papers/cramming/paper
created 05-10-2001
revised last 05-29-2001

delim @@

# The Strategy of Cramming*

*Larry Wos*

Mathematics and Computer Science Division
Argonne National Laboratory
Argonne, IL 60439-4801

**Abstract**

The problem in focus in this article concerns proof shortening, and featured prominently is a strategy, *cramming*, for addressing that problem. The literature shows that this problem was indeed of interest to some of the masters of logic, masters that included, c. A. Meredith, A. Prior, and I. Thomas. Until very recently, unknown to those masters and to mathematicians and logicians as a group was the interest in proof simplification by the greatest mathematician of the twentieth century, Hilbert. Hilbert expressed his interest by posing a twenty-fourth problem, found by R. Thiele in his examination of the Hilbert files in Goettingen. Yes, Hilbert's problems are not limited to twenty-three as was believed for almost a century. One aspect of proof simplification concerns proof length, proof shortening or refinement. The cramming strategy (introduced in this article) was formulated to seek shorter proofs, starting with some known proof, usually the shortest offered by the literature. The most impressive success with the use of this strategy concerns an abridgment of the Meredith-Prior abridging of the Lukasiewicz proof for his shortest single axiom for the implicational fragment of two-valued sentential (or propositional)l calculus.

## 1. Perspective and Wellspring

In the early 1990s, my colleague William McCune and I embarked on a study designed to produce a proof relying solely on condensed detachment showing that the fifth of Lukasiewicz's five axioms for infinite-valued sentential calculus is in fact dependent on the other four. C. A. Meredith had already proved the theorem, but not purely in terms of the cited inference rule; equality-oriented reasoning was involved. What we did not know at the time—and only learned here in 2001—was that, apparently, the literature offers no proof of the type we sought. Our effort was rewarded: Each independently found the desired proof. In other words, we unknowingly had found a proof absent from the literature, a Hilbert-style, axiomatic proof relying solely on condensed detachment. In retrospect, we would have been well-advised to attempt to add our success to the literature.

Each of us then turned to an attempt at finding a means for OTTER to produce a proof of the theorem under discussion, where no guidance was provided by the researcher. We each succeeded. A comparison of the two resulting proofs showed they were virtually identical, essentially proofs of length 63 (applications of condensed detachment. And McCune took one road while I took another.

Indeed, that success quickly led to my study aimed at finding a shorter proof. (The reason: It simply struck my fancy, perhaps influenced by my studies at the University of Chicago decades earlier.) If memory serves, at the time I had no methodology in hand and, certainly, no notion of where this road would take me almost a decade later. In particular, especially the late 1990s until now have

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

witnessed the formulation of various approaches to proof refinment with respect to length (as well as to other properties such as formula complexity).

One additional occurrence completes the perspective for the strategy introduced in this article. In the late 1990s, Branden Fitelson contacted me by e-mail in the context of his finding a new proof (with McCune's OTTER)for a result of Frege. And we have sought new proofs ever since. As part of our effort, Fitelson posed the problem of finding a proof better than that found and published by Meredith and Prior. They had found ''an abridgment'' of a proof that Lukasiewicz provided for his shortest single axiom for the implicational fragment of two-valued sentential (or propositional) calculus. In other words, Fitelson suggested I seek an abridgment of the Meredith-Prior abridgment—which was the wellspring for the formulation of the *cramming strategy* featured here.

Indeed, a single theorem was the force for the birth of this new strategy. As detailed later, the goal was reached: The desired abridgment was found by OTTER. As required, the offering of this strategy would be premature were it not for its successful use in other contexts, also to be detailed. Its existence has perhaps added import. Specifically, Rudiger Thiele's discovery that Hilbert in fact offered twenty-four problems—not twenty-three as was believed for almost a century. That twenty-fourth problem is concern with proof simplification. One obvious way to simplify a proof is to find a shorter proof, and (as will be discussed) the cramming strategy provides one avenue to pursue. (For those less familiar with the literature of mathematics, a proof can be shortened with the price of a sharp decrease in simplicity to the point of being obscure, either by omitting many steps, or by relying on far more complex formulas or equations. The latter is even true for proofs produced by an automated reasoning program. The rules of the game here are that all of the deduced steps are within the theory under consideration, one or more specific inference rules account for deductions, and steps are not omitted—as Hilbert might indeed find appropriate.)

## 2. The Nature and Object of Cramming

This new strategy derives its name from its nature. In effect, the strategy focuses on a chosen subproof of the total proof in hand of some conjunction and attempts to cram steps of that subproof into other subproofs needed to complete a new and shorter total proof. Because of the basic mechanism that is employed, level saturation, it seems almost certain that none of the masters applied such an approach, whereas cramming is indeed well suited to the actions of an automated reasoning program. Alternatively, with the emphasis on a chosen integer *j*, the cramming strategy derives its name from the attempt to find subproofs of members of a conjunction to be considered with a given subproof of one of the members such that their lengths can be so-to-speak crammed into *j*. Typically, *jR is chosen to be less than the length of the shortest proof known for the theorem under study.*

As My colleague Robert Veroff s nicely phrases the problem, the seeking of a shortest proof for a conjunction asks for the minimization of the set-theoretic union of the subproofs of its members. That problem offers unexpected obstacles, the main one captured by the following maxim. