The Flowering of Automated Reasoning*

Larry Wos
Mathematics and Computer Science Division
Argonne National Laboratory
Argonne, IL 60439-4801
e-mail: wos@mcs.anl.gov

Abstract

This article celebrates with obvious joy the role automated reasoning now plays for mathematics and logic. Simultaneously, this article evidences the realization of a dream thought impossible just four decades ago by almost all. But there were believers, including Joerg Siekmann to whom this article is dedicated in honor of his sixtieth birthday. Indeed, today (in the year 2001) a researcher can enlist the aid of an automated reasoning program often with the reward of a new proof or a better proof in some significant aspect. The contributions to mathematics and logic made with an automated reasoning assistant are many, diverse, often significant, and of the type Hilbert would indeed have found most pleasurable. The proofs discovered by W. McCune's OTTER (as well as other programs) are Hilbert-style axiomatic. Further, some of them address Hilbert's twenty-fourth problem (recently unearthed by Rudiger Thiele), which focuses on the completion of simpler proofs. In that regard, as well as others, I offer challenges and open questions, frequently providing appropriate clauses to provide a beginning.

1 From Miniscule to Monumental

The style of this article is narrative, interweaving elements of history with developments in automated reasoning that presage a stirring and rewarding future. Here one meets a new generation of researchers in the field and learns of incarnations of Hilbert's recently discovered (by Rudiger Thiele) twenty-fourth problem. By presenting diverse results from a wide spectrum focusing on mathematics and logic, the explicit intent is to interest new researchers, as well as the experienced, in experimentation and application. To further that objective, this article includes very short stories, featuring research nuggets, methodologies,

*This work was supported 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 strategies to extend, modify, and experiment with. The number of topics covered and the length of this paper pay tribute to the long, long career of Joerg Siekmann. Although this article in no way comes close to being a thorough survey, it does offer some startling successes that would almost certainly have been out of reach were it not for the reliance on an automated reasoning program.

I choose to mark the birth of the field now known to many as automated reasoning with J. A. Robinson's introduction of binary resolution. Of course, earlier work on geometry and logic in the context of proving theorems was indeed important, but the era of concern here begins with Robinson's contribution. The inference rule binary resolution beautifully generalizes both modus ponens and syllogism. I prefer (because of questions of effective encoding) the definition of binary resolution that focuses on a single literal in each of two premisses and, therefore, prefer the inclusion of the inference rule factoring as well.

The cited rule of reasoning is strongly connected with a particular language. Indeed, a simply brilliant stroke was the choice of the clause language for representing questions and problems to the type of reasoning program that would evolve. Although on the surface one might find this language quite poor in comparison with natural language, its poverty actually facilitates the formulation of effective means for conclusion drawing and, perhaps of greater importance, facilitates the formulation of powerful strategies to control the reasoning. Without strategy, it seems to me, what I discuss shortly as successes would most likely not have occurred.

Almost at the beginning, when the Argonne paradigm (with its heavy emphasis on the use of strategy and on experimentation) was in its infancy, a key test problem was a five-clause example, the Davis-Putnam problem. Finding a proof for this inconsistent set of clauses by a computer program in negligible time marked an achievement but—especially in retrospect—a minuscule one that in no way foretold what was to come.

A bit more significant was our success with a classroom exercise in group theory. The theorem (actually too strong a classification) asks for a proof of commutativity for groups of exponent 2 (those in which the square of x is the identity e). Its proof was completed in less than 3 CPU-seconds on what today would be called a very, very slow machine. What did count was the first instance of what would become the Argonne paradigm. In that paradigm, the use of strategy is indispensable, clause retention plays a key role, and experimentation is most heavily emphasized, where the targets are actual results from mathematics and logic as opposed to syntactic problems. When the question is open, far better. Regarding the retention of newly deduced information, sometimes the sought-after proof is discovered after the keeping of more than 1,000,000 clauses and the generation of tens of millions. As for the occasionally-voiced objection that an unaided master would not proceed in such a manner, note that the past few years have witnessed the discovery of proofs that had eluded experts for decades; see Section 4.

This section has, till now, focused on the miniscule; next in order is the monumental,

a classification virtually demanding some justification and explanation before evidence is presented. In both mathematics and logic, at the simplest level, advances in the context of proof can be partitioned into first proofs and improved proofs. A beautiful example of the first class was the establishment by an automated reasoning program that every Robbins algebra is a Boolean algebra [McCune1997]. A charming example of the second class was the discovery (by a different automated reasoning program) of a proof focusing on Meredith's 21-letter formula for two-valued sentential (or propositional) calculus [Meredith1953], the following expressed in clause notation.

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

The proof of concern shows that this formula is a single axiom, but, more important, the proof consists of 38 applications of condensed detachment in contrast to the (in effect) 41-step proof of Meredith himself. I shall return to these two examples, as well as others, when I present in Section 4 evidence that "monumental" is well deserved.

Of a different nature, and perhaps more significant, is that element focusing on the source of the implied question or that on the expert who has shown interest in such a question. No less a person than Alfred Tarski devoted energy to the Robbins algebra question, offering it in a book [Henkin1971], having his students attempt to answer it, and posing it repeatedly to researchers. Regarding Meredith, he clearly was concerned with proof length, and he was a master at such simplification, in part attested to by his publishing (with Prior) an abridgment of a result of Lukasiewicz. But I can offer (to many) a far more persuasive view, justification, and explanation for employing the term "monumental"—indeed, Hilbert himself enters the picture.

In his 1900 Paris talk, Hilbert offered twenty-three problems that have had and continue to have a profound effect on mathematics. This greatest mathematician of the twentieth century also had a dramatic effect on logic. His emphasis on axiomatic proofs is but one example. The type of program I prefer (of which OTTER is the finest example in my view) produces Hilbert-style axiomatic proofs. But far more gold was yielded by that most powerful mind.

Only recently, because of the scholarship of Rudiger Thiele, has the world learned that in fact a twenty-fourth problem was offered by Hilbert (in one of his notebooks) [Thiele2001]. As Hilbert himself said, the problem was not included in the famous Paris talk because he had not as yet adequately formulated it. The problem focuses on the finding of simpler proofs. As he notes, mathematicians should simplify mathematics, not complicate it. But what makes a proof simpler, and in which specific ways is one proof simpler than another?

Ceteris paribus, for the most obvious case, proof **P** is simpler than proof **Q** if **P** is strictly shorter than **Q**. A second case focuses on formula (or equation) complexity, where the *formula complexity* (*equation complexity*) of a proof is k if and only if one of its deduced

steps consists of k symbols and all other deduced steps consist of k or fewer. A proof **P** is simpler than a proof **Q**, all things being equal, if the formula (equation) complexity of **P** is strictly less than that of **Q**. Where the *size* (gleaned from a conversation with D. Ulrich) of a proof equals the total number of symbols in its deduced steps, **P** is simpler than **Q** when **P** has strictly smaller size. A fourth and subtler case concerns lemma presence in a proof. The proof **P** is simpler than the proof **Q** when **P** avoids the use of some powerful lemma **L** whereas **Q** relies on the use of **L**. Finally, among the other cases, I find interesting that concerning the type of term present. In particular, for example, **P** is simpler than **Q** when **P** avoids the use of *double-negation terms* whereas **Q** relies on such. A double-negation term t is a term of the form $n(n(s))$ for some term s, where the function n denotes negation (as occurs, for example, in two-valued sentential calculus).

