The Arrival of Automated Reasoning

Larry Wos,[1]
M. Spinks[2]
[1]Mathematics and Computer Science Division
Argonne National Laboratory, Argonne, IL 60439
wos@mcs.anl.gov
[2]Department of Philosophy
La Trobe University, Bundoora, Vic 3083 Australia
mspinksau@yahoo.com.au

Abstract

For some, the object of automated reasoning is the design and implementation of a program that offers sufficient power to enable one to contribute new and significant results to mathematics and to logic, as well as elsewhere. One measure of success rests with the number and quality of the results obtained with the assistance of the program in focus. A less obvious measure (heavily in focus here) rests with the ability of a novice, in the domain under investigation, to make significant contributions to one or more fields of science by relying heavily on a given reasoning program. For example, if one who is totally unfamiliar with the area of study but skilled in automated reasoning can discover with an automated reasoning program impressive proofs, previously unknown axiom dependencies, and far more, then the field of automated reasoning has indeed arrived. This article details such—how one novice, with much experience with W. McCune's program OTTER but no knowledge of the domains under investigation, obtained startling results in the study of areas of logic that include the *BCSK* logic and various extensions of that logic. Among those results was the discovery of a *variety* weaker than has been studied from what we know, a variety that appears to merit serious study, as, for example, does the study of semigroups when compared with that of the study of groups. A quite different result concerns the discovery of a most unexpected dependency in two extensions of the *BCSK* logic.

1 Setting the Stage

When a researcher, who is a master of some field, uses an automated reasoning program and finds a proof of a significant theorem in said field, applause is more than appropriate. That success contributes to the mystique of automated reasoning, providing yet one more bit of evidence that substantial progress has occurred. Evidence of this type exists that includes studies of K. Kunen [Kun92], D. Phillips [Phi], and J. Belinfante [Bel01]. An expert, if the program in use provides the appropriate means, can give advice, make enlightened conjectures, and otherwise restrict and direct the program's attack in a manner that sharply increases the likelihood of success. For example, through the use of R. Veroff's hints strategy [Ver96] or with the resonance strategy [Wos95], one can guide the program toward or away from paths of reasoning. In addition, the researcher can restrict the program's attack by instructing it to avoid certain lemmas and certain types of term (through the use of demodulation) and, most effective, block the program (with the set of support strategy [WRC65]) from applying inference rules to sets of hypotheses whose consideration could bury the program in irrelevant conclusions. Sometimes a paper results, stating clearly that automated reasoning played an important role, perhaps a vital role. Even better is the case when the paper is published in a journal devoted to mathematics or to logic rather than to automated reasoning.

In contrast, one might consider the case in which a novice, an amateur, in the field of study in focus makes important discoveries by relying on a reasoning program. (By a novice is meant one who knows nothing of the area under investigation, but one who may know much about automated reasoning.) If the discoveries include impressive proofs, previously unknown axiom dependencies, and far more, a landmark has been reached, one that predicts greatness for the future of automated reasoning. This article offers a story of such discoveries, a story of a novice studying the *BCSK* logic, as well as extensions of that logic, with absolutely no knowledge of the fields under study.

At this point, we briefly provide some of the underlying formalism. Recall from [BP94] that the *fixedpoint discriminator* on a set A is the function $f : A^3 \rightarrow A$ defined for all a, b, c in A by

```
f(a, b, c) = c if a = b
             1 otherwise
```

for some element $1 \in A$. The element 1 is called the *discriminating element*. The fixedpoint discriminator arises naturally in algebraic logic as a generalization of the ternary discriminator; see, for instance, [BP94].

The *generic fixedpoint discriminator variety*, in symbols FPD_1, is the variety generated by the class of all algebras $\langle A; f, 1\rangle$ of type $\langle 3, 0\rangle$, where f is the fixedpoint discriminator on A and 1 is a nullary operation, the range of which is the discriminating element of f. The 1-assertional logic of FPD_1, in symbols $S(FPD_1, 1)$, is the consequence relation from sets of terms to terms determined by the equivalence $\Gamma|-_{S(FPD_1,1)}\phi$ if and only if $\psi = 1 : \psi \in \Gamma \models_{FPD_1} \phi$. Since FPD_1 is a variety $|-_{S(FPD_1,1)}$ and is both finitary and substitution invariant, and hence is a deductive system in the sense of Blok and Pigozzi [BP99], our interest in *BCSK* logic stems from the observation, made in [BSV], that it is formula equivalent to $S(FPD_1, 1)$.

The following nine axioms, for the *BCSK* logic, initiated the study, where the functions i and j denote *strong* and *weak* implication, respectively.

```
P(i(x,i(y,x))).                          %  A1
P(i(i(x,i(y,z)),i(i(x,y),i(x,z)))).      %  A2
P(i(i(i(x,y),x),x)).                     %  A3
P(i(x,j(y,x))).                          %  A4
P(i(j(x,j(y,z)),j(j(x,y),j(x,z)))).      %  A5
P(i(j(x,j(y,z)),j(y,j(x,z)))).           %  A6
P(i(j(j(x,y),x),x)).                     %  A7
P(i(j(i(x,y),y),j(i(y,x),x))).           %  A8
P(j(i(x,y),j(x,y))).                     %  A9
```

The nature of the contributions in focus here strongly suggests that automated reasoning has indeed arrived.

