Hilbert's Twenty-Fourth Problem

Rudiger Thiele

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

and

*Larry Wos**

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

Abstract

For almost a century, a treasure lay hidden in a library in Germany, hidden until a remarkable discovery was made. Indeed, for most of the twentieth century, all of science thought that Hilbert had posed twenty-three problems, and no others. In the mid-1990s, however, as a result of a thorough reading of Hilbert's files, a twenty-fourth problem was found (in a notebook, in file Cod. Ms. D. Hilbert 600:3), a problem that might have a profound effect on research. This newly discovered problem focuses on the finding of simpler proofs. A proof may be simpler than previously known on one or more ways that include length, size (measured in terms of total symbol count), term structure, and the like. This article presents Hilbert's twenty-fourth problem, discusses its relation to certain studies in automated reasoning, and offers researchers with varying interests the challenge of addressing this newly discovered problem. In particular, we include open questions to be attacked, questions that (in different ways and with diverse proof refinements as the focus) may prove of substantial interest to mathematicians, to logicians, and (perhaps in a slightly different manner) to those researchers primarily concerned with automated reasoning.

1. Background and Perspective

Featured here is a remarkable discovery, one that is pertinent to mathematics, to logic, and (of course) to automated reasoning. The consequences of this discovery and the possible research it may spawn cannot be estimated at this time.

Among mathematicians of the twentieth century, Hilbert often (if not always) receives the highest ranking. The problems he offered in his Paris lecture in 1900 have occupied fine minds for decades. Marked progress in solving any of them results in substantial acclaim for the researcher.

If asked about the number of problems Hilbert posed, even a master would have answered—until very recently—twenty-three. Indeed, that is the number of problems Hilbert offered in his Paris lecture. But a twenty-fourth problem does exist, a problem that was not presented in 1900, a problem that remained hidden in his massive files in Germany—until it was discovered in the mid-1990s. With details given in Section 2, the focus of that intriguing problem is to find *simpler proofs*, sometimes measured in the number of deduced steps, sometimes measured in terms of total number of symbols, and sometimes with the focus on to some other refinement. As evidence of Hilbert's concern for finding simpler proofs, in 1917 Hilbert cited the problem in one of his lectures. Near the end of his life, he also cited the problem once again when he indexed his files.

---

*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 the German Academy Leopoldina, Halle, under Contract LPD 1996.

The most obvious class of simpler proofs concerns proof length. Indeed, ceteris paribus, the shorter the proof, the simpler the proof. Hilbert would indeed have enjoyed the presentation, accompanied by an appropriate sound argument, that the shortest proof had been found for some given theorem. For example, he would have been pleased to learn of R. Veroff's unusual use of linked inference to show, for proofs of moderate level, that no shorter proof can exist when presented with an appropriate proof [Veroff2001a].

As strong evidence of the naturalness and the significance of Hilbert's twenty-fourth problem in the context of proof length—and ironic in that this deep problem had not yet been discovered—the last half of the twentieth century witnessed fine minds embarked on the treasure hunt for shorter and yet shorter proofs. For a first excellent example, in the 1950s, the school of Lesniewski addressed deductive systems with ''shortest and simplest axioms''. A decade later, C. A. Meredith and A. Prior [Meredith1963] published ''an abridgment'' of the Lukasiewicz proof for his shortest single axiom for the implicational fragment of two-valued sentential (or propositional) calculus [Lukasiewicz1970]. More recently, I. Thomas [Thomas1975] published shorter proofs than the literature had offered for various theorems, and (still later) D. Ulrich devoted significant effort in that direction. Even here in the year 2001—although the researchers were unaware of the discovery of the Hilbert problem when they began their research—B. Fitelson and L. Wos [Fitelson2001] still intensely study the question of the existence of shorter proofs. Among their successes was the finding of a proof shorter than the cited Meredith-Prior proof, a discovery made possible through the use of a newly formulated strategy (the *cramming strategy*, introduced in [Wos2001]).

This last success nicely and sharply connects Hilbert's twenty-fourth problem to automated reasoning. Indeed, McCune's program OTTER [McCune1994] played an indispensable role in the discovery of the abridgment of the Meredith-Prior abridgment. (In Section 4 of this article, additional details regarding the use of an automated reasoning program will be given concerning proof refinement, in the context of length, complexity, and other aspects.)

The pursuit of simpler proofs is by no means confined to proof length. Quite relevant is proof complexity, the property that focuses on the longest equation or formula among the deduced steps. Another concern (suggested by Ulrich) is proof size, the total number of symbols of the proof (or one can restrict the measure to the deduced steps). The search for simple proofs may also focus on the maximum number of distinct variables that occur in any of the deduced steps. A less apparent aspect of simplicity concerns the nature of the terms that are present, for example, the presence or the absence of terms, say, of the form $n(n(t))$ for some term t, where the function n denotes negation.

Even in the context of the innocent-sounding quest for the simplest proof with respect to length, one soon encounters perhaps unforeseen obstacles. As an example of the unexpected, one might naturally conjecture that the prize would be won by merely conducting a breadth-first or level-saturation search. To see why this conjecture fails to hold, begin with the following informal definition. The level of an axiom or assumption pertinent to the theorem under study is 0, and the level of a deduced conclusion is one greater than the maximum levels of the parents of the conclusion. Then note that for many problems the size of the levels to be explored grows more or less exponentially (which can be verified through experimentation). Finally, to finish the refutation of the conjecture, note that some interesting proofs unfortunately have level 70 and greater. In other words, in the vast majority of situations that might be imagined, a breadth-first search aimed at finding with certainty a shortest proof is highly impractical, even with the aid of today's impressively powerful computers.

For a far subtler and even (to some) diabolical obstacle to the seeking of shorter proofs, we cite the aphorism ''Shorter subproofs do not necessarily a shorter total proof make''. To understand why this aphorism holds, consider the following. First assume that the object is to prove from a single formula the conjunction of A, B, and C, as is the case, for example, when showing that the Meredith single axiom for two-valued sentential (or propositional) calculus implies the three-axiom system of Lukasiewicz. Second, assume that the goal is to find a proof of the conjunction that is shorter than the proof in hand. Third, assume that within the given proof $\mathbf{P}$, the subproof $\mathbf{Q}$ of A has length j. Fourth, assume that the attempt at finding a shorter proof of the conjunction apparently leads to an advance, yielding a proof $\mathbf{R}$ of A of length strictly less than j. The replacement of $\mathbf{Q}$ by $\mathbf{R}$—rather than producing a shorter proof—may in fact produce a longer proof of the conjunction. Such can happen when many of the deduced steps of $\mathbf{R}$ serve no purpose other than to prove A, whereas the deduced steps of

Q assist in the proof of *B*. For the curious or less experienced with proof finding, we note that such occurrences are with surprising frequency.

