems*, 1, no. 3 (1953), 155-164.

[Meredith1963] Meredith, C. A., and Prior, A., ''Notes on the Axiomatics of the Propositional Calculus'', *Notre Dame J. Formal Logic*, 4, no. 3 (1963) 171-187.

[Meredith1964] Meredith, C. A., and Prior, A. N., ''Investigations into Implicational S5'', *Z. Math. Logik Grundlagen Math.*, 10 (1964) 203-220.

[Thiele2001] Thiele, R., and Wos, L., ''The Strategy of Cramming,'' Preprint ANL/MCS-P898-0801, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL, 2001.

[Veroff1996] Veroff, R., ''Using Hints to Increase the Effectiveness of an Automated Reasoning Program: Case Studies'', *J. Automated Reasoning*, 16, no. 3 (1996) 223-239.

[Wos1995] Wos, L., ''The Resonance Strategy'', *Computers and Mathematics with Applications*, 29, no. 2 (February 1995) 133-178.