Detailed is the adventure that began with a set of axioms and target theorems in logic, a collaboration of one expert in that logic with another expert in automated reasoning—but truly a novice in the logic—and some hard-to-find proofs mostly supplied by Veroff using his powerful sketches [Ver01] approach. (More generally, an individual knowing essentially nothing about a field to be studied—but having much knowledge and experience with a reasoning program in hand—can make significant contributions to that field. Further, if the corresponding logical formulation is available, one who is a novice in logic or mathematics can fearlessly seek one valuable proof after another with the expectation of some or much success by relying on a program that offers a variety of strategy. At the other end of the spectrum, one who possesses substantial knowledge of the field to be studied but knows little of automated reasoning can also succeed.) A novice in the field under study has the advantage of not being trapped by knowledge of how one typically proceeds (perhaps, implicitly, must proceed) and can, therefore, follow paths not previously explored. The automated reasoning program has this advantage, for it knows nothing of any field and lacks bias or orientation—such a program is indeed a novice. For example, one can instruct the program to totally avoid some type of term or some lemma that the literature suggests must be relied upon. Such instruction can lead to most satisfying and wondrous discoveries. Well demonstrated here is the current state of automated reasoning in that, eventually, (as one learns here) startling results were obtained.

Our objective was to find "short" proofs, and seeking that objective led to marvelous and unexpected discoveries, which is the basis of the story to unfold. Not surprising, especially to the person familiar with the new book *Automated Reasoning and the Discovery of Missing and Elegant Proofs* [WP03] and the strategies and methodologies it offers, a number of short proofs were in fact completed. More pertinent to this article, unknown axiom dependencies were found, a new concept was formulated—*proof dependence*—and a variety was unearthed that may merit serious study. In particular, by way of a foretaste, of the nine axioms that prompted the original study, two were proved to be dependent, $A3$ and $A6$. These two axioms as well as $A7$ were shown to be totally avoidable (even as deduced formulas) for completion of the proofs being sought, which revealed a promising weaker variety to consider. This variety is axiomatized with axioms 1, 2, 4, 5, 8, and 9.

When one is introduced to some field of mathematics or logic, one is typically presented with a set of axioms from which the theorems are deducible. And that is how this story begins, with nine axioms of the *BCSK* logic. A glance at the set of axioms (of the area in focus) often does not readily reveal which, if any, are dependent on the remaining. For example, if one is introduced to group theory with the axiom set consisting of associativity of product, the existence of a two-sided identity element e, and, with respect to the identity, a two-sided inverse, one might not immediately see that dependencies exist among the given five axioms. But, they are present. Indeed, the axioms of right inverse and right identity are each dependent on the remaining three; equally, those of left inverse and left identity are dependent.

The proofs of the cited dependencies are well within reach of various automated reasoning programs or well within reach of the unaided researcher. With W. McCune's program OTTER, one can simply negate the axiom to be proved dependent, place the negation in the passive list, place the other axioms in the initial set of support list, and seek (when the notation is equational) with the inference rule paramodulation a proof by contradiction, which, at least for group theory, will be in hand almost at once.

Rather than the deducibility from the remaining axioms, the key focus for this article about dependence is that, at least axiomatically, the dependent item is not needed (in the input). However, a dependent axiom might still be required to complete one or more proofs of interest, needed at the deduced level. Here, as one learns, we are concerned with items that are not needed even at the deduced level, a topic that is featured as we introduce the notion of *proof dependent.* For a foretaste of what is to come, we note that the total avoidance of some thought-to-be indispensable lemma when seeking to complete a proof of a theorem of substantial interest can be challenging. Because of the nature of dependence and the importance of axioms, to totally avoid the use of an independent axiom may be far more challenging and, if successful, may mark the beginning of a study of a weaker variety (field), as in the case of the study of groups versus semigroups. Further, when compared with avoiding the use of some lemma, more interest may rest with the total avoidance of some axiom; after all, axioms are not typically thought of as lemmas.

The axiom in focus need not be a dependent axiom. For example, we might begin with a three-axiom system consisting of independent axioms and seek proofs in which one of the three is selected to be avoided. If we find such proofs, for each of the corresponding theorems, we say we have established proof dependence, because we have shown the selected item to be unnecessary. (The term *proof dependent* is intended to suggest to one that its establishment for a particular formula or equation depends on finding an appropriate proof.) When the selected item is an independent axiom and we are, nevertheless, able to prove one significant theorem after another without its participation (at the deduced level)—proof dependence is present—then we might be in the presence of a weaker theory that merits study. For a well-known situation, one need only consider group theory and its weakening to that of the theory of semigroups, where certain group-theory axioms are dropped. In this article, by offering proofs that totally avoid the use of a key independent axiom ($A7$), we offer a theory (that might merit study) weaker than the *BCSK* logic. For extensions of that logic, we found and offer proofs in which $A7$ is totally avoided, as well as a proof of its dependence that was indeed unexpected. These proofs provide powerful evidence that automated reasoning has arrived and that one with little or no knowledge can find treasure.

A second example, relevant to proof dependency, nicely illustrates one of the limiting points. Let us consider a logic in which condensed detachment is the only rule of inference such that the logic is studied in terms of a single axiom A. Let F denote the formula obtained by applying condensed detachment to two copies of A. Every proof of length greater than or equal to 1 in this study must have as its first step F. In other words for proofs of nonzero length, one cannot dispense with F. Therefore, F is never proof dependent (because it is always needed), regardless of the theorem under consideration when its proof requires at least one deduced step for its completion.