The preceding small example shows that an obvious shortcut to finding a shortest proof will, most of the time, fail. That shortcut is, of course, simply focusing on one member of a conjunction to be proved and seeking a shortest proof of that member. However, although the ideal goal of showing without doubt that a shortest proof has been found is most elusive and apparently out of reach in most cases, the opportunity for addressing Hilbert's twenty-fourth problem in a significant way in the context of proof length still exists. In particular, substantial satisfaction can be derived from finding a proof shorter than that found by a master who, in turn, has exhibited concern for such matters. To provide a beginning for the interested researcher, we devote Section 6 of this article to challenges and open questions in the context of finding a proof shorter than known. In that section and also germane to Hilbert's twenty-fourth problem, we also offer proof-refinement challenges in the context of fewer distinct variables, in the context of complexity, and in other contexts.

The challenges we offer do not demand an interest in automated reasoning nor the use of a reasoning program. Nevertheless, judicious use of such a program (specifically, McCune's program OTTER) has yielded many, many successes of the type in which Hilbert almost certainly would have derived pleasure. With the conscious objective of stimulating research in proof refinement, we provide (in Section 4) for those new to automated reasoning as well as for those quite experienced some of the details regarding effective methodologies. For an example of the gold that can be mined, we focus (in Section 5) on the discovery of a 38-step proof of Meredith's single axiom for two-valued sentential (or propositional) calculus and note that his 41-step proof had provided what appeared to be an impenetrable barrier. In that section, we also present a newly discovered 30-step proof of an axiom dependence in infinite-valued sentential calculus, found in fact during the writing of this article. This plum concerns a deep theorem first proved by Meredith; our proof is the shortest known at this time.

With the background in hand and the perspective examined, we now turn to the historical aspects of the astounding revelation concerning Hilbert's *new* problem. An awareness of this problem may spark new avenues of thought and effort. For those who use a reasoning program, the information we present concerning the weapons that are pertinent to addressing the new twenty-fourth Hilbert problem may indeed prove intriguing.

2. The History of Hilbert's Newly Discovered Twenty-Fourth Problem

At the dawn of the last century, Hilbert—one of the most famous mathematicians of the day—offered a list of problems to the mathematical community at the International Mathematical Congress (ICM) in Paris. The spirit of a great deal of contemporary mathematics is foreshadowed in Hilbert's problems.

Having this impact in mind, one may be surprised at how quickly Hilbert created his collection. Near the end of 1899, Hilbert was invited to give a major lecture at the forthcoming ICM. Hilbert wavered about whether he should reply to a talk of Poincare given at the 1897 ICM in Zurich or should present a list of important unsolved problems. Not until the end of spring 1900 did he decide to attempt a look into the future of mathematical research, and in July he surprised his friends Hurwitz and Minkowski with the proofs of the printed version of his forthcoming Paris lecture (see the letters from Minkowski to Hilbert ref to be added]).

As a matter of fact, Hilbert composed that famous list during the summer term of 1900, in which he lectured ten hours a week (his normal task). Therefore, the preparation of his Paris talk was also interwoven with his current research.
###from my colleague: I am not sure what this means: he presented an integral and then formulated it as a problem? Why is an integral a problem? Does he mean hilbert formulated a problem based on this integral? ## Indeed, the twenty-third problem has its roots in his summer lecture course on Flaechentheorie (theory of surfaces), during which he presented his famous invariant integral for the first time and then—just a few weeks later—presented this discovery as a formal problem.

To be more precise, Hilbert did not actually lecture on that problem.
###from my colleague: Perhaps you could add a phrase or two in the next sentence, to provide a hint about how Hilbert did present the twelve. ## Rather, because of the lack of time, he presented it as he

did twelve other problems of the collection in the cited printed version in the Goettingen Nachrichten. Moreover, he canceled a twenty-fourth problem in the lecture, as well as in all the later published versions of the lecture that appeared revised and translated (bibliography, see also Grattan-Guinness, Notices of the AMS [ref]). The omitted problem is recorded, however, in his ''Mathematisches Notizbuch'', preserved in the Niedersaechsische Staats- und Universitaets-bibliothek Goettingen, Handschriftenabteilung (Cod. Ms. D. Hilbert 600).

Like many other mathematicians, Hilbert had a ''scientific diary'', a most remarkable three-volume document that, unfortunately, has not yet been published or evaluated. In it are enough unpublished ideas to have made dozens of reputations! Hilbert apparently had too many ideas to work out himself (like his predecessor Carl Friedrich Gauss in Goettingen) or to give to his collaborators, and that is why he recorded some of the ideas from 1895 on.

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 he expressed in the following.

> 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. Quite generally, if there are two proofs for a theorem, you must keep going until you have derived each from the other, or until it becomes quite evident what variant conditions (and aids) have been used in the two proofs. Given two routes, it is not right to take either of these two or look for a third; it is necessary to investigate the area lying between the two routes. Attempts at judging the simplicity of a proof are in my examination of syzygies and syzygies [Hilbert made a slip in writing] between syzygies. The use or the knowledge of a syzygy simplifies a proof essentially that a certain identity is true. Because any process of addition [is] an application of the commutative law of addition etc. [and because] this always corresponds to geometric theorems or logical conclusions, one can count these [processes] and, for instance, in proving certain theorems of elementary geometry (the Pythagorean theorem on remarkable points of triangles) one can very well decide which of the proofs is the most simple. [Part of the last sentence is not only barely legible but also grammatically incorrect. Some corrections and insertions that Hilbert made in this entry frequently show that he wrote down the problem in haste.] (Notebook Cod. Ms. Hilbert 600:3, pp. 25-26; trans. R. Thiele, 2000)

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. The (almost) twenty-fourth problem belongs to foundations in general. For Hilbert, generally speaking, the subject of foundations falls into two main branches: proof theory and metamathematics. We will see in more detail how he dealt with the problem later.

Up to 1904, the year of the Heidelberg ICM, besides the calculus of variations and integral equations, Hilbert also conducted research in logic and in the foundations of arithmetic. Then, he made a surprising turn, focusing for a long period on mathematical physics rather than logic. Of course, from 1914 to 1918, World War I hampered scientific life in general.
###from my colleague: Would an example or two of the next sentence be of interest?### But in 1917 Hilbert gave a lecture in Zurich, Switzerland, in which he mentioned just those examples he had added to the idea of simplest proof in the Notebook.