Based on the preceding, the reader may have formulated a most natural conjecture concerning what is about to be written:

- Automated reasoning programs can and do answer open questions by producing a first proof and answers others by discovering simpler proofs.

With the evidence to be presented in Section 4, many—or perhaps many, many— will join me in the attribution of *monumental* to the field of automated reasoning. Before such a presentation, in order is the nature of a *marvelous dream* and some details concerning the journey culminating in reaching that dream.

2 An Impossible Dream

In 1963 at Argonne National Laboratory, I began my study of what was then called *mechanical theorem proving*, a rather distasteful name for theorem proving certainly was not attacked mechanically. I was not fresh from graduate school, having been at the laboratory for six years. As for background, in 1954 I received my masters from the University of Chicago and in 1957 my Ph.D. from the University of Illinois (in group theory, under R. Baer). So I was well schooled in mathematics. Perhaps because of that schooling, I would (I am almost certain) have conjectured in 1960 that an attempt to find proofs with a computer seemed totally out of reach. Indeed, my few years of computing would most likely have enabled me to conclude that no way existed to communicate concepts such as commutativity, associativity, group, ring, and the like. And, even if such was possible, how would a program set about to search for and then find a proof? One lesson to be learned: Do not attempt to estimate the power of the mind, especially that of a researcher.

The lesser of two dreams—which had to be realized if the second and impossible dream was to be pursued—was to provide the means for a computer program to accept the assignment of searching for a proof. The second and impossible dream was to rely on such means,

embellish them, extend them, and then make significant contributions to mathematics and to logic. Indeed, could a computer emulate so well the mind of a researcher that new results would come into existence? Could the knowledge, experience, and intuition of a master be encoded? And how would a program recognize, among its drawn conclusions, those that would merit accolades?

To answer these questions, and others that merited answering, I embarked on an exciting journey—one that continues today. The stage will then be set for presenting the promised evidence.

3 A Lengthy Journey of Almost Four Decades

I shall in the main confine my treatment—from there (the early 1960s) to here (late 2001)—to items featuring the Argonne extended group of researchers. Of course, the work of others counts substantially. For example, one merely glances at the Boyer-Moore incredible success with program verification to experience amazement. A successor to the original program is now being used by the chip manufacturer AMD. Many deep theorems have been proved by students and by researchers with reliance on one of the incarnations of the Boyer-Moore effort. But I am sketching only a fraction of an almost-four-decade journey, not a full history.

For reasons not to be covered here, the 1960s witnessed relatively little experimentation; few reasoning programs existed. That paucity, according to me, hindered the needed advances. In contrast, from the beginning, Argonne emphasized experiments, featuring theorems taken from the literature. At first, and for many years, the group focused on finding proofs of known results. The explanation that was given—and I think a reasonable one—was that, if our program could not find proofs of theorems already proven, then how could we expect to answer open questions. Yes, I think it accurate to say that we were pursuing the impossible dream of eventually making important contributions, adding new knowledge in the form of new proofs. Nevertheless, I would have predicted, and in effect did so, that my lifetime would see little or no such successes. How wrong I was!

A more than casual observer, examining the evidence of the early and mid-1960s, would also have made the prediction that little or nothing would ever come of the effort devoted to eventually proving significant theorems. For example, although we at Argonne were somewhat gratified when our program proved that groups of exponent 2 are commutative, the theorem is very easy to prove and lacks much depth. When we turned to ring theory and found that, for the program to prove that minus x times minus y equals xy, we were forced to include two lemmas concerning the product of 0 and x, we were not filled with optimism. Our observer would have most likely scoffed with justification at the weakness of our automated search for proofs. For a third bit of negative evidence, for the program

to complete a proof that the square root of 2 is irrational, we were again forced to include a crucial lemma. Given this sampling of data, who would have predicted eventual success?

Sure, we had not yet automated any form of equality-oriented reasoning (paramodulation was our choice eventually), nor had demodulation been formulated. Equality was simply treated as just another relation, no different than the relation of containment, subset, or the like. Yes, hyperresolution had been formulated, but the program was still forced to rely on appropriate clauses to capture equality substitution and other properties. The three theorems just cited yield far more readily to treating equality as built in, as "understood". The set of support strategy—which had been formulated to conquer the exponent-2 theorem and which is still considered today to be the most powerful restriction strategy—was useful but not sufficient to crush the other two theorems. The unsatisfying study of the cited theorem from ring theory produced nothing of consequence. In contrast, the study of the number theory problem did lead to the formulation and introduction of demodulation; however, the power offered by a reasoning program was still indeed inadequate. So why did we press on; what made us believe the far distant future was more than promising?

I never asked others in the Argonne group that included (before 1980) George Robinson, Dan Carson, Lee Shalla, Ross Overbeek, Ewing Lusk, Robert Veroff, Brian Smith, Steve Winker. Therefore, I can speak only for myself, and then not with certainty. Clearly, strategy fascinated me. The improved power offered by my introduction of the unit preference strategy and the set of support strategy was indeed gratifying. I was often heard to say from the 1960s throughout the 1980s that new strategies were needed, and I expressed puzzlement at the lack of intense research devoted to their formulation—not that I myself was making many contributions in this context.

Also, at least for me, there was the excellent companionship of fine minds, sharing the belief in experimentation. The almost exclusive emphasis was on proving theorems from mathematics and logic, the exception being the work of Smith and Veroff, which was primarily concerned with program verification. Finally, and perhaps accurately, we accepted the nature of basic research, understood that the main objectives would be difficult to reach and possibly years away.

The value and importance of Overbeek's joining the group in the early 1970s cannot be overestimated. He was, and still is, a master at systems design, one of the best in the entire world of computing. He brought with him his program for proving theorems, followed by many later and improved versions. By then the field had survived being called "automatic theorem proving", which it certainly was not, and had more accurately become known as "automated theorem proving". The mid-1970s witnessed the publication of a paper devoted to experiments in theorem proving (in part coauthored by John McCharen, a member of the extended group for but a short time). But even then (in my view) far, far too little experimentation was featured throughout the world of automated theorem proving.

With Overbeek as the motivating force, in the late 1970s our journey became most haz-

ardous: We ventured into the dangerous and often unrewarding domain of *open questions*. As far as I know, with the exception of SAM's lemma proved in the mid-1960s, the field had provided nothing new to mathematics and to logic. We at Argonne were about to change that aspect to a small extent. Indeed, we first answered a set of small open questions from ternary Boolean algebra [Winker1978]. Prompted by that minor achievement and motivated by a conversation I had with I. J. Kaplansky in which I requested a reasonable target, we then answered a far more interesting question concerning involutions, antiautomorphisms, and semigroups [Winker1981]. Winker played the key role in both achievements.

Next, the logician John Kalman (expert in equivalential calculus) visited Argonne, bringing with him seven open questions concerning the status of certain formulas in the context of being single axioms for the cited area of logic. Winker and I answered six of the seven questions, with the status of the formula XCB even at this time (2001) still uncertain. Around the time Winker and I were collaborating (1980), I had introduced the term "automated reasoning", in part because of our study of conjectures, program verification, puzzle solving, and, of course, theorem proving. History shows that the term has lived on and gained rather wide acceptance.

In the early and mid-1980s, a vigorous debate rose concerning the value of clause retention. The Argonne paradigm insisted on retaining ever-growing sets of conclusions in the form of clauses. Various other paradigms did not share our enthusiasm for this aspect. We still maintain that such retention is crucial if a program is to provide substantial assistance in answering deep questions and solving hard problems.