Hilbert himself might have been interested in proof dependence. Indeed, many of us learned as students of the famous 1900 lecture by Hilbert in Paris, a talk in which he offered twenty-three problems for study. As it turns out, a twenty-fourth problem exists, one he said in his notes that he did not have time to adequately formulate for the Paris lecture. (For that find, thanks goes to R. Thiele [TW02] and his thorough examination of Hilbert's notebooks.) That problem focuses on finding simpler proofs. A proof can be simplified in many ways, including shortening, removing a messy formula or equation, or avoiding some type of term. Also, pertinent to this article, a proof can be simplified by avoiding in all senses some axiom; if said axiom is independent, so much the better and more intriguing. We have a 30-step proof for the dependence of the fifth of the five Łukasiewicz axioms for his infinite-valued sentential calculus that is simpler than the original Meredith proof in various ways. For example, it is shorter, and, surprising to many, it avoids the use of any *double-negation terms*, terms of the form $n(n(t))$ for some term t. (The book citeWos2003b, offers the 30-step proof and many others of its type and features in detail various refinement methodologies. The book also offers open questions and challenges, in Chapter 7, that readers may find interesting.)

In the spirit of Hilbert's twenty-fourth problem is finding a proof that relies on fewer axioms than that in hand. If, for example, one has a proof P of a theorem T that relies on a set of axioms that include dependent axioms, and if one removes the dependent axioms to produce a set S of independent axioms, then there must exist a proof Q from S of the theorem T. The proof Q is simpler than the proof P in an axiomatic sense.

Of course, the absence of a dependent axiom in a set of hypotheses says nothing about its absence in a proof obtained from the so-called smaller (independent) axiom set. In this article, offered is methodology for finding proofs in which the dependent axioms not only are avoided as hypotheses but also are absent among the deduced steps of the proof, are, therefore, proof dependent. The inference rule or rules being employed are taken into account. This article details how such proof-dependent items were found with substantial aid from OTTER. The method that is given is extended to finding proof-dependent items even when the axiom in focus is in fact independent. When one finds that an independent axiom is proof dependent for a number of interesting theorems, then (as noted) the variety obtained by its omission may merit serious study. Proofs that avoid reliance on independent axioms might indeed have been of interest to Hilbert for they are simpler in an important way.

Finally, good fortune, occurring during the search (in the spirit of the new Hilbert problem) for short proofs, takes center stage as the experiments are discussed that led to the discovery of certain dependencies among the original nine axioms that in turn provided the wellspring for the study reported here. Here the article offers proofs that establish proof dependence for various axioms of the *BCSK* logic, as well as for some of its extensions. These proofs support the position that a weaker logic, obtained by the omission of the axiom $A7$, might offer unexpected power and interest.

2 A Wellspring for Ideas

The entire article came into being because of an attempt to find pleasing proofs for the following three theses (theorems), each given in its negated form.

```
-P(i(i(A,B),j(A,B))) | $ANS(THESIS_1).
-P(j(i(A,B),i(j(B,C),j(A,C)))) | $ANS(THESIS_2).
-P(j(i(B,C),i(j(A,B),j(A,C)))) | $ANS(THESIS_3).
```

The study was based on the nine axioms (for the *BCSK* logic that were given in Section 1) as hypotheses. (One might find interesting the fact that the first three axioms serve well for a complete system for the implicational fragment of intuitionistic logic.) Rather than seeking first proofs—three were in hand from Veroff—the object was to find shorter proofs, perhaps far shorter. Because of the nature of the first thesis, namely, it plays the role of a key lemma in a paper under consideration [BSV], the goal was to prove it by itself. The other main goal was to find a short proof of the join of the other two theses, 2 and 3.

Of the aspects of the approach to proof refinement with respect to length, two were prominent. First, we used ancestor subsumption, which causes the program to compare pairs of paths to a conclusion, preferring the shorter derivation. Second, we used demodulation to block steps of a proof, one at a time, to prevent their participation. (Demodulation is typically used for simplification and canonicalization.) By blocking the use of some given step, the program is forced to seek other paths to a proof—and it often is a most effective move to make when seeking shorter proofs.

3 Consequences of the Refinement Phase

As we made progress in finding ever shorter proofs, one of them was of particular note. Specifically, it failed to rely on $A3$ as a hypotheses. In other words, we had a proof from a smaller set of axioms, a set in which $A3$ was omitted.

The next move was to remove $A3$ from the axioms (by commenting it out in the initial set of support list). Somewhat later, we had in hand an even shorter proof, one with a most unexpected property. This proof relied on $A3$—as a deduced step. OTTER thus had established $A3$ to be dependent on the remaining eight of the nine original axioms. Now one sees why, near the close of Section 1, a reference was made to the discovery by good fortune of axiom dependencies.

Because we had found a satisfying proof in which $A3$ was not relied upon as an axiom, but was relied upon as a deduced step, we decided to seek a proof that totally avoided its presence—and the concept of *proof dependent* was born. The approach chosen, which succeeded, was to block, by demodulating unwanted new conclusions to "junk", the retention of $A3$ when and if it was deduced. OTTER later found a 14-step proof (which we give) of the dependency of $A3$, a proof relying on but six of the nine original axioms, omitting $A3$ (of course) but also omitting $A6$ and $A7$.

Stimulated by the discovery of a dependency within the original set of nine axioms, we sought to find other dependencies, focusing on $A6$ perhaps because of its position within the axiom set. In particular, $A3$ is the third of the given axioms concerned exclusively with the function i, and $A6$ is the third of those concerned almost exclusively with the function j. Again, our approach was to comment out $A6$ in the input, and we found appropriate proofs. We thus knew that $A6$ was not needed, at the axiomatic level, to find proofs of the three given theses. Eventually, we had a nice proof of the first of the three theses in which neither $A3$ nor $A6$ was used as an axiom. In that proof $A3$ was not present as a deduced step, but $A6$ was.

We paused before resuming the main journey to seek a nice proof of the dependence of $A6$ on seven of the nine original axioms, with $A3$ not participating. OTTER found one, a proof of length 27 (not given here) relying (as was the case for $A3$) on but six of the nine original axioms, omitting totally the use of $A3$, $A6$, and $A7$.