Hilbert continued his research in the foundations of arithmetic; but it is one of life's little ironies that when he delivered his last great speech on the foundations of mathematics in Koenigsberg in 1930, a young mathematician named Kurt Goedel came into the story and proved results Hilbert had not expected (incompleteness theorems). Although old and sick, Hilbert tried to react and to save his program; however, his creative time was gone. Nevertheless, at the end of his life, looking back and revisiting his research, Hilbert made and inserted a short ''index'' into his Notebooks; and, among the few key words on this one-page index, we find the twenty-fourth problem again.

All in all, we see that Hilbert regarded the problem of simplest proof as an important one throughout his life. Moreover, 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.

The most interesting question is: Why did Hilbert not announce this problem in Paris or later elsewhere? We think there are 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 of simplest proofs in precise mathematical concepts is not at all easy. (Especially Section 4 of this article will discuss the aspects and scope of this important concept.) 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 special fields.

Somewhere A. Weil said, ‘‘Great problems furnish the daily bread on which mathematics thrives’’. As Hilbert's pupil and biographer Blumenthal put it, Hilbert was a man of problems. Specifically, Hilbert started with clear and plausible examples and then extended the mathematical state to a theory. That is why he was (and we, too, can be) confident that there exists a reliable ground for the truth of his conjectures and ideas. In any restricted or more precisely given form, there should be an answer. We hope that this paper will encourage readers to devote research to finding the answer, to studying the question of simpler proofs.

3. On Hilbert Himself

For a better understanding of Hilbert's attitude, we add some remarks on his background. Around the turn of the last century, there were diverse opinions about how to found mathematics. Three schools dominated the discussions: the formalists, the logicists, and the intuitionists. The philosophy of the formalists, led by Hilbert, is usually explained in the following way. The different branches of mathematics are described by deductive systems each of which has its specific axioms and definitions. But the objects of such systems are only symbolic elements that no longer reflect any intuition.

Characterizing formalism in this way, one may overlook the fact that Hilbert knew that such a formalistic reasoning cannot prove the consistency of a deductive system. For Hilbert, nature and thinking were finite, and the infinite was only an idealization of our mind in an extended and nonformalistic sense. Hence, he attempted to create his metamathematics: to found mathematics on the intuitive ground of finiteness. Nevertheless, we must start with the formalization of a branch in order to keep ideas and reasoning strong.

We think that Hilbert believed that ‘‘the proof of proof: that it must always be possible to arrive at a proof’’ (Notebook 600:3, p. 95) rests on this philosophy of finiteness. To support our belief, we again quote Hilbert himself: ‘‘All our effort, investigation, and thinking bases on the belief that there can be but one valid view (maximum)’’ (*Ibid.*).

Of course, Hilbert was acquainted with the complexity of finite problems. (A modern example would be the four-color problem.) In the Notebook he asked, for example, ‘‘What is the 10 to the 10 to the 10th decimal of pi?’’ With respect to the complexity, he regarded the mathematician's function to be to simplify the intricate (*not* to complicate what is simple and then call it ‘‘generalizing’’) (*Ibid.*, p. 45). He advised, ‘‘Always endeavor to make a proof with the least elementary means, for that way mastery of the subject comes best to the fore’’ (*Ibid.*, p. 105). He optimistically concluded, ‘‘That there is no Ignorabimus in mathematics can probably [be] proved by my theory of logical arithmetic’’ (*Ibid.*). In contrast to the ‘‘Ignorabimus’’ (‘‘we will not know’’) in mathematics, he demanded ‘‘Noscemus’’ (‘‘we will recognize/will know’’). Moreover, for Hilbert the question of simplicity is—without doubt—linked with the question of the unity of mathematics. And it was just the unity of all branches of mathematics he obstinately insisted on in all his mathematical activities. There is but one science of mathematics. If we cannot recognize the association of mathematical branches, theories, and so on, we have not simplified enough the nature of the ideas under inspection. Simplicity not only is a practical demand; it also glues mathematical thinking together. From this viewpoint, simplicity clearly is a fundamental concept in Hilbert's work.

John Barrow, professor of astronomy at the University of Sussex, has expressed astonishment about simplicity: ‘‘It is enigma enough that the world is described by mathematics but by simple mathematics, . . . this is a mystery within an enigma'' (*The World within the World,* p. 349). For Hilbert, however, every object of thinking was also an object of mathematics: no essential difference exists between thinking and being (pre-established harmony) (*Ibid.,* p. 95). That is why we find simple theories not only in mathematics but also in natural sciences. Similarly, Albert Einstein turned to simplicity in constructing mathematical theories, stating, ''Our experience hitherto justifies us in believing that nature is the realization of the simplest conceivable mathematical ideas'' (see [Norton2000]).

4. Methodologies for Addressing the New Hilbert Problem

If one would enjoy attacking Hilbert's new problem, an automated reasoning program might provide substantial assistance, depending of course on the arsenal of weapons the program offers. Especially regarding proof simplification in the context of length, an array of weapons is needed because of the difficulty and subtlety of seeking shorter proofs (as discussed in Section 1). From such an array, one can formulate various methodologies, of which the following are a sample. (For a far fuller treatment of methodologies, see the forthcoming book entitled *Automated Reasoning and the Discovery of Missing and Elegant Proofs* [Wos2002].) The various methodologies for seeking shorter proofs are driven each by focusing on one or more proofs in hand.

4.1. Proof-Step-Blocking Methodology

In its simplest incarnation, the proof-step-blocking methodology attempts to find a shorter proof than the best in hand by blocking the use of its deduced steps one at a time. At one end of the spectrum is the case in which a shorter proof exists such that all of its steps form a proper subset of the deduced steps of the longer proof driving the search. The definition of proof requires that no so-called extra steps be present. Therefore, if a proof of the type just cited exists, at least one of its steps must have a different parentage from that in the longer proof.

A fine example of this phenomenon is provided by two proofs taken from two-valued sentential (or propositional) calculus, both of which use as hypothesis the Lukasiewicz system, and both of which derive the Church axiom system. For the curious, the longer proof has length 22, and the shorter proof has length 21. The shorter proof was discovered by relying on the hot list strategy, on resonators corresponding to the deduced steps of the 22-step proof, and on an assignment to max_weight that discourages the program from retaining new conclusions unless they match (are similar in shape to) one of the twenty-two resonators. (For a full treatment of the just-cited items, see two books, *A Fascinating Country in the World of Computing* [Wos1999] and the *Collected Works of Larry Wos* [Wos2000].)

In lieu of formal definitions and fine detail, the basic idea in the case under discussion is to attempt to force the program to rely solely on various steps of the 22-step proof and instruct the program to visit and revisit continually the axioms used to drive the attack. The deduced steps of the shorter 21-step proof are in fact contained among those of the longer 22-step proof; the omitted step is the second deduced step of the longer proof. That step is used to deduce the sixth step of the 22-step proof and no other. That sixth step also appears as the sixth step of the shorter 21-step proof, but its parents are different from those in the longer proof. Indeed, in the longer proof, its sixth step relies on two deduced steps as parents, whereas the sixth step in the shorter proof relies on a deduced step and on one of the axioms (which is promoted by the use of the hot list strategy). (The proofs exhibit other differences as well.)

A less used but sometimes effective incarnation of the methodology under discussion concerns blocking two or more steps at a time. A third incarnation relaxes the max_weight (by assigning a somewhat generous value) to permit the use of deduced conclusions not formerly allowed to participate. The proofs found with this incarnation can be at the other end of the spectrum, For example, when the max_weight was assigned the value 28 (in the context of the 22-step proof), OTTER found various proofs, including one of length 54. That proof avoided the use of six steps of the 22-step proof. More dramatic is the case concerning two 22-step proof that differ by twelve steps, both deducing the Church axiom system, and both using the Lukasiewicz axiom system as hypotheses.

The essence of blocking proof steps is to require the program to seek a proof different from that in hand. Various additional mechanisms are used to cause the program to explore previously unexplored regions of the space of deduced conclusions. The actual blocking is typically by means of a nonstandard use of demodulation or a nonstandard use of weighting.

4.2. The Methodology of Related Proofs

A reasonable conjecture asserts that some proofs found by an unaided researcher are obtained by conscious imitation of an earlier success. Just how such imitation is effected remains essentially a mystery. Although the automation of imitation is clearly not the goal of automated reasoning as practiced in the Argonne paradigm, nevertheless a vaguely related approach can be put to good use in some extraordinary cases. (R. Overbeek suggests that a good research project focuses on a modification to a reasoning program that does admit a form of imitation, where one collects all proof steps of some author and has the program focus almost exclusively on such formulas or equations. Access to such a mechanism might indeed enable a researcher to imitate a master, for example, Hilbert himself.)

Specifically, when the goal is that of finding a shorter proof (especially in the context of a deep theorem), the methodology of related proofs can win the game occasionally. To apply this methodology to a theorem **T** of the form P implies Q—although not required—a proof of **T** serves nicely, at least as a wellspring. If such a proof is in hand, and even better if that proof was offered by a master who has evidenced interest in proof refinement with respect to length, one begins by choosing some of its properties on which to focus with the intent of finding other proofs *not* sharing those properties. For example, the variable richness of the proof in focus might be k, where k is the maximum number of distinct variables present in one of its deduced steps. Also, the proof might rely on occurrences of some type of term such as that of double negation, $n(n(t))$ for some term t with the function n denoting negation. Required for the methodology is one or more proofs that violate one or more chosen properties of the proof in hand.

For example, consider Meredith's (in effect) 41-step proof showing that his single axiom as hypothesis suffices to derive the Lukasiewicz axiom system for two-valued logic, where the sole inference rule in use is condensed detachment. Two steps of that proof rely on seven distinct variables (its variable richness is 7), and seventeen of the 41 deduced steps rely on double negation. Meredith's proof played a key role in finding two new rather different and indeed interesting proofs, one of which has variable richness 6, and one of which is totally free of double-negation terms.

The methodology featured in this section has the program key on such new proofs with the objective of discovering a proof shorter than the so-called old proof that prompted the study. Although not necessary, one usually seeks a proof with the same properties (such as variable richness and term structure) because the likelihood of success is increased. In the context of the just-given example, one would attempt to use the two new proofs, differing, respectively, in variable richness and in term structure, from the proof to be refined, in a manner to direct a reasoning program's attack. The goal for us was to find a proof for Meredith's single axiom of length strictly less than 41, allowing it to have variable richness 7 and to rely on deduced steps containing double-negation terms. As noted in Section 1, we in fact succeeded in mining the sought-after gold, discovering a 38-step proof.

The mechanism for making such an attempt is a layered resonance approach, which consists (in this case) of including two sets of resonators. The first set corresponds to the deduced steps of the first new proof, each assigned a small value, say, i; the second set corresponds to the deduced steps of the second new proof, each assigned a small value, say, i+1. The assigned value to max_weight must be greater than i+1. (Resonators are formulas or equations whose variables are treated as indistinguishable.) When the program deduces a new conclusion, the conclusion is assigned a priority equal to the assigned value of the first resonator it matches (if such exists), where all variables are treated as indistinguishable. The lower the assigned value, the higher the priority given to the formula or equation for being chosen to direct the program's attack. If one prefers to focus on items that subsume or are subsumed by the steps of new proofs, rather than on items that are similar in shape to proof steps, R. Veroff's hints strategy serves one well [Veroff1996].

4.3. The Target-Replacement Methodology

Occasionally one can (as we in fact did) find a shorter proof by replacing the so-called target of the best proof in hand by a different target. For example, consider Meredith's proof, which completes with the deduction of three formulas, the Lukasiewicz axiom system for two-valued sentential calculus.

With the methodology taking center stage here, one might be interested in finding *any* shorter proof showing that Meredith's formula suffices, perhaps a proof having the Church system as target rather than the Lukasiewicz system. The goal might be to refine the Meredith 41-step proof or, instead, to refine a 38-step proof discovered by OTTER.

The target-replacement strategy adds to the interest in any new axiom system that is found. To prove that some set of formulas or equations is, in fact, a new axiom system, one typically derives from it some known axiom system (sometimes called a basis). When Z. Ernst (in early 2001) found seven new (proposed) axiom systems for *C*5 (the implicational fragment of *S*5, which is itself a modal logic), he set in motion a series of studies whose objective was to deduce from each of the seven new systems one or more known bases for *C*5. Most likely unexpected, he also provided researchers with the opportunity of finding shorter proofs establishing that Meredith's single axiom for *C*5 is sufficient—new targets.

Indeed, prior to the use of the target-replacement methodology, members of the Argonne research team had found a 28-step proof showing that the Meredith single axiom for *C*5 implies a known 3-basis, a proof shorter than offered by the literature. One can abridge one's own work; indeed, quite often a person finds a so-called first proof and, before submitting it for publication, simplifies that proof in the context of length and other aspects (as Hilbert would approve of). (In fact, many months after we found the cited 28-step proof, possibly because of advances in our methodologies, we revisited the problem. That revisiting eventually produced a 23-step proof, found shortly before Ernst added his new axioms to the pool and before the target-replacement strategy was formulated.)

That addition enabled us to apply the target-replacement methodology. By using one of the Ernst new single axioms for *C*5, we were able to discover with OTTER an 18-step proof establishing the Meredith formula to be a full axiomatization for *C*5, and we had a success with the new methodology.

4.4. Other Methodologies, Other Refinements

Hilbert's twenty-fourth problem offers challenges in other areas than proof length. As mentioned in the preceding section, *variable richness* is one such aspect. Just as an axiom system is more appealing because of exhibiting smaller variable richness than other axiom systems—none of its members depend on more than *j* distinct variables, whereas earlier systems require more richness—so it also is the case for proof. A reduction in variable richness corresponds to a type of simplification.

For example, Meredith's 41-step proof for his single axiom for two-valued sentential calculus exhibits a variable richness of seven in two steps, and a small improvement is found in each of our three 38-step proofs for his axiom in that each contains but one formula relying on seven distinct variables (in part discussed in the preceding sections). The 38-step proofs were discovered with OTTER applying various methodologies.

We note that the first step of any proof focusing exclusively on the Meredith axiom, regardless of the target, relies on five distinct variables and is common to all such proofs because it arises from applying condensed detachment to the axiom with itself. Therefore, if a simplification of Meredith's 41-step proof with respect to variable richness is possible, the best proof in that regard would exhibit a richness of five.

Hilbert would (almost certainly) consider the seeking of such a proof was merited because one would have attained the limiting case. OTTER offers just what is needed to address this aspect of simplification, even (much of the time) in the context of the limiting case. The relevant parameter is max_distinct_vars. The value assigned to this parameter instructs the program to discard any newly deduced conclusion if its richness exceeds that value.

With various methodologies, OTTER did in fact find an encouraging 49-step proof (of Meredith's single axiom) with variable richness six, whose forty-second step is the only step relying on six distinct variables. Typically, however, a refinement in one aspect or parameter results in a cost or loss in some other parameter. Indeed, that 49-step proof of variable richness six (which is the shortest of its type known to us) contains nineteen steps relying on double-negation terms. When a proof with richness of five was sought, success eventually occurred, discovering a 69-step proof, eight of whose steps rely on five distinct variables, and eighteen of whose steps rely on double negation.

If, instead of variable richness, the aspect of Hilbert's twenty-fourth problem to be addressed concerns *formula complexity* (the number of symbols in the longest formula or equation among the deduced steps), OTTER offers the max_weight parameter. The program automatically discards any conclusion whose weight strictly exceeds the value assigned to max_weight, where the weight of a conclusion is measured in terms of symbol count unless otherwise determined. One must of course be aware that included resonators or other weight templates can cause a deduced conclusion to have an assigned weight other than its symbol count.

Among other aspects of simplification pertinent to the new Hilbert problem is that of proof size (as suggested by D. Ulrich). The size of a proof (restricted to its deduced steps) is the total number of symbols present. A reduction in proof length obviously does not guarantee a reduction in proof size. Nevertheless, if size is the aspect of simplification of interest, perhaps the most effective attack rests with focusing on proof length. Although not as subtle and deep as simplification regarding length of proof, we find size to present some vaguely similar obstacles. Indeed, we know of no direct approach to addressing the problem, and OTTER offers nothing clearly germane.

5. Successes

A small sampling of successes might encourage those who find Hilbert's twenty-fourth problem intriguing and who, at the same time, suspect it to be too formidable to attack. These successes also in part pave the way for offering challenges and open questions in Section 6. For the researcher who wishes more than a taste of what can be done (which is the focus in this section), we reserve the needed formulas and equations for that section.

When, as is the case with Meredith, a single axiom is the shortest known for some area of logic, the axiom merits intense consideration. With crucial assistance from OTTER, we have found satisfying simplifications of Meredith's 41-step proof for his single axiom for two-valued sentential calculus.

- Length. In contrast to the Meredith proof, we have found three proofs of length 38.

- Size. With the predicate symbol as part of the count, the size of the 41-step proof is 696, and the respective sizes of the three shorter proofs are 624, 633, and 644 (coincidentally, in the order of their discovery). We know of no proof with size less than 624.

- Variable richness. Whereas the variable richness of Meredith's proof is seven, OTTER found a 69-step proof with richness equal to five and present in but eight steps relying on that many distinct variables.

- Term structure. The Meredith proof contains seventeen steps relying on double-negation terms; we have in hand a 51-step proof free of double negation.

- Formula complexity. Meredith's proof has formula complexity 34 (not counting the predicate symbol) in but one of its steps. Each of the three 38-step proofs also has complexity 34, but each has two such steps among the deduced steps.

A brief glance at the structure of Meredith's 41-step proof compared with that of any of the 38-step proofs might provide some insight into how to attack an implied open question (given in the next section). Meredith's proof proves the Lukasiewicz axioms in the order 3, 2, 1, and uses 2 to prove 1 (which intuitively is the most challenging). In sharp contrast, the 38-step proof (of level 24) proves the axioms in the order 3, 1, 2, and the proof uses 1 to prove 2 and uses 3 later in the proof as well. Such counterintuitive proofs present little or no problem for OTTER—the program lacks intuition, except for that which the researcher (in effect) gives it through the use of resonators and the like. In other words, the lack of intuition for program or researcher can, and sometimes does, lead to a proof simplification that Hilbert might have thoroughly enjoyed.

The next success focuses on axiom dependence, a topic that also is of substantial interest to mathematicians and logicians, and one that presents somewhat different obstacles than does the study of a proposed single axiom. The area of concern is infinite-valued sentential calculus, weaker than is two-valued, and originally axiomatized by Lukasiewicz with the following five formulas expressed as clauses.

P(i(x,i(y,x))).

P(i(i(x,y),i(i(y,z),i(x,z)))).
P(i(i(i(x,y),y),i(i(y,x),x))).
P(i(i(n(x),n(y)),i(y,x))).
P(i(i(i(x,y),i(y,x)),i(y,x))).