The early 1980s marked yet another significant development: William McCune joined the group. He has proved to be a master of systems design—his program OTTER is (in my view) currently the most versatile and powerful reasoning program in existence. His monumental success (with his program EQP) in answering the Robbins algebra question (that had remained open for more than six decades) [McCune1997] and his monograph on open questions answered (with his colleague R. Padmanabhan) [McCune1996] attest to the vital role he has played and continues to play in automated reasoning. He and I (in the late 1980s) answered open questions in combinatory logic posed by R. Smullyan [Wos1993;McCune1987]. Roughly at the same time, Lusk with his colleague McFadden answered some tough questions concerning semigroup order [Lusk1987]. All of the questions we have cited and all we will cite were attacked with indispensable assistance from one of our reasoning programs. Their diversity provides strong evidence that the field was and is moving forward with increasing speed.

These varied and, in some cases, impressive successes in no way silenced the skeptics. Indeed, in a 1992 article published in the *New York Times*, mathematicians from various fields expressed a negative view of automated reasoning. Some asserted that, among the flaws, one could not learn from proofs produced through automated means. Of course, such is clearly not the case. For a first example, OTTER has discovered proofs of significant

theorems that avoid the reliance on thought-to-be-indispensable lemmas. For a second example, OTTER has found proofs in which so-called double-negation terms are absent, terms of the form $n(n(t))$ for some term t. A glance at the literature strongly suggests that logicians were unaware of such proofs and, further, may have believed such was not possible in many of the cases in which OTTER has succeeded. More generally, Michael Beeson has proved that, for certain axiom systems in certain areas of logic, one is guaranteed the existence of such double-negation free proofs when the theorem is free of double negation. A study of the proofs that avoid various lemmas or that avoid various classes of term can indeed be instructive and can teach much to both the student and the experienced researcher.

With the new millennium, the journey continues. New members have joined the extended Argonne group, among whom are Branden Fitelson, Kenneth Harris, and Zachary Ernst. Although not in the field, each alone as well as together have found the field of automated reasoning most engrossing and the use of OTTER so intriguing. Mathematicians are now assisted by a reasoning program, and chip designers have theorem-proving groups.

4 Recent Successes

Far more detail, including proofs and input files, relevant to the material of this section will be found in the planned book entitled *Automated Reasoning and the Discovery of Missing and Elegant Proofs.*

Still surprising to me, (from what it appears) not all researchers in automated reasoning consider applications the most important aspect. Although I clearly emphasize the proving of theorems in mathematics and in logic, I suspect that, from the viewpoint of science in general, program verification is the most important application. A close second is circuit and chip validation and design. However, as programs are continually enhanced to offer more and more power with the evidence of proving deeper and deeper theorems, the two cited applications will benefit. This section offers evidence of significant advances and is the pinnacle of this article in its demonstration of the power and value of today's automated reasoning program. In the main, I focus on successes that have occurred since McCune conquered the Robbins algebra problem and also answered numerous open questions discussed in his monograph. Challenges and open questions are offered in this section.

In both mathematics and logic, much energy has been devoted to finding axiom systems with appealing characteristics such as independence and smallness in number. Regarding the latter, often the limiting case has been reached, the discovery of a single axiom. In group theory, for example, McCune contributed (among others) the following single axiom [McCune1993], where the functions f and g respectively denote product and inverse.

```
f(x,g(f(y,f(f(f(z,g(z)),g(f(u,y))),x)))) = u.  % McCune (3.1.1)
```

Kunen then offered the following single axiom, an improvement in that its variable richness is three rather than four [Kunen1992].

```
f(g(f(y,g(y))),f(f(g(y),z),g(f(g(f(y,x)),z)))) = x.  % Kunen
```

Ideally, one might prefer a single axiom in which product, inverse, and the identity e were present, but such cannot be done, as proved by Tarski. However, researchers are offered an interesting open question. Does there exist a single axiom whose variable richness is three and whose length is equal to the McCune offering (shorter than that of Kunen)?

Lattice theory also admits single axioms. Until very recently, the shortest known (as far as I can determine) had length 79 with variable richness six, where "v" denotes union and "^" denotes intersection.

```
(((x ^ y) v (y ^ (x v y))) ^ z) v (((x ^ (((u ^ y) v (y ^ v)) v y)) v
(((y ^ (((u v (y v v)) ^ (w v y)) ^ y)) v
(v6 ^ (y v (((u v (y v v)) ^ (w v y)) ^ y)))) ^ (x v (((u ^ y) v
(y ^ v)) v y)))) ^ (((x ^ y) v (y ^ (x v y))) v z)) = y.
```

McCune and Veroff, with our colleague R. Padmanabhan, laid the groundwork for seeking a far shorter single axiom. McCune has succeeded, finding the following 29-symbol axiom of variable richness eight.

```
(((y v x)^x) v (((z^ (x v x)) v (u^x))^x1) )^ (w v ((v6 v x)^ (x v v7))) = x.
```

Immediately three open questions arise. First, does there exist a shorter single axiom? Second, does there exist a single axiom with less variable richness than eight whose length is, say, less than or equal to forty? Third, in the spirit of Hilbert's twenty-fourth problem focusing on proof simplification, does there exist for the McCune axiom a proof of length strictly less than 59?

The last question merits a bit more detail. In particular, I have a preference for proofs relying solely on forward reasoning, the denial being used to complete the proof. Such proofs have the appealing property of explicitly deducing the goal or goals of the theorem, or generalizations of such. I also prefer proofs in which demodulation is not present. To me (as well as other mathematicians with whom I have spoken) its absence often facilitates more insight into the nature of the proof. In contrast, a Knuth-Bendix bidirectional proof is usually easier to complete, sometimes far easier. Therefore, if one has the objective of producing a forward-reasoning demodulation-free proof, one might be wise to first seek a

Knuth-Bendix bidirectional proof and then embark on some form of proof translation. The process of proof translation is often rather difficult, presenting various obstacles.

I have found with OTTER's assistance a 59-step proof of the type I prefer. That proof has variable richness nine, which leads to yet another challenge. Can one find a proof for the given single axiom for lattice theory such that the proof has variable richness that does not exceed eight?

Next in order is Boolean algebra, a field that is studied in terms of various operators. Among the studies is that focusing on disjunction coupled with negation and that focusing on the Sheffer stroke. Regarding the latter, S. Wolfram suggested as candidate axiom systems a pair of equations and twenty-five equations to be considered separately [Veroff2001b]. Robert Veroff, with an intense effort relying on his use of his ingenious methodology called *sketches*, showed that the pair is indeed a 2-basis, an axiom system for Boolean algebra when presented in terms of the Sheffer stroke. Still using OTTER, he then showed that, if commutativity was coupled with any of the twenty-five candidates, an axiom system resulted. McCune then showed that commutativity could be proved dependent for two of the twenty-five, yielding two different single axioms. McCune then proved that their mirror images are also single axioms. But the study did not stop there.

Indeed, two of the newer members of the extended Argonne group, Branden Fitelson and Kenneth Harris, showed that no shorter single axiom in terms of the Sheffer stroke existed, shorter than length 15, the length of the cited single axioms [McCune2001]. Further, with the contribution of a summer student Andrew Feist, members of the group showed with models that seven of the twenty-five are too weak. Summarizing, open questions remain, each regarding the possible axiomatic status of sixteen of the twenty-five. Of those proved to be single axioms, where the function f denotes the Sheffer stroke, my favorite is the following.

```
f(f(f(x,f(y,x)),x),f(y,f(z,x))) = y.
```

But Boolean algebra had additional treasure to mine.