We therefore resumed the main journey, seeking a proof in which $A6$ was totally avoided, again relying on demodulating unwanted formulas to junk. The various attempts failed, which (in effect) brings us to the methodology that was promised for this article.

To put all in perspective, a review is in order. OTTER had succeeded in completing satisfying short proofs of the dependency of both $A3$ and $A6$ on the remaining seven axioms of the nine that prompted the study. We had a proof in which $A3$ participated in no way, $A6$ was not relied upon as an axiom, but $A6$ was present as a deduced step. Further, all attempts at completing a proof in which $A6$ was totally absent and all of the other given conditions were met had failed—with the numerous standard approaches we take.

We were thus forced to depart from our usual practice, that of paying little or no attention (in the vast majority of studies) to the actual proofs themselves. More precisely, our typical approach does not call for a close examination of a completed proof, in detail or as a whole. Instead, we rely on years of experimentation for a feel for which options and which values, if assigned to parameters, are likely to enable the program to complete a given assignment. In other words, we have found that the reading of a proof usually sheds little or no light on how one might proceed to refine it. Instead, such a reading can play a role in the formulation of new strategies and new methodologies that apply to many areas.

The so-to-speak forced inspection of the proof in hand that was the focus of attention showed that $A6$ was used as a parent for only one formula that followed its derivation. In that none of the standard approaches had enabled OTTER to find the sought-after proof, the obvious conjecture asserts that a number of steps greater than 1 might be needed to obtain the child of $A6$, where the formula $A6$ was not allowed to participate. Indeed, intuitively, removing one of the two parents of a deduced conclusion, especially when the removed parent is itself a deduced conclusion late in a proof—in the case under discussion, the 46th step in a 53-step proof—can cause havoc. Our choice was to invoke the use of the command set(sos_queue) to cause OTTER to conduct a breadth-first search for a proof of the child of $A6$. We placed in the initial set of support (in addition to the axioms $A1$, $A2$ $A4$, $A5$, $A7$, $A8$, and $A9$) the first 45 deduced steps of the proof in hand up to but not including $A6$. The target, negated and placed in list(passive), was the child of $A6$. (The approach we took is indeed reminiscent of the cramming strategy [Wos03], a strategy that enables the program to force or cram formulas in the initial set of support into the desired proof.) Just for total clarity, with almost certainty, additional deduced steps would be needed. After a thorough level-saturation search through level 1, at level 2 the desired proof of the child of $A6$ was completed, a proof of length 2.

We pause briefly to note that the approach just given would have merited use even if $A6$ had been the parent of more than one formula that followed its derivation. Iteration would be the way to proceed. One would proceed as we did but now with the negation of the first child of $A6$ placed in list(passive) with the goal of obtaining the needed proof that culminates with the derivation of the first child and without allowing $A6$ to participate in any manner. Then one would amend further the list(sos) with the new proof steps (that led to the derivation of the first child of $A6$ without $A6$ participating), as well as proof steps of the original proof preceding the second child and not dependent on $A6$, and used as target the second child, placing its negation in list(passive), now with the goal of deriving the second child and with the given constraints. One would proceed in this manner, gathering proof steps along the way, until the last child of $A6$ was proved. Of course, the method we are presenting is useful when the goal is to avoid any unwanted formula or equation and replace its role by other formulas or equations, whether establishing proof dependence is the intention or not.

We had the components that almost guaranteed we could complete a proof that avoided the use of $A3$ and $A6$ as axioms and, more important in the context of proof dependence, avoided the use of those two formulas even as deduced steps. To enable OTTER to return the proof of interest, we placed in the initial set of support the original nine axioms but with $A3$ and $A6$ commented out. In list(passive), we still placed the negation of thesis_1 and, for monitoring purposes, the negations, respectively, of thesis_2 and thesis_3. In list(usable), we placed the two rules for condensed detachment, one for the function i and one for the function j, and the negation of the join of theses 2 and 3. To ensure that both $A3$ and $A6$ would not participate in any proof, we included the following.

```
list(demodulators).
(P(i(i(i(x,y),x),x)) = junk).            %  A3
(P(i(j(x,j(y,z)),j(y,j(x,z)))) = junk).  %  A6
(i(x,junk) = junk).
(i(junk,x) = junk).
(j(x,junk) = junk).
(j(junk,x) = junk).
(P(junk) = $T).
end_of_list.
```

The crucial move directed OTTER to the proof we expected it to find, a proof quite like that which relied on both $A6$ as a deduced clause and exactly one of its children. Throughout the experiments, we had relied upon the use of *resonators* to direct the program's search for one or more proofs. A resonator [Wos95] is a formula or an equation that does not itself take on the value **true** or **false**. Instead, its functional pattern is the key, where all variables within a resonator are treated as indistinguishable from each other, just denoting that a variable occurs in the corresponding position, and where the value assigned to a resonator reflects its conjecture importance (the smaller the value, the higher the priority given to similar deduced items). For the resonators intended to guide the program to the expected goal, we used the set that had led to the proof relying on $A6$, and, to enable the program to find the newer proof (not depending on $A6$) of the child of $A6$, we included the two resonators that corresponded to the proof found with level saturation.

As expected, OTTER was successful, and we had established both $A3$ and $A6$ proof dependent and, of course, not relying on either at the axiomatic level. We immediately attempted to further prune the original nine with regard to axiom dependencies and, more relevant to this article, seek proofs establishing additional proof dependence than that in hand. We did not expect that $A1$, $A2$, $A4$, or $A5$ would extend what we had in hand so far, in part because they appeared to be vital. However, $A7$ did look promising.

Therefore, we began a study with $A7$ commented out, as well as $A3$ and $A6$ not accessible as axioms or as deduced formulas.

The capture was quickly made: we had proofs in which neither $A3$ nor $A6$ nor $A7$ was present, as an axiom or as a deduced formula. We therefore turned to an attempt to prove $A7$ dependent on but six of the nine axioms. The experiment failed. Z. Ernst came to the rescue—or perhaps rescue is the wrong word in that we would have preferred $A7$ to be dependent—finding the following three-element model (with Mace4; see the Web www.mcs.anl.gov/AR/mace4) showing that $A7$ is in fact independent of the six.

```
-------- Model 1 at 0.01 seconds --------

 a : 1

 b : 2

 i :
      | 0 1 2
    --+------
    0 | 0 1 2
    1 | 0 0 2
    2 | 0 1 0

 j :
      | 0 1 2
    --+------
    0 | 0 1 2
    1 | 0 0 2
    2 | 0 0 0

 P :
        0 1 2
    ---------
        1 0 0

-------- end of model --------
```

Nevertheless, we now had in hand an example of an extension of the original concept of proof dependence in that we had considered an independent axiom. Specifically, although $A7$ is in fact independent, we had in hand proofs establishing each of $A3$, $A6$, and $A7$ to be proof dependent, with $A3$, $A6$, and $A7$ absent from the axiom system. We were thus ready for a serious effort at proof refinement in the context of length, within the given constraints, seeking "short" proofs of thesis_1, the join of theses 2 and 3, the dependence of $A3$, and the dependence of $A6$.

4 Pleasing Proofs

Our main effort in the context of proof refinement was aimed at proof shortening; shorter proofs are usually more pleasing than longer. As noted earlier, of the various aspects that can be brought to bear, two played the key role: ancestor subsumption and demodulation (to block the retention of conclusions one classes as unwanted). In particular, we took each proof in hand and instructed OTTER to block its steps one at a time, forcing it to seek a somewhat different proof, occasionally a sharply different proof. One might immediately conjecture that a direct attack on finding shorter proofs is in order, some type of exhaustive search, for example. The task cannot be so subtle, or can it?

Of course, one would prefer applying an algorithm that simply seeks and finds the shortest proof that exists for any given theorem and given axiom set. Studies of more than a decade prove (to me) that such an algorithm in many, many cases does not exist. Further, an unexpected obstacle (illustrating the cited subtlety) exists when seeking a proof shorter than that in hand. The following aphorism (found in some books and papers) nicely captures the obstacle. "Shorter subproofs do not necessarily a shorter total proof make." For the curious, how can this aphorism hold? An example is in order.

Let us consider a proof of, say, 20 steps in which the tenth step is proved by using steps 6 through 9. In other words, the length of the subproof concluding with step 10 is five. Now let us assume that OTTER or some person finds a proof of 10 that relies on 6a and 8a, a proof of length three. With ancestor subsumption in use, the program will prefer this second derivation because its length is three rather than five. A program or a person might then attempt to complete a proof relying on the three-step shorter subproof, with the expectation that the total proof (of step 20) will clearly be shorter. Such may not occur, for example, in the event that steps 6 through 9 play a vital role in the twenty-step proof. If a proof is completed that uses the cited three steps (of the shorter subproof), the resulting proof may have length at least 22—and far worse may occur.

Our efforts were indeed rewarded, as seen with the following proofs.

The given proofs are the shortest, for their respective conclusions, we have been able to complete. (For the curious, we note that the inference rule regarding the function i can be dispensed with; it is a derived inference rule. Its inclusion enables the program to find shorter proofs. In the presence of *A1,* *A*2, *A*4, *A*5, *A*8, and *A*9, OTTER finds a two-step proof showing that the corresponding clause is dependent. With the cited axiom system and the derived inference rule present, OTTER finds a 14-step proof showing *A*3 to be dependent; when the derived inference rule is removed, the best proof we have found has length 20.)

A 14-Step Proof of the Dependency of A3

```
----- Otter 3.3d, April 2004 -----
The process was started by wos on jaguar.mcs.anl.gov,
Thu May 27 10:43:03 2004
The command was "otter".  The process ID is 31886.
----> UNIT CONFLICT at   0.02 sec ----> 150 [binary,149.1,17.1] $ANS(a3).
```

```
Length of proof is 14.  Level of proof is 10.

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

6 [] -P(i(x,y)) | -P(x) | P(y).
7 [] -P(j(x,y)) | -P(x) | P(y).
9 [] P(i(x,i(y,x))).
10 [] P(i(i(x,i(y,z)),i(i(x,y),i(x,z)))).
11 [] P(i(x,j(y,x))).
12 [] P(i(j(x,j(y,z)),j(j(x,y),j(x,z)))).
13 [] P(i(j(i(x,y),y),j(i(y,x),x))).
14 [] P(j(i(x,y),j(x,y))).
17 [] -P(i(i(i(a1,a2),a1),a1)) | $ANS(a3).
24 [hyper,6,9,9] P(i(x,i(y,i(z,y)))).
38 [hyper,6,10,10] P(i(i(i(x,i(y,z)),i(x,y)),i(i(x,i(y,z)),i(x,z)))).
40 [hyper,6,10,24] P(i(i(x,y),i(x,i(z,y)))).
41 [hyper,6,10,9] P(i(i(x,y),i(x,x))).
46 [hyper,7,14,40] P(j(i(x,y),i(x,i(z,y)))).
58 [hyper,6,38,41] P(i(i(x,i(x,y)),i(x,y))).
97 [hyper,7,14,58] P(j(i(x,i(x,y)),i(x,y))).
115 [hyper,6,13,97] P(j(i(i(x,y),x),x)).
120 [hyper,6,11,115] P(j(x,j(i(i(y,z),y),y))).
124 [hyper,6,12,120] P(j(j(x,i(i(y,z),y)),j(x,y))).
129 [hyper,7,124,46] P(j(i(i(i(x,y),z),y),i(x,y))).
138 [hyper,7,124,129] P(j(i(i(i(i(x,y),x),z),x),x)).
144 [hyper,6,13,138] P(j(i(x,i(i(i(x,y),x),z)),i(i(i(x,y),x),z))).
149 [hyper,7,144,9] P(i(i(i(x,y),x),x)).
```

5 Extending the Logic

At this point, one might ask about the power of the abbreviated axiom set consisting of $A1$, $A2$, $A4$, $A5$, $A8$, and $A9$. For example, with $A3$ and $A6$ and $A7$ omitted, can we prove significant theorems when the logic is extended by adjoining yet another axiom of interest or by adjoining a set of axioms focusing on different functions? Obviously, the omission of both $A3$ and $A6$ presents no problem at the axiomatic level in that they have been proved dependent on the set consisting of $A1$, $A2$, $A4$, $A5$, $A8$, and $A9$. However, as noted, $A7$ is independent of that set. Further, perhaps $A3$ or $A6$ or both will be needed at the deduced level, and $A7$ will be needed at the axiomatic level or at the deduced level.

With the following formula, $A10$, we have such an extended logic, *BCSK+*.

```
P(i(j(j(x,y),y),j(j(y,x),x))).  % A10
```

An interesting theorem to prove is captured, in its negated form, with the following clause; the formula to be proved is equivalent to $A10$, and the proof found by OTTER avoids totally $A3$, $A6$, and $A7$.

```
-P(i(j(A,B),i(A,B))) | $ANS(thm).
```

To complete proof of the equivalence, OTTER found an 18-step proof that deduces $A10$ from the following formula in clause notation, and, again, a proof completely free of reliance on $A3$, $A6$, and $A7$.

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

An appropriate move to test the power of the abbreviated axiom system, now consisting of seven axioms with the cited addition of $A10$, is to give OTTER an input file whose initial set of support consists of the seven axioms. The demodulator list contains equalities that, respectively, block the retention of $A3$, $A6$, and $A7$ if and when each is deduced. After all, for example, $A7$ might now be dependent on the seven-axiom system. From Veroff, we had in hand a 42-step proof of the theorem under consideration to initiate the study, a proof that does depend on the three axioms we intended to avoid, (at both the axiomatic and deduced levels). The original goal was to shorten that proof. More pertinent from the viewpoint of this article, we sought to find a proof establishing each of $A3$, $A6$, and $A7$ to be proof dependent simultaneously.

All went smoothly, with the discovery of a 23-step proof. In examining the proof, we observed that the added axiom, $A10$, is used but once. This observation caused us to ask about the independence of $A7$ in this extended logic. After all, perhaps the use of the added axiom ($A10$) leads to a proof of the dependence of $A7$. Therefore, we turned after a short time to studying this possible dependence. Is $A7$ independent or dependent in the extended logic? The effort paid off: OTTER found a proof of dependence, the following in which both $A3$ and $A6$ are totally absent.

A 24-Step Proof of the Dependence of A7

```
----- Otter 3.3g-work, Jan 2005 -----
The process was started by wos on jaguar.mcs.anl.gov,
Tue Mar  8 16:10:03 2005
The command was "otter".  The process ID is 4337.
----> UNIT CONFLICT at   0.13 sec ----> 679 [binary,678.1,17.1] $ANS(a7).

Length of proof is 24.  Level of proof is 11.

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

1 [] -P(i(x,y)) | -P(x) | P(y).
2 [] -P(j(x,y)) | -P(x) | P(y).
5 [] P(i(x,i(y,x))).
6 [] P(i(i(x,i(y,z)),i(i(x,y),i(x,z)))).
7 [] P(i(x,j(y,x))).
8 [] P(i(j(x,j(y,z)),j(j(x,y),j(x,z)))).
9 [] P(i(j(i(x,y),y),j(i(y,x),x))).
10 [] P(j(i(x,y),j(x,y))).
11 [] P(i(j(j(x,y),y),j(j(y,x),x))).
17 [] -P(i(j(j(a1,a2),a1),a1)) | $ANS(a7).
```

```
57 [hyper,1,6,6] P(i(i(i(x,i(y,z)),i(x,y)),i(i(x,i(y,z)),i(x,z)))).
58 [hyper,1,5,6] P(i(x,i(i(y,i(z,u)),i(i(y,z),i(y,u))))).
59 [hyper,1,6,5] P(i(i(x,y),i(x,x))).
61 [hyper,1,5,7] P(i(x,i(y,j(z,y)))).
69 [hyper,2,10,5] P(j(x,i(y,x))).
80 [hyper,1,6,58] P(i(i(x,i(y,i(z,u))),i(x,i(i(y,z),i(y,u))))).
83 [hyper,1,57,59] P(i(i(x,i(x,y)),i(x,y))).
93 [hyper,1,7,69] P(j(x,j(y,i(z,y)))).
124 [hyper,1,80,5] P(i(i(x,y),i(i(z,x),i(z,y)))).
127 [hyper,2,10,83] P(j(i(x,i(x,y)),i(x,y))).
136 [hyper,1,8,93] P(j(j(x,y),j(x,i(z,y)))).
188 [hyper,1,6,124] P(i(i(i(x,y),i(z,x)),i(i(x,y),i(z,y)))).
198 [hyper,1,9,127] P(j(i(i(x,y),x),x)).
288 [hyper,1,188,61] P(i(i(j(x,y),z),i(y,z))).
344 [hyper,1,288,11] P(i(x,j(j(x,y),y))).
395 [hyper,2,10,344] P(j(x,j(j(x,y),y))).
398 [hyper,1,124,344] P(i(i(x,y),i(x,j(j(y,z),z)))).
550 [hyper,1,8,395] P(j(j(x,j(x,y)),j(x,y))).
564 [hyper,1,398,288] P(i(i(j(x,y),z),j(j(i(y,z),u),u))).
615 [hyper,1,11,550] P(j(j(j(x,y),x),x)).
618 [hyper,2,198,564] P(j(j(i(x,y),x),x)).
632 [hyper,2,136,615] P(j(j(j(x,y),x),i(z,x))).
650 [hyper,1,11,618] P(j(j(x,i(x,y)),i(x,y))).
678 [hyper,2,650,632] P(i(j(j(x,y),x),x)).
```

A second and more intriguing extension of the original logic, *SBPC*, was studied with the goal of determining the need, at the deduced level, of $A3$, $A6$, and $A7$. For the study, we began again with the now so-to-speak famous six axiom system, that consisting of $A1$, $A2$, $A4$, $A5$, $A8$, and $A9$, and adjoined the following six axioms (expressed in clause notation), where the function a denotes logical **and** and the function o denotes logical **or**.

```
P(j(x,o(x,y))).                           %  A11
P(i(y,o(x,y))).                           %  A12
P(j(j(x,z),j(j(y,z),j(o(x,y),z)))).       %  A13
P(i(a(x,y),x)).                           %  A14
P(j(a(x,y),y)).                           %  A15
P(i(i(x,y),i(i(x,z),i(x,a(y,z))))).       %  A16
```

In this extended logic, we attempted to find proofs, preferably short ones, of the following four theorems, each given in its negated form, and, as one might predict, we sought proofs in which $A3$, $A6$, and $A7$ are totally absent.

```
-P(j(i(A,B),i(o(A,C),o(B,C)))) | $ANS(1).
-P(j(i(A,B),i(o(C,A),o(C,B)))) | $ANS(2).
-P(j(i(A,B),j(i(B,A),i(a(A,C),a(B,C))))) | $ANS(3).
-P(j(i(A,B),i(a(C,A),a(C,B)))) | $ANS(4).
```

We began the study of the four theorems with proofs supplied by Veroff, obtained by him using his powerful technique called *sketches*. Perhaps because of the goal of finding appropriate proofs, four of them, establishing proof dependence for the three unwanted axioms, we were unable to complete the studies until we relied on a 92-step proof that deduced (without using $A3$ in any way) a (former) child of $A3$. In other words, the earlier studies of proof dependence came into play, enabling us (and OTTER) to overcome an obstacle. Success eventually was the result. OTTER returned a 53-step proof of the first of the four theorems, a 64-step proof of the second, a 104-step proof of the third and a 99-step proof of the fourth. Many experiments were required, as well as much use of refinement methodology detailed in the book [WP03]. The last significant reductions in proof length (of the proofs of the third and fourth theorems) were obtained by heavy reliance on cramming. Briefly, OTTER was given proofs of steps near the end of the proofs in hand and asked to (in effect) force their proof steps into (we hoped) shorter proofs of the targets.

The discovery that $A7$ is dependent in the *BCSK*+ logic (obtained by adding $A10$ to the original nine axioms, then removing any use of $A3$ and $A6$) led us to consider the possibility that that formula is dependent in this second extension of the *BCSK* logic. Indeed, would it not be more than piquant to find that $A7$ is independent in the original study and then find it dependent in two extensions of the logic? And, as the following proof shows—the shortest so far discovered—that is exactly what was found.