Meredith eventually proved that the fifth of the five is in fact dependent upon the other four axioms. The theorem is difficult to prove; indeed, a search of the literature strongly suggests that until the early 1990s no purely Hilbert-style axiomatic proof was offered. Instead, the various approaches were based in part on the use of equality, not strictly within the logic. As far as can be determined, the first axiomatic proofs were each obtained with OTTER, two proofs each of length 63 (applications of condensed detachment). The relevance to Hilbert's twenty-fourth problem rests with the following.

In particular, in the spirit of the Hilbert problem featured in this article—even though its existence was unknown at the time—we began a study in the early 1990s to find a proof shorter than length 63. The next few years witnessed the eventual discovery of a 39-step proof, closely followed by a turn of interest. Specifically, also in the spirit of Hilbert's new problem focusing on simplification, but concerning term structure rather than length, we next began a study with the goal of finding a proof of the axiom dependence such that double-negation terms were avoided. The result was the discovery of a 32-step proof with the sought-after term-avoidance property. An examination of the literature shows that such an avoidance might have been thought unlikely or perhaps even impossible.

Still later, a third aspect of simplification entered the picture, that of blocking the use of lemmas that had played a key role for the masters. Our effort yielded a 34-step proof, one free of double negation and not dependent on three key lemmas. We experienced little disappointment with the slight (two-step) increase in proof length because simultaneous refinements typically require a relaxation of some type.

During the writing of this section, and most likely in part because of the writing, we revisited this deep theorem of infinite-valued sentential calculus. With not much optimism, the intent was to seek an even more elegant proof. And, because of the effectiveness of the various methodologies and the power of OTTER's arsenal of weapons, the following most unexpected 30-step proof was discovered. (For OTTER, the clause notation relies upon ‘‘-’’ to denote logical **not** and on ‘‘ | ’’ to denote logical **or**.)

A Thirty-Step Proof of the Dependence within the Lukasiewicz Five-Axiom System

```
----- Otter 3.1-b0, May 2000 -----
The process was started by wos on myrtis.mcs.anl.gov,
Mon May 14 21:08:24 2001
The command was "otter". The process ID is 22654.
----> UNIT CONFLICT at   0.72 sec ----> 1627 [binary,1626.1,9.1] $ANS(MV_5).

Length of proof is 30. Level of proof is 21.

---------------- PROOF ----------------

1 [] -P(i(x,y)) | -P(x) | P(y).
2 [] P(i(x,i(y,x))).
3 [] P(i(i(x,y),i(i(y,z),i(x,z)))).
4 [] P(i(i(i(x,y),y),i(i(y,x),x))).
5 [] P(i(i(n(x),n(y)),i(y,x))).
9 [] -P(i(i(i(a,b),i(b,a)),i(b,a))) | $ANS(MV_5).
26 [hyper,1,3,3] P(i(i(i(i(x,y),i(z,y)),u),i(i(z,x),u))).
28 [hyper,1,3,2] P(i(i(i(x,y),z),i(y,z))).
31 [hyper,1,3,5] P(i(i(i(x,y),z),i(i(n(y),n(x)),z))).
33 [hyper,1,26,26] P(i(i(x,i(y,z)),i(i(u,y),i(x,i(u,z))))).
42 [hyper,1,28,4] P(i(x,i(i(x,y),y))).
47 [hyper,1,3,31] P(i(i(i(i(n(x),n(y)),z),u),i(i(i(y,x),z),u))).
49 [hyper,1,31,28] P(i(i(n(x),n(i(y,z))),i(z,x))).
```

63 [hyper,1,33,42] P(i(i(x,i(y,z)),i(y,i(x,z)))).
85 [hyper,1,3,63] P(i(i(i(x,i(y,z)),u),i(i(y,i(x,z)),u))).
95 [hyper,1,85,33] P(i(i(x,i(y,z)),i(i(u,x),i(y,i(u,z))))).
124 [hyper,1,95,4] P(i(i(x,i(i(y,z),z)),i(i(z,y),i(x,y)))).
125 [hyper,1,95,3] P(i(i(x,i(y,z)),i(i(z,u),i(x,i(y,u))))).
162 [hyper,1,125,124] P(i(i(i(x,y),z),i(i(x,i(i(y,u),u)),i(i(u,y),z)))).
212 [hyper,1,162,5] P(i(i(n(x),i(i(n(y),z),z)),i(i(z,n(y)),i(y,x)))).
767 [hyper,1,47,28] P(i(i(i(x,y),z),i(n(x),z))).
791 [hyper,1,3,767] P(i(i(i(n(x),y),z),i(i(i(x,u),y),z))).
861 [hyper,1,791,212] P(i(i(i(x,y),i(i(n(z),u),u)),i(i(u,n(z)),i(z,x)))).
943 [hyper,1,861,5] P(i(i(x,n(y)),i(y,n(x)))).
970 [hyper,1,95,943] P(i(i(x,i(y,n(z))),i(z,i(x,n(y))))).
988 [hyper,1,791,970] P(i(i(i(x,y),i(z,n(u))),i(u,i(n(x),n(z))))).
1048 [hyper,1,26,988] P(i(i(x,y),i(z,i(n(y),n(x))))).
1112 [hyper,1,124,1048] P(i(i(i(n(x),n(y)),z),i(i(y,x),z))).
1113 [hyper,1,95,1048] P(i(i(x,i(y,z)),i(u,i(x,i(n(z),n(y)))))).
1261 [hyper,1,1113,42] P(i(x,i(y,i(n(z),n(i(y,z)))))).
1311 [hyper,1,1261,1261] P(i(x,i(n(y),n(i(x,y))))).
1345 [hyper,1,95,1311] P(i(i(x,y),i(n(z),i(x,n(i(y,z)))))).
1531 [hyper,1,1345,49] P(i(n(x),i(i(n(y),n(i(z,u))),n(i(i(u,y),x))))).
1553 [hyper,1,212,1531] P(i(i(n(i(i(x,y),x)),n(y)),i(y,x))).
1575 [hyper,1,1112,1553] P(i(i(x,i(i(y,x),y)),i(x,y))).
1626 [hyper,1,85,1575] P(i(i(i(x,y),i(y,x)),i(y,x))).

This proof is the shortest so far found for the axiom dependence in infinite-valued logic, free of double negation, and independent of three lemmas that might have been considered indispensable. The size of this 30-step proof (excluding predicate symbol occurrences) is 430, whereas one of the 32-step proofs (that found first in the years of study) has size 438, which suggests that some formula complexity has been introduced in that the length has been reduced by two. Indeed, such is the case: The 30-step proof contains two formulas of complexity 19 (excluding the predicate symbol), whereas the 32-step proof contains (with most complexity) two 17-symbol formulas. Nevertheless, nature is still unusually generous: The 30-step proof simultaneously exhibits the finest properties in many respects of simplicity—all relevant to Hilbert's twenty-fourth problem.