In particular, McCune conducted a lengthy study of this algebra in terms of disjunction and negation. His effort was rewarded, and mathematics was enriched. He found various 22-symbol axioms, including the following, where + denotes **or** and ~ denotes **not**.

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

Two open questions merit study. Does there exist a shorter single axiom than length 22 of this type? Does there exist a forward-reasoning, demodulation-free proof shorter than length 57 for the McCune axiom? The reader's conclusion is correct: I have such a 57-step step proof.

The preceding citations are generally placed under the province of mathematics rather than under logic. Consistent with my earlier and frequent references to mathematics *and* logic, the time has come for the latter to take center stage. Although not required, to provide more diverse evidence of the advances that have occurred, I shall feature theorems in which equality plays no role, in contrast to the preceding.

The researchers that played the main role in what is to be presented now are the newest generation, the newer members of the extended Argonne group. The first of these, Fitelson, entered the picture through an e-mail to me, asking my view of a rather intricate input file he had produced to prove a theorem of Lukasiewicz concerning the dependence of an axiom of Frege [Lukasiewicz1970]. Fitelson's (and OTTER's) proof were markedly different from that of Lukasiewicz. He had become interested in automated reasoning, so he informed me, because of reading my book entitled *The Automation of Reasoning: An Experimenter's Notebook and OTTER Tutorial.* I was so impressed by the nature of his included input file that I contacted him by phone, and we have conducted research together almost continually the past three years.

For a splendid example of the interworkings of our group, Veroff, learning of Fitelson's new 8-step proof of the cited Lukasiewicz theorem, made an important contribution. Specifically, Veroff showed that, quite likely, no shorter proof could be found and that there are a number of 8-step proofs, including the original Lukasiewicz proof [Veroff2001a]. The inference rule used is condensed detachment, encoded with a single clause of three literals and the use of hyperresolution. Veroff formulated a new use of linked inference rules to produce the cited result concerning different proofs and the likelihood that the shortest had been found.

Shortly thereafter, Fitelson answered an open question posed by Epstein. Quite different from the search for a single axiom, the question focused on possible axiom dependence among the following six axioms for two-valued sentential (or propositional) calculus, asking which (if any) of the six axioms is dependent on the remaining [Epstein1995]. The functions a, i, and n denote, respectively, conjunction, implication, and negation.

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

To give the reader (unaided or aided by some reasoning program) time to attack the at-one-time-open Epstein question uninfluenced by its answer, I shall delay presentation of the facts for a short while.

Early in our collaboration, Fitelson provided proofs, obtained with OTTER, of the two halves of an associativity relation that holds in infinite-valued sentential calculus [Fitelson2001]. As far as I know, these proofs marked an advance for logic in that each was the first such relying solely on condensed detachment for drawing conclusions. From the viewpoint of automated reasoning, the pair of proofs illustrates the value of strategy. Specifically, when Fitelson was able to obtain one of the two proofs but unable to obtain the other even after many CPU-hours, he turned to the use of the resonance strategy. In particular, he used the proof steps of the new proof (of one half of the relation) to direct the program's attack on finding the proof of the other half—and OTTER quickly succeeded. When equality is not involved (as is the case in the typical axiomatic treatment of infinite-valued sentential calculus), proofs based solely on condensed detachment are preferred. From Hilbert's writings, one might surmise he would indeed have recommended that proofs remain within the theory. In an important sense, such proofs are simpler than those in which the author instead relies on the use of equality-oriented reasoning.

With this vignette completed, it is time now for the answer to the open question offered by Epstein as to which (if any) of the given six axioms are dependent on the remaining.

The fourth axiom is dependent, and I now have a 10-step proof solely in terms of condensed detachment of its dependency. Does there exist a shorter proof relying solely on condensed detachment? As for the other five axioms, they form (as Fitelson showed with automated model generation) an independent set, which answers another of Epstein's questions. For a challenge, one might attempt to prove that the independent five-axiom set indeed axiomatizes two-valued sentential calculus by deriving from that set some previously known axiomatization. I have a 12-step proof that derives the following Lukasiewicz three-axiom system; a shorter proof in that context might exist.

```
% Lukasiewicz 1 2 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))).
```

The proof has added interest in the context of simpler proofs (which is relevant to the Hilbert twenty-fourth problem, but not regarding proof length). Indeed, although nine of the twelve steps rely on the presence of the function n for negation, the proof is free of double-negation terms, terms of the form $n(n(t))$ for any term t. I shall later return to that aspect of proof simplification in part because of the Hilbert problem, in part because of the emphasis on evidence of field advancement, and in part to illustrate what can be done with an automated reasoning program in the context of term avoidance and also the context of lemma avoidance.

Having found condensed-detachment proofs for an associativity law, Fitelson turned his attention to the far more difficult task of finding a condensed-detachment proof of a distribu-

tivity law in infinite-valued sentential calculus. He enlisted the collaboration of his long-term colleague Kenneth Harris, who is now yet another member of the extended Argonne group. Based on the literature and their collective insight, they decided that their likelihood of success would be increased if they began their attack by relying on equality-oriented reasoning. Therefore, rather than using condensed detachment and hyperresolution, Fitelson and Harris turned to paramodulation. After substantial effort in part by hand and in part by program, they had their proof in terms of equality. However, much work remained before the desired condensed-detachment proof would appear, for their proof relied on heavy use of demodulation and on bidirectional reasoning.

The presence of demodulation and steps that assert that s does not equal t for terms s and t and that result from reasoning backward from the denial each presented a rather severe obstacle. Fortunately, the group contains a master at proof conversion, Robert Veroff. He took the Fitelson-Harris proof and applied his methodologies to their proof, using his extended version of OTTER, and produced a forward-reasoning, demodulation-free proof relying solely on paramodulation. But the goal had not yet been reached, clearly.

McCune supplied the next key piece to the puzzle: He provided an algorithm to convert the paramodulation proof to a condensed-detachment proof. And the game was essentially won [Harris2001]. What remained was my role, which was to apply various methodologies to simplify the resulting proof in the context of length and term structure. In particular, regarding the latter, the final proof is free of double-negation terms, resulting from a move not often endorsed by all the members of the group. Indeed, our group can trace much of its success to the ability to disagree sharply about various aspects of the field and still perform as a well-integrated team. Such disagreements have in fact played a key role in the contributions in the areas of inference rule, strategy, implementation, and the like for which we have been credited.