A 35-Step Proof of the Dependence of A7 in a Second Extension

```
----- Otter 3.3g-work, Jan 2005 -----
The process was started by wos on theorem.mcs.anl.gov,
Sun Mar 20 12:31:56 2005
The command was "otter".  The process ID is 20352.
----> UNIT CONFLICT at   0.09 sec ----> 904 [binary,903.1,24.1] $ANS(a7).

Length of proof is 35.  Level of proof is 20.

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

10 [] -P(i(x,y)) | -P(x) | P(y).
11 [] -P(j(x,y)) | -P(x) | P(y).
12 [] P(i(x,i(y,x))).
13 [] P(i(i(x,i(y,z)),i(i(x,y),i(x,z)))).
14 [] P(i(x,j(y,x))).
15 [] P(i(j(x,j(y,z)),j(j(x,y),j(x,z)))).
16 [] P(i(j(i(x,y),y),j(i(y,x),x))).
17 [] P(j(i(x,y),j(x,y))).
18 [] P(j(x,o(x,y))).
19 [] P(i(y,o(x,y))).
20 [] P(j(j(x,z),j(j(y,z),j(o(x,y),z)))).
24 [] -P(i(j(j(a1,a2),a1),a1)) | $ANS(a7).
130 [hyper,10,13,13] P(i(i(i(x,i(y,z)),i(x,y)),i(i(x,i(y,z)),i(x,z)))).
```

```
133 [hyper,10,12,14] P(i(x,i(y,j(z,y)))).
135 [hyper,10,12,15] P(i(x,i(j(y,j(z,u)),j(j(y,z),j(y,u))))).
138 [hyper,11,17,16] P(j(j(i(x,y),y),j(i(y,x),x))).
139 [hyper,11,17,15] P(j(j(x,j(y,z)),j(j(x,y),j(x,z)))).
140 [hyper,11,17,14] P(j(x,j(y,x))).
142 [hyper,11,17,12] P(j(x,i(y,x))).
180 [hyper,10,130,133] P(i(i(x,i(j(y,x),z)),i(x,z))).
197 [hyper,11,140,140] P(j(x,j(y,j(z,y)))).
244 [hyper,10,180,135] P(i(j(x,y),j(j(z,x),j(z,y)))).
285 [hyper,11,17,244] P(j(j(x,y),j(j(z,x),j(z,y)))).
290 [hyper,10,244,142] P(j(j(x,y),j(x,i(z,y)))).
298 [hyper,10,15,285] P(j(j(j(x,y),j(z,x)),j(j(x,y),j(z,y)))).
347 [hyper,11,298,197] P(j(j(j(x,y),z),j(y,z))).
362 [hyper,11,285,347] P(j(j(x,j(j(y,z),u)),j(x,j(z,u)))).
371 [hyper,11,347,138] P(j(x,j(i(x,y),y))).
416 [hyper,11,362,139] P(j(j(x,j(y,z)),j(y,j(x,z)))).
449 [hyper,11,285,371] P(j(j(x,y),j(x,j(i(y,z),z)))).
488 [hyper,11,416,416] P(j(x,j(j(y,j(x,z)),j(y,z)))).
506 [hyper,11,416,138] P(j(i(x,y),j(j(i(y,x),x),y))).
559 [hyper,11,449,18] P(j(x,j(i(o(x,y),z),z))).
584 [hyper,11,139,488] P(j(j(x,j(y,j(x,z))),j(x,j(y,z)))).
603 [hyper,11,506,19] P(j(j(i(o(x,y),y),y),o(x,y))).
604 [hyper,11,506,14] P(j(j(i(j(x,y),y),y),j(x,y))).
657 [hyper,11,584,559] P(j(x,j(i(o(x,y),j(x,z)),z))).
681 [hyper,11,416,657] P(j(i(o(x,y),j(x,z)),j(x,z))).
698 [hyper,11,603,681] P(o(x,j(x,y))).
709 [hyper,11,488,698] P(j(j(x,j(o(y,j(y,z)),u)),j(x,u))).
727 [hyper,11,709,140] P(j(x,x)).
738 [hyper,11,20,727] P(j(j(x,y),j(o(y,x),y))).
767 [hyper,11,709,738] P(j(j(j(x,y),x),x)).
814 [hyper,11,285,767] P(j(j(x,j(j(y,z),y)),j(x,y))).
845 [hyper,11,814,604] P(j(j(i(j(j(x,y),x),x),x),x)).
851 [hyper,11,814,290] P(j(j(j(i(x,y),z),y),i(x,y))).
903 [hyper,11,851,845] P(i(j(j(x,y),x),x)).
```

For the curious, the first study yielded a 39-step proof, a proof that the usual methods were unable to improve upon. However, with a most unsophisticated form of cramming, the given 35-step proof was found. In particular, rather than relying on a subproof of one of the late steps, OTTER was merely given the first 34 steps of the 39-step proof and told to apply level saturation. In other words, no attention was paid to the possible presence of steps among the thirty-four that were not used in the proof of the thirty-fourth step.

6 Summary

In this article, we have extended the notion of axiom dependence to one of *proof dependence.* Briefly, a formula or equation is proof dependent if it can be dispensed with, even as a deduced item; in other words there exists at least one proof that shows the item to be totally unnecessary. The new term, proof dependent, was chosen because of the nature of a dependent axiom, namely, one that is unnecessary at the so-called input level (from the viewpoint of automated reasoning). We have given methodology for finding an appropriate proof, one that completely avoids the use of some selected item, even when the item is in fact an independent axiom.

We have included various proofs discovered with indispensable aid from McCune's OTTER, the shortest proofs that we could discover, given the conditions to be satisfied. Such conditions included total avoidance of one or more items. Among our successes was the discovery of various axiom dependencies. In two of the three logics we studied, both extensions of the *BCSK* logic, we found (most unexpectedly) that a key axiom, $A7$, is dependent, although it is independent among the axioms for *BCSK.*

Acknowledgments

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

References

[Bel01] J. Belinfante. Computer assisted proofs in set theory, 2001. Web site http://www.math.gatech.edu/∼belinfan/research/autoreas/index.html.

[BP94] W. J. Blok and D. Pigozzi. On the structure of varieties with equationally definable principal congruences III. *Algebra Universalis*, 32:545–608, 1994.

[BP99] W. J. Blok and D. Pigozzi. Algebraisable logics. *Mem. Amer. Math. Soc.*, 77, 1999.

[BSV] R. J. Bignall, M. Spinks, and R. Veroff. On the assertional logics of the generic pointed discriminator and generic pointed fixedpoint discriminator varieties. Preprint, 2003.

[Kun92] K. Kunen. Single axioms for groups. *J. Automated Reasoning*, 9:291–308, 1992.

[Phi] J. D. Phillips. Private Communication, Argonne Workshop on Automated Reasoning and Deduction (AWARD), 2003.

[TW02] R. Thiele and L. Wos. Hilbert's twenty-fourth problem. *J. Automated Reasoning*, 29(1):67–89, 2002.

[Ver96] R. Veroff. Using hints to increase the effectiveness of an automated reasoning program: Case studies. *J. Automated Reasoning*, 16(3):223–239, 1996.

[Ver01] R. Veroff. Solving open questions and other challenge problems using proof sketches. *J. Automated Reasoning*, 27(2):175–199, 2001.

[Wos95] L. Wos. The resonance strategy. *Computers and Mathematics with Applications*, 29(2):133–178, 1995.

[Wos03] L. Wos. The strategy of cramming. *J. Automated Reasoning*, 30(2):179–204, 2003.

[WP03] L. Wos and G. W. Pieper. *Automated Reasoning and the Discovery of Missing and Elegant Proofs*. Rinton Press, Paramus, N.J., 2003.

[WRC65] L. Wos, G. Robinson, and D. Carson. Efficiency and completeness of the set of support strategy in theorem proving. *J. ACM*, 12:536–541, 1965.