Among other successes, Boolean algebra, lattice theory, and group theory have yielded results that address the Hilbert twenty-fourth problem [Wos1998]. In that equality-oriented reasoning replaced condensed detachment, but many of the same methodologies were relied upon, one sees that this intriguing problem can profitably be attacked with a reasoning program.

6. Challenges and Open Questions

Before turning to some intriguing open questions (most of which are directly pertinent to the Hilbert problem featured in this article), we offer challenges regarding automated assistance. We obviously delight in attacking one open question after another—wondering how any activity can compare with such an endeavor—and strongly embrace the view that many, many discoveries would elude mathematics and logic were it not for the role played by a powerful reasoning program.

As noted earlier, an automated reasoning program that offers a variety of weapons has proved to be invaluable in addressing various aspects of Hilbert's twenty-fourth problem. However, also as noted, little is offered (at least by OTTER) in directly addressing proof size. Therefore, a challenging problem asks for some methodology (in the spirit of automated reasoning) that assists in directly seeking proofs of smaller size than the best in hand.

The second challenging problem concerns the choice of resonators. The choice of which formulas or equations to include, used to direct a program toward fertile ground, is currently left entirely to the researcher. Although no substitute exists for the expertise of a fine researcher, (as McCune observes) one might be able to write a program that would take instructions about sets and sequences of resonators for automatically running a series of experiments. We have in mind a program in the spirit of, say, OTTER's super-loop, which enables the user to automatically run a series of experiments in which chosen parameters vary over assigned values. With such a new program, one could far more easily compare the work of various masters by focusing on sets of resonators taken from published proofs. Instead, or in addition, one could more quickly run a series of experiments that differed mainly in the advice given in the form of sets of sets of resonators chosen from some pool of sets extracted from earlier successes. As Overbeek suggested, when a new area was under investigation, one could begin by proving simple theorems and gradually, from their proofs, construct a library of sets of resonators. This library, even for a distantly related area, might prove most useful for attacking open questions, including those we now offer.

The first question asks whether there exists a proof of 37 or fewer steps (of condensed detachment) that establishes the following Meredith formula (expressed in clause notation) to be sufficient for two-valued sentential (or propositional) calculus.

% Meredith's single axiom
P(i(i(i(i(i(x,y),i(n(z),n(u))),z),v),i(i(v,x),i(u,x)))).

As noted earlier, we have discovered three 38-step proofs showing that the given formula provides a complete axiomatization for two-valued (or classical) logic. In each case, the target was the following three-axiom system of Lukasiewicz.

% Lukasiewicz 1, 2, and 3
P(i(i(x,y),i(i(y,z),i(x,z)))).
P(i(i(n(x),x),x)).
P(i(x,i(n(x),y))).

To answer the question in the affirmative, one is not required to deduce the Lukasiewicz given system; for other possible targets, see the forthcoming book *Automated Reasoning and the Discovery of Missing and Elegant Proofs* [Wos2002].

Although not directly in the spirit of addressing Hilbert's twenty-fourth problem, the following merits serious study. specifically, the question of whether the Meredith 21-letter single axiom is indeed the shortest such for classical logic remains open. Especially for the intuitive researcher who occasionally identifies a subtle and hard-to-detect symbol pattern, we note that (decades earlier) Lukasiewicz supplied the following 23-letter single axiom (also expressed as a clause).

P(i(i(i(x,y),i(i(i(n(z),n(u)),v),z)),i(w,i(i(z,x),i(u,x))))).

The shortest proof we have found has length 56 (applications of condensed detachment); its formula complexity is 30, not counting the predicate symbol; and its size is 818, not counting predicate symbol occurrences. Implied are corresponding challenges regarding proof simplification.

Returning to a question clearly in the context of proof simplification but focusing on a logic weaker than classical, we offer the challenge of finding (if possible) a proof of 29 or fewer applications of condensed detachment that shows the fifth of the Lukasiewicz axioms (given in Section 5) for infinite-valued sentential calculus to be dependent on the first four. No restriction is placed on the type of term present, the use of particular lemmas, or the variable richness. For example, a shorter proof may exist in which double-negation terms are present, or some lemma from earlier literature is derived as a subproof, or one or more deduced steps rely on six or more distinct variables.

Again, one has the opportunity of answering a most difficult open question and, if successful, making a significant contribution to logic. Specifically, no single axiom currently is known for infinite-valued sentential calculus.

For the next question, we revisit *C*5, the implicational fragment of the modal logic *S*5. That logic was formulated in part to capture the usually accepted notion of implication. As noted, the study of *C*5 provides a fine example of abridging one's own work and also a fine example of discovering a shorter proof by replacing the usual target. An open question to consider asks for a proof strictly shorter than length 18 showing that the Meredith single axiom for *C*5 (the following) is sufficient, regardless of the target.

P(i(i(i(i(x,x),y),z),i(u,v)),i(i(v,y),i(w,i(u,y))))).

For an open question that addresses the Hilbert problem in focus in the context of equality, we turn to Boolean algebra. The question concerns the possible existence of a proof whose variable richness does not exceed four. The following remarkable equation (due to McCune) in terms of **or** and **not** can be proved to be a single axiom by deducing the Robbins 3-basis for Boolean algebra. (Union or logical or is denoted by ‘‘ + ’’, complement is denoted by ‘‘ ~ ’’, and not equal by ‘‘ != ’’.)

```
~(~(~(x + y) + z) + ~(x + ~(~z + ~(z + u)))) = z.
```

For completeness, we give the following negation of the Robbins basis.

```
B+A != A+B | (A+B)+C != A+ (B+C) | ~(~(A+B)+ ~(~A+B)) != B | $ANS(Robbins_basis).
```

With OTTER, we have found a 57-step proof whose variable richness is five. Since we know of no shorter proof, an additional challenge is offered in the context of refinement with respect to proof length. The question does not demand that the Robbins basis be used as the target to complete a proof of sufficiency.

A distantly related and most difficult open question asks about the existence of a single axiom (in terms of **or** and **not**) whose length is 21 or less. Because of combinatoric considerations, this question is even harder than its correspondent in which **or** and **not** are replaced by the Sheffer stroke [McCune2001]. That question was answered by finding four single axioms, each of length fifteen, a splendid success whose wellspring was Veroff’s impressive research focusing on finding simpler bases [Veroff2001b]. Regarding the Sheffer stroke, we offer yet another open question concerning proof refinement in the context of length. For the following equation in the Sheffer stroke, does there exist a proof of length strictly less than sixty-two showing it to be a single axiom for Boolean algebra? As implied, we have in hand (because of OTTER) a 62-step proof relying solely on the inference rule paramodulation (with no use of demodulation) that deduces a Sheffer 3-basis. The following give, respectively, one of the 15-symbol single axioms and the negation of the Sheffer basis.

```
f(f(x,f(f(y,x),x)),f(y,f(z,x))) = y.
f(f(a,a),f(a,a)) != a | f(a,f(b,f(b,b))) != f(a,a) | f(f(f(b,b),a),f(f(c,c),a)) != f(f(a,f(b,c)),f(a,f(b,c))).
```

To attack the question, one may find it profitable to use a target other than the given Sheffer basis.

for the final question, we turn to quasilattices, axiomatized in the following manner (suitable for consideration with OTTER) by the first nine equations. The tenth equation is a self-dual form of modularity, and the conclusion to be proved is a standard form of modularity. The problem is referred to as QLT-5 in [McCune1996]. (Join is denoted by ‘‘ v ’’ and meet by ‘‘ ^ ’’.)

```
x = x.
x ^ x = x.
x ^ y = y ^ x.
(x ^ y) ^ z = x ^ (y ^ z).
x v x = x.
x v y = y v x.
(x v y) v z = x v (y v z).
(x ^ (y v z)) v (x ^ y) = x ^ (y v z).
(x v (y ^ z)) ^ (x v y) = x v (y ^ z).
(x ^ y) v (z ^ (x v y)) = (x v y) ^ (z v (x ^ y)).
```

The challenge is to find a proof of length strictly less than 35 (applications of paramodulation, without intermediate demodulation) of the following, given in its negated form.

```
(A ^ B) v (A ^ C) != A ^ (B v (A ^ C)).
```

7. The Hilbert New Problem and Automated Reasoning: A Symbiosis

The careful examination in the late 1990s of Hilbert’s writings did indeed yield a treasure, his twenty-fourth problem. The focus of that problem is proof simplification. Hilbert clearly recognized the difficulty of defining simplicity, which probably accounted for his not offering the problem in his Paris lecture. He was, as noted earlier, continually interested in this problem, and his views are in part captured with the following observation: ‘‘Always endeavor to make a proof with the least elementary means, for that way mastery of the subject comes best to the fore.’’ With respect to the complexity, Hilbert regarded the mathematician's function to be to simplify the intricate (*not* to complicate what is simple and then call it ''generalizing'') (*Ibid.,* p. 45).

Among Hilbert's observations, the following sheds even more light on his twenty-fourth problem. ''To formulate the intuitively obvious question of simplest proofs in precise mathematical concepts is not at all easy.'' Despite the imprecise nature of proof simplification, ordinarily shorter proofs are simpler than longer. Also, among other properties, a proof with fewer symbols in total (smaller size) is simpler than one with a greater number. The cases for refinement with respect to variable richness, formula complexity, and the like are somewhat subtler but still hold.

Depending on one's particular criterion for an aspect of proof—for example, the avoidance of thought-to-be-indispensable lemmas—the possibilities for addressing this (in effect) new Hilbert problem are endless. The appeal of this problem will have a slightly different shading for the unaided mathematician, compared with the unaided logician, compared with the researcher relying on the assistance of an automated reasoning program.

Perhaps the last of these three groups is where the appeal will be the strongest because of the ability of a reasoning program to explore huge spaces of conclusions and do so, when instructed to, in a most unintuitive manner. The avoidance of the use of double negation is but one example of an unintuitive attack.

The nature and variety of weaponry and the diversity of methodology offered by a program such as OTTER invite vigorously addressing Hilbert's twenty-fourth problem through automated reasoning. Conversely, the scope and depth of that problem almost demand in many cases the use of such a program. Further, the type of proof discovered by this type of program is indeed a Hilbert-style, axiomatic proof. One might assert with some satisfaction that a symbiosis exists between the newly discovered Hilbert problem and the recently enhanced uses of an automated reasoning assistant.

References

[Barrow????] Barrow, J. *The World within the World*, publisher, place, ?

[Ernst2001] Ernst, Z. ???

[Fitelson2001] Fitelson, B., and Wos, L. ''Missing Proofs Found'', *J. Automated Reasoning*, special issue on Advances in Logic through Automated Reasoning, 2001.

[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.

[McCune1996] McCune, W., and Padmanabhan, R., *Automated Deduction in Equational Logic and Cubic Curves,* Lecture Notes in Computer Science, Vol. 1095, Springer-Verlag: Heidelberg, 1996. See http://www.mcs.anl.gov/home/mccune/ar/monograph/ for additional information.

[McCune2001] McCune, W., ????

[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.

[Minkowski????]

[Norton2000] Norton, J. D. ''Nature Is the Realization of the Simplest Conceivable Mathematical Ideas'', *Stud. Hist. Mod. Phys.* **31**, no. 2 (2000) 135-170.

[Thomas1975] Thomas, I., ‘‘Shorter Development of an Axiom’’, *Notre Dame J. Formal Logic* **16** (1975) 378.

[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.

[Veroff2001a] Veroff, R. ‘‘Finding Shortest Proofs: An Application of Linked Inference Rules’’, *J. Automated Reasoning* **21**, no. 2 (2001) 123-139

[Veroff2001b] Veroff, R. ‘‘Solving Open Questions and Other Challenge Problems Using Proof Sketches’’, *J. Automated Reasoning* **21**, no. 2 (2001) 157-174

[Wos1998] Wos, L., ‘‘Automating the Search for Elegant Proofs’’, *J. Automated Reasoning* **21** (1998) 135-175.

[Wos1999] Wos, L., and Pieper, G. W., *A Fascinating Country in the World of Computing: Your Guide to Automated Reasoning*, Singapore, World Scientific, 1999.

[Wos2000] Wos, L., and Pieper, G. W. *The Collected Works of Larry Wos*, Singapore, World Scientific, 2000.

[Wos2001] Wos, L., ‘‘The Strategy of Cramming’’, Preprint ANL/MCS-P898, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Illinois, 2001.

[Wos2002] Wos, L., and Fitelson, B. *Automated Reasoning and the Discovery of Missing and Elegant Proofs*, in preparation.