Possibly because of McCune's success in finding single axioms for Boolean algebra in terms of disjunction and negation, Fitelson with Harris made a similar study for the implicational fragment of infinite-valued sentential calculus. Harris produced the following 69-symbol single axiom (not counting the predicate symbol).

```
P(i(i(i(x,i(y,x)),i(i(i(i(i(i(i(i(z,u),i(i(v,z),i(v,u))),
i(i(w,i(v6,w)),v7)), v7),i(i(i(i(v8,v9),v9),i(i(v9,v8),v8)),v10)),v10),
i(i(i(i(v11,v12),i(v12,v11)),i(v12,v11)),v13)),v13),
i(i(v14,i(v15,v14)),v16))),v16)).
```

Immediately two open questions come to mind. Is there a shorter single axiom? Is there a single axiom in strictly fewer than seventeen variables? Given my 15-step proof that uses the Harris axiom to derive the standard four-axiom system for the implicational fragment, one wonders if a shorter proof in that context exists.

```
%  the four-axiom system.
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(i(x,y),i(y,x)),i(y,x))).
```

Some months before this success, Fitelson and Harris had found new single axioms for two-valued sentential calculus in terms of the Sheffer stroke, among which is the following.

```
P((D(D(x,D(y,z)),D(D(x,D(y,z)),D(D(u,z),D(D(z,u),D(x,u))))))).
```

One might enjoy accepting the challenge of proving that the given formula is in fact a single axiom by deriving Nicod's single axiom, the following negated.

```
-P(D(D(a,D(b,c)),D(D(e,D(e,e)),D(D(f,b),D(D(a,f),D(a,f)))))) | $ANS(NICOD).
```

They proved that no shorter single axiom in terms of the Sheffer stroke exists.

The Fitelson-Harris studies naturally included attempts to prove that no shorter single axiom existed for various areas of logic, such as propositional logic. That area is axiomatized by the following single formula of Meredith.

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

Meredith's finding of this single axiom answered a question posed by Lukasiewicz when he offered in the mid-1930s his 23-letter single axiom for propositional calculus. Specifically, Lukasiewicz noted that three years had been devoted to finding the 23-letter formula, and he would leave it to others to find, if such existed, a shorter single axiom [Lukasiewicz1970]. Whether an axiom with strictly fewer than 21 letters (the length of the Meredith axiom) exists for this area of logic is still open. Fitelson and Harris have made substantial progress on this problem, but the work is not complete.

However, in their efforts, they found most valuable the addition of another colleague, Zachary Ernst, who is now yet one more member of the Argonne extended group. Ernst took to the type of problem under discussion beautifully. He also became enthralled with the use of OTTER. One of the areas Fitelson, Harris, and Ernst attacked was $C5$, the implicational fragment of $S5$. The logic known as $S5$ is a modal logic, in part formulated to capture more closely the widely accepted notion of implication, in contrast to that typically featured in other areas of logic. Meredith and others had found axiom systems for $C5$, among which is the following Meredith single axiom.

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

Ernst, through diligence, thoroughness, and brilliance, found six additional single axioms [Ernst2001b], among which is the following that exhibits a most interesting property.

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

The reader might enjoy comparing the two axioms with the objective of identifying the property I am about to discuss.

Before stating explicitly what that property is, I shall set the stage by quickly touching on the *resonance strategy* introduced in the early 1990s. My formulation of that strategy can be traced directly to Dana Scott's offering 68 theses (theorems of Lukasiewicz) for consideration by OTTER. A *resonator* is a formula or equation whose variables are treated as indistinguishable; its functional pattern is what matters. The researcher assigns a value to each included resonator. The program's reasoning is then directed accordingly, assigning the corresponding priority to any formula or equation that matches a resonator. Low assigned values give high priority to matching conclusions. The Nicod and Fitelson single axioms for propositional calculus in terms of the Sheffer stroke are in the same equivalence class, that class defined by taking either and treating it as a resonator. And the stage is set for answering the posed question concerning the two given single axioms (Meredith's and Ernst's) for $C5$.

The cited Ernst formula is the fourth of six new single axioms he found. The first three as well as that of Meredith all are in the same resonator-equivalence class. The fourth is not; its functional pattern differs from that of Meredith—indeed most gratifying!

But, where the *size* of a basis (axiom system) is measured in terms of the total number of symbols present, the Meredith 1-basis and each of the Ernst 1-bases have size 21, no difference. Naturally, one wonders about the existence of a smaller-sized basis, when no restriction is placed on the number of elements (axioms) in the basis. Ernst examined that question with the assistance of OTTER and produced a startling result. A basis of size less than 21 (the size of, for example, the Meredith basis) does exist, the following.

```
P(i(i(x,y),i(i(i(i(y,z),w),z),i(x,z)))).
P(i(x,x)).
```

This basis has size 18 (not counting predicate symbol occurrences). And, ensuring that no loose ends remained, the trio of Ernst, Fitelson, and Harris finished the game well, proving that no smaller in size (total symbol count) basis for $C5$ can exist. They then turned their attention to $C4$.

Apparently Meredith made a serious attempt to find a single axiom for $C4$ to no avail. Clearly, the problem was extremely difficult, if indeed such an axiom did exist. But, as it

turned out, Ernst, Fitelson, and Harris conquered the problem [Ernst2001b], finding this elusive single axiom, the following.

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

To prove that the given formula suffices, one might accept the challenge of attempting to deduce from it the following two-axiom system for $C4$.

```
P(i(i(x,i(y,z)),i(i(x,y),i(u,i(x,z))))).
P(i(x,i(y,y))).
```

The shortest proof I have found at this point has length 33. Their success leads naturally to an intriguing open question. Does there exist another single axiom for $C4$, or have my colleagues found the only such?

Inspired by the $C4$ success, Ernst turned to another area of logic, conducting an intense study of the implicational fragment of Dunn's classical relevance logic RM, which is called $RM \rightarrow$. Meyer and Parks provided the following impressive and independent basis for $RM \rightarrow$.

```
P(i(i(x,y),i(i(y,z),i(x,z)))).  % suffixing
P(i(x,i(i(x,y),y))).            % assertion
P(i(i(x,i(x,y)),i(x,y))).       % contraction
% Following is Parks's axiom -- GOAL
P(i(i(i(i(i(x,y),y),x),z),i(i(i(i(i(y,x),x),y),z),z))).
```

The system $RM \rightarrow$ is equivalent to the implicational fragment of Sobocinski's three-valued logic S. Perhaps a more satisfying basis exists, possibly of the same cardinality, of smaller size. In particular, perhaps the fourth and most complex axiom could be replaced by a less complex axiom.

The Ernst study did just that, yielding the following four-axiom system.

```
P(i(i(x,y),i(i(y,z),i(x,z)))).  % suffixing
P(i(x,i(i(x,y),y))).            % assertion
P(i(i(x,i(x,y)),i(x,y))).       % contraction
P(i(i(i(i(i(x,y),z),i(y,x)),z),z)).
```

He and OTTER found a replacement for the 21-symbol axiom, namely, an axiom of complexity thirteen. His research yielded even more, a second 4-basis, where the last given formula is replaced by the following.

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

Quite interesting, the two 13-symbol axioms are not in the same resonator-equivalence class. But the story is not yet complete.

Indeed, Fitelson, with his continual curiosity concerning axiom dependence, considered the two new Ernst bases for $RM \rightarrow$. And the world of logic was treated to a marvelous result: Fitelson showed that the axiom of contraction is dependent on the remaining three basis elements, for each of the Ernst bases [Ernst2001a]. Ernst, Fitelson, and Harris presented to logic a 3-basis of size 31 to replace the well-known 4-basis of size 48.

Immediately, two open questions offer themselves. Does there exist a basis of smaller size than 31? Does there exist a basis of two or fewer members? If the reader enjoys challenges regarding proof length, I note that I have in hand a 38-step proof that relies on the first Ernst basis and deduces the contraction axiom and the Parks axiom. For the second Ernst basis, I have found a 37-step proof.

At this point, I now turn to results directly pertinent to the Hilbert twenty-fourth problem, various types of proof simplification. In the context of proof length, Meredith and Prior were clearly interested [Meredith1963]. Indeed, they published an "abridgement" of a proof of Lukasiewicz, for the Lukasiewicz shortest single axiom for the implicational fragment f two-valued logic. The Meredith-Prior proof has length 33 (applications of condensed detachment), and the Lukasiewicz proof has length 34. One might, before reading any further, enjoy the challenge of using as hypothesis the following 13-symbol Lukasiewicz formula and attempting to deduce the Tarski-Bernays system.

```
%  following is shortest single axiom for the implicational fragment
P(i(i(i(x,y),z),i(i(z,x),i(u,x)))).
-P(i(p,i(q,p))) | -P(i(i(i(p,q),p),p)) | -P(i(i(p,q),i(i(q,r),i(p,r)))) |
  $ANS(TARSKI_BERNAYS).
```

Because Fitelson had witnessed a number of successes on my part in proof refinement with regard to length (some of which I report here), he brought the Meredith-Prior result to my attention. He posed for me the problem of finding a further abridgment. As one justifiably would predict, I thought the chances small in view of the mastery of logic evident in Meredith's and Prior's works. Notwithstanding, the problem was most intriguing, and I had OTTER for a companion and a powerful assistant.

The sought-after proof of length 32 or less did indeed prove elusive. The use of the resonance strategy, of Veroff's hints strategy, of a methodology that blocks steps of a given proof one at a time—none of these won the game. However, by using as resonators the Meredith-Prior proof and the Lukasiewicz proof and various methodologies, OTTER did complete another 33-step proof. That proof contained as a subproof a 30-step proof of the

third member of the cited Tarski-Bernays system. And an idea was born, a possible new strategy, one that would be called *cramming*.

Intuitively, in the first incarnation, the object of the strategy is to reduce the length of the proof of a target conjunction by focusing on the proof of one member and attempting to "cram" as many of its steps into the needed remaining proofs. In the case of the target three-axiom Tarski-Bernays system, the plan (if successful) was to focus on the 30-step proof of the third member and cram so many of its steps into the desired proofs of the other two members that only two additional formulas would be needed. For this to occur, there must exist two pairs of clauses among the thirty such that condensed detachment applied to the pairs yields, respectively, the first and second members of the three-axiom system—highly unlikely, but possible. Fortunately, OTTER offers just what is needed to examine all pairs, including pairs in which the two elements are identical. In particular, with the command set(sos_queue), the program conducts a breadth-first search. Therefore, one merely places the (in this case) thirty clauses that prove the third member of the Tarski-Bernays system in the initial set of support. One additional move must be made to attempt to prevent the program from retaining unwanted conclusions, conclusions other than the first and second members of the three-axiom system. The move consists of placing their correspondents in a hints list and assigning a very small maximum on the complexity of newly retained conclusions.

If the preceding had failed, clearly with almost certainty this episode would not be included here. Indeed, the story as expected ends with success, with the completion of a 32-step proof that is an abridgment of the Meredith-Prior abridgment. Especially for the thorough historian who enjoys the vagaries of science, I note that later experiments yielded different 30-step proofs of the third member, none of which permitted completion of the desired 32-step proof. Because of the charm of this result and its significance to both automated reasoning and logic, an open question is virtually demanded. Does there exist a proof that uses the Lukasiewicz shortest single axiom as hypothesis and that has as target the Tarski-Bernays system with strictly fewer than thirty-two applications of condensed detachment?

But reduction in proof length is just one refinement reflecting Hilbert's intest in simpler proofs. Quite different is that concerning *variable richness*, where the variable richness of a proof is k if and only if one of its deduced steps relies on exactly k distinct variables and all other deduced steps rely on k or fewer. If one has in hand a proof of variable richness, say, k and wishes a refinement of that proof whose richness is strictly less than k, OTTER offers precisely what is needed. One can include assign(max_distinct_vars,j), and the program will retain a newly deduced conclusion only if it relies on j or fewer distinct variables. Deepak Kapur conducted an investigation with the objective of finding a proof for Meredith's single axiom for two-valued logic that completes with the deduction of the cited Lukasiewicz three-axiom system such that its variable richness is six. Meredith's proof has richness seven, containing two deduced steps out of forty-one in which seven distinct variables are present.

Kapur in fact succeeded in this context of proof refinement, producing with OTTER a 63-step proof of richness six. In answer to an anticipated question, I now have in hand a 49-step proof of variable richness six, only one of whose steps has that richness. As for answers to the next probable series of questions, no proof exists with richness strictly less than five, which can be seen by noting that the first condensed detachment step has that richness and is the result of applying condensed detachment to two copies of the Meredith single axiom. Yes, there does exist a proof of variable richness five, and I have in hand (because of OTTER) one of length 68.

Of still a different nature is proof refinement with respect to *term structure*. For example, all things being equal, a second proof is simpler than the first when the second avoids the use of double-negation terms, whereas the first relies on them heavily. Meredith's (in effect) 41-step proof for his single axiom for two-valued logic includes seventeen steps relying on double negation. If a researcher wishes to avoid such terms or to avoid some other class, OTTER again comes to the rescue through the use of demodulation or weighting. For example, by including the demodulator of the form (n(n(x)) = junk) with others to propagate the corresponding rewrite, double-negation terms can be avoided among retained clauses. I have in hand many, many proofs that the literature suggests require the use of double negation but that avoid it entirely. In fact, almost never did I fail to find such a proof for the theorem under consideration. I was thus curious about the conditions that guaranteed the existence of a double-negation-free proof.

D. Ulrich beautifully refined my concern by asking whether there existed an axiom system for two-valued logic such that, whenever the theorem to be proved was free of double negation, a proof could be found that was also free of double negation. Thus began a collaboration that first included the newest member of the extended Argonne automated reasoning group, Michael Beeson; later Veroff joined in the collaboration. Beeson answered the Ulrich question (which astounded me) by providing most of the details for a proof showing that the cited three-axiom system of Lukasiewicz has the desired property. Veroff and I supplied proofs of some needed lemmas, through (of course) OTTER's cooperation. Perhaps more astounding, Beeson (with some assistance by Veroff and me) then proved the corresponding theorem for infinite-valued sentential calculus with the following axiom system.

```
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))).
```

One might quickly conjecture that proof simplification in one aspect comes at the expense of simplification in another. For example, the removal of double-negation terms most likely results in a rather longer proof than that in hand. Similarly, the avoidance of thought-to-be-indispensable lemmas, a nice refinement, quite likely lengthens the resulting

proof. I can report with satisfaction and with some amount of awe that, often, such is not the case. For but one example, infinite-valued sentential calculus takes center stage, with the preceding four axioms in focus. A fifth axiom (the following) once thought necessary but later proved by that master Meredith dependent was the target.

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

Before OTTER and I attacked the problem, the literature offered a 37-step proof, one that relied upon double negation. Further, a review of the literature suggested that the following three lemmas might be indispensable.

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

After years of study, frequently interrupted by other investigations, I finally have in hand a 30-step proof, free of double negation, and avoiding the use of all three cited lemmas. I know of no shorter proof than that of length 30, regardless of its other properties. In other words—and this example is by no means isolated—a refinement in one aspect does not necessitate a cost in one or more other aspects.

The student, the experienced researcher, the simply curious—each might wonder why such proofs were missing from the literature and, apparently in many cases, perhaps out of reach of the unaided. Perhaps today's powerful and fast computers are essentially the answer. In my view (shared by others), such is indeed not the case. Rather, the answer lies in the use of strategy and methodology that permits and even encourages the program to search in the vast space of conclusions where no one had searched before. Indeed, some of the approaches that have led to many of the cited successes (and others not presented here) could be classed as counterintuitive.

5 Highlights and Perspective

This article is written in honor of Joerg Siekmann's sixtieth birthday; Joerg and I have known each other for decades. To put all in perspective, I sample a bit of history (at one end of the spectrum), and (at the other) I focus on recent significant contributions to mathematics and to logic that resulted directly from the use of an automated reasoning program. The results would have been out of reach, so it appears, were it not for the introduction of various strategies and methodologies that depend on them. Heavy experimentation was the key as well as the ability to retain hundreds of thousands of new conclusions.

The article is intended to stimulate a wide audience, including those not primarily concerned with automated reasoning. Perhaps not obvious, that audience includes those mainly interested in circuit design and validation and those interested in program synthesis and verification. Indeed, for a small hint as to why such may be the case, I note that the reduction in the length of a constructive proof typically corresponds to a reduction in the complexity of the object being constructed, a circuit or a bit of computer code, for example. Proof simplification in the context of length, as well as in other contexts such as variable richness, are directly pertinent to the newly discovered Hilbert's twenty-fourth problem. To add to the possible usefulness of this article, I include open questions, challenges, some detail concerning strategy, some concerning methodology, and more.

I mark the birth of the current incarnation of automated reasoning with J. A. Robinson's introduction of binary resolution. In its beginning, the field could boast only of conquering the miniscule. However—and I say this in part directly to Joerg, who clearly shared with me and others the dream of automation—here in the year 2001, automated reasoning in the Olympics of science merits a gold medal for its achievements. They are many, and they are diverse.

At one end of the spectrum is the use of reasoning programs by chip manufacturers that include Intel and AMD; Robert Boyer and J Moore pioneered this important subfield of program verification. At the other end of the spectrum—but of equal significance in the long run—reasoning programs are making important contributions to both mathematics and logic. For a taste of the fecundity of the field, one need only turn to the Web site of Johan Belinfante, finding there hundreds and hundreds of proofs for theorems from set theory [Belinfante2001], proofs essentially out of reach but a decade ago. In that regard, for the researcher seeking a deep problem on which to work, (as suggested by Boyer) one might seek a small set of axioms, say, sixty or fewer, that captures most of the set-theoretic reasoning needed for mathematics and logic outside set theory. For a taste of the breadth of automated reasoning's successes, one need only sample the material in this article or browse among the following highlights.

5.1 Axiom Systems with Various Properties

In both mathematics and logic, fine minds have devoted substantial research to the discovery (if such exists) of single axioms. In that context, automated reasoning programs (specifically McCune's OTTER) have proved to be a powerful reasoning assistant, and sometimes even a colleague, as the following highlights from the work of the Argonne extended group indicate.

- Boolean algebra, in terms of the Sheffer stroke (Veroff, McCune, Fitelson, Harris, Feist)
  - from 25 candidate equations (supplied by Wolfram), proof that 2 are single ax-

ioms, and then proved that their mirror images are also
  - proof that 7 of the 25 are too weak to be single axioms
  - proof that there exists no shorter axiom than length 15

- Boolean algebra, in terms of disjunction and negation (McCune)
  - discovery of ten 22-symbol single axioms

- Lattice theory (McCune)
  - discovery of a 29-symbol single axiom for lattices: far more attractive than the 40,000,000-symbol single axiom found algorithmically

In the preceding successes, equality is featured. As the experienced experimenter knows, equality presents various problems because of its nature. Paramodulation, for example, is a term-oriented inference rule in contrast to, say, hyperresolution, which is a literal-oriented rule. In logic, condensed detachment also is literal oriented, and its study therefore involves fewer obstacles than when equality is dominant. Nevertheless, problems in which condensed detachment is the sole rule of inference are most challenging. The following examples highlight some of our recent achievements.

- C5 (Ernst, Fitelson, Harris)
  - six new single axioms, three in the same resonator-equivalence class as Meredith's single axiom
  - shortest possible 2-basis
  - proof that no single axiom shorter than 21 symbols exists
  - proof that a basis with 18 symbols exists, the shortest possible

- C4 (Ernst, Fitelson, and Harris)
  - first single axiom
  - proof that a single axiom with 21 symbols exists, the shortest possible
  - proof that a 20-symbol 2-basis exists, the shortest possible

- RM$\rightarrow$ (Ernst, Fitelson, Harris)
  - smaller 4-basis than that of Meyer and Parks
  - 3-basis of size 31 (replacing 4-basis of size 48 of Meyer and Parks)

- Propositional logic, in terms of the Sheffer stroke (Fitelson and Harris)
  - new single axioms all of length 23, and no shorter exists

5.2 Proofs with Various Properties

Of an apparently different cast from the pursuit of axiom systems with various properties is the pursuit of proofs with various properties. I say "apparent" because size of basis (among other properties) and cardinality of basis have their obvious counterparts in proof, namely, the total number of symbols in the deduced steps of a proof and the precise number of such steps. For the past decade, I have devoted much research to such proof properties and to proof refinement (in the spirit of Hilbert's twenty-fourth problem). Here I highlight a few of my successes.

- Proof length
  - first proof (length 200) and shortest proof (length 50) deducing Lukasiewicz's three-axiom system from his 23-letter formula
  - 38-step proof, improving on Meredith's 41-step proof of his single axiom for two-valued logic
  - 30-step proof of the dependence of one of Lukasiewicz's five axioms for infinite-valued sentential calculus
- Term and lemma avoidance
  - cited shortest known proof (length 30) of axiom dependence in infinite-valued sentential calculus (1) avoids three seemingly crucial lemmas and (2) avoids double-negation terms

Intrigued by my success with the avoidance of double negation terms, M. Beeson (with assistance from Veroff and me) proved a charming metatheorem for two-valued sentential calculus: If the theorem **T** to be proved is free of double negation, then there must exist a proof **P** of **T** in which double negation is totally absent that relies on condensed detachment alone and that uses as hypotheses the Lukasiewicz three-axiom system. Beeson (with assistance from Veroff and from me) also showed that the independent four-axiom system for infinite-valued sentential calculus presented earlier in this article has the corresponding property to that of the Lukasiewicz three-axiom system.

All of these successes are directly related to Hilbert's twenty-fourth problem (as discovered by R. Thiele in his examination of Hilbert's files). The focus of that newly discovered problem is *simpler proofs*—whether of proof length or size, of term avoidance, of formula complexity, or of variable richness. Imagine my delight when I learned that my decade-long fascination with proof simplification had been a subject of considerable interest also to one of the masters of mathematics!

5.3 Techniques for Refining Proofs

The key to proof simplification is not, as some might suppose, the CPU speed of today's computer; rather, it is the availability of powerful strategies and effective methodologies. The past few years of my research have resulted in several new approaches that often aid in proof refinement—particularly with respect to proof length (applications of condensed detachment). The approaches also are effective where equality is the sole or dominant relation, and they are effective often in finding a first proof or settling a conjecture. These approaches include *blocking* proof steps one at a time (through demodulation or weighting), directing the program's reasoning with the McCune's *ratio strategy* (which blends breadth first with choosing the initiator of an inference rule application by conclusion complexity), and *cramming* the steps of a proof of one member of a conjunction into the needed proofs of the other members.

Interesting—and indeed counterintuitive—is the fact that a reduction in proof length may result from the study of proof refinement in some other aspect, such as lemma avoidance or term avoidance. The avoidance of thought-to-be-indispensable lemmas, when successful, is a sometimes overlooked aspect of proof simplification, as is the avoidance of some class of term. For example, the cited proof of axiom dependence in infinite-valued sentential calculus (the one that avoids the use of three lemmas and double-negation terms) possesses one additional remarkable property: It has length 30 (applications of condensed detachment), the shortest proof I know of.

Is the reduction in length the result of the successful refinements in the two other aspects? I strongly suspect so. Indeed, were it not for instructing OTTER to avoid the so-called key three lemmas and avoid double-negation terms (each through the use of demodulation), I believe that the 30-step proof may have lay hidden forever. Its discovery (so it appears) was because of forcing the program to explore within the space of conclusions where it would ordinarily not spend the majority of its time. That part of the space would typically not be explored by the unaided master, I believe, because of its counterintuitive nature, avoiding key lemmas and avoiding double negation. I suspect that the density of proofs of interest is greater in the restricted space of conclusions than in the entire space of conclusions.

For a glimpse of what I think is the case, one might imagine spending one's research life in algebras in which associativity is present and then being asked to (intuitively) study a nonassociative algebra. One's experiences would almost certainly interfere with sought-after proofs. Put a slightly different way, one of the marvelous features of using a program such as OTTER is the capability to *explore where no researcher has gone before.*

5.4 Evolution and Revolution

The field has come a long way. Even its name has evolved, from mechanical theorem proving to automatic theorem proving to automated theorem proving and, most accurately today, to automated reasoning. At first, some satisfaction was drawn from finding proofs for essentially syntactic examples; now research is often rewarded with a significant contribution to mathematics, logic, or some other discipline. For but one example of the latter, McCune's work on ortholattices [McCune1998] has served physicists well. He was asked to find a small ortholattice in which the following equation fails, where v denotes join, ^ denotes meet, and ′ denotes complement.

```
((x v y') ^ ((( x ^ y) v (x' ^ y)) v (x' ^ y'))) = ((x ^ y) v (x' ^ y'))
```

MACE (a model generation program of McCune) found a size-12 example. This example was very useful to Pavicic and Megill [Pavicic1999]. More generally, the Argonne paradigm that emphasizes the use of strategy, conclusion retention, reliance on the clause language, and vast amounts of experimentation is triumphing.

I think one can say with certainty that Hilbert would have found great pleasure in the successes of our field, not the least of which are those concerning proof simplification. Hilbert's recently discovered twenty-fourth problem may stimulate experimentation and research, *within* and *without* automated reasoning. More open questions are needed, more proofs to refine, more challenges to address. I for one would enjoy receiving such, preferably from mathematics or logic.

Joerg, the field is winning and, perhaps even more important, the second derivative of progress is positive. As for the future, the field will never lose sight of the goal, that of proving ever deeper theorems and contributing to design and verification in other disciplines. Finally, researchers have access to programs that often function as automated reasoning colleagues.

References

[Belinfante2001] Belinfante, J., Computer Assisted Proofs in Set Theory, Web site http://www.math.gatech.edu/ belinfan/research/autoreas/index.html, 2001.

[Epstein1995] Epstein, R., *Propositional Logics: The Semantical Foundations of Logic*, Oxford University Press, Oxford, 1995.

[Ernst2001a] Ernst, Z., "A Concise Axiomatization of RM→", *Bulletin of the Section of*

Logic, to appear.

[Ernst2001b] Ernst, Z., Fitelson, B., Harris, K., and Wos, L., "Shortest Axiomatizations of Implicational S4 and S5", Preprint ANL/MCS-P919-1201, December 2001.

[Fitelson2001] Fitelson, B., and Wos, L. " Missing Proofs Found", *J. Automated Reasoning* **27**, no. 2 (August 2001) 201–225.

[Harris2001] Harris, K., and Fitelson, B. "Distributivity in L-Aleph-0 and other Sentential Logics", *J. Automated Reasoning* **27** (2001) 141–156.

[Henkin1971] Henkin, L., Monk, J. D., and Tarski, A., *Cylindric Algebras I*, North-Holland, Amsterdam, 1971.

[Kunen1992] Kunen, K., "Single Axioms for Groups", *J. Automated Reasoning* **9**, no. 3 (December 1992) 291–308.

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

[Lusk1987] Lusk, E., and McFadden, R., "Using Automated Reasoning Tools: A Study of the Semigroup F_2B_2", *Semigroup Forum* **36**, no. 1 (1987) 75–88.

[McCune1987] McCune, W., and Wos, L., "A Case Study in Automated Theorem Proving: Finding Sages in Combinatory Logic", *J. Automated Reasoning* **3**, no. 1 (February 1987) 91–108.

[McCune1993] McCune, W., "Single Axioms for Groups and Abelian Groups with Various Operations", *J. Automated Reasoning* **10**, no. 1 (1993) 1–13.

[McCune1996] McCune, M., and Padmanabhan, R., *Automated Deduction in Equational Logic and Cubic Curves*, *Lectures Notes in Computer Science 1095*, Springer-Verlag, New York, 1996.

[McCune1997] McCune, W., "Solution of the Robbins Problem", *J. Automated Reasoning* **19**, no. 3 (1997) 263–276.

[McCune1998] McCune, W., "Automatic Proofs and Counterexamples for Some Ortholattice Identities", *Information Processing Letters* **65** (1998) 285–291.

[McCune2001] McCune, W., Veroff, R., Fitelson, B., Harris, K., Feist, A., and Wos, L., "Short Single Axioms for Boolean Algebra", *J. Automated Reasoning* (accepted for publication).

[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.

[Pavicic1999] Pavicic, M., and Megill, N., "Non-Orthomodular Models for Both Standard Quantum Logic and Standard Classical Logic: Repercussions for Quantum Computers", *Helv. Phys. Acta* **72**, no. 3 (1999) 189–210.

[Thiele2001] Thiele, R., and Wos, L., "Hilbert's Twenty-Fourth Problem", Preprint ANL/MCS-P899-0801, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL, 2001.

[Veroff2001a] Veroff, R., "Finding Shortest Proofs: An Application of Linked Inference Rules", *J. Automated Reasoning* **27**, no. 2 (August 2001) 123–139.

[Veroff2001b] Veroff., "Solving Open Questions and Other Challenge Problems Using Proof Sketches", *J. Automated Reasoning* **27**, no. 2 (August 2001) 157–174.

[Winker1978] Winker, S., and Wos, L., "Automated Generation of Models and Counterexamples and Its Application to Open Questions in Ternary Boolean Algebra", in *Proceedings of the Eighth International Symposium on Multiple-Valued Logic, Rosemont, Illinois,* IEEE and ACM, May 1978, pp. 251–256; reprinted in [Wos2000, pp. 286–297].

[Winker1981] Winker, S., Wos, L., and Lusk, E., "Semigroups, Antiautomorphisms, and Involutions: A Computer Solution to an Open Problem, I", *Mathematics of Computation* **37**, no. 156 (October 1981) 533–545 (October 1981); reprinted in [Wos2000, pp. 315–329].

[Wos1993] Wos, L., "The Kernel Strategy and Its Use for the Study of Combinatory Logic". *J. Automated Reasoning* **10**, no. 3 (1993) 287–343; reprinted in [Wos2000, pp. 1221–1287].

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

[Wos2000] Wos, L., with Pieper, G. W., *Collected Works of Larry Wos*, 2 vols., World

Scientific, Singapore, 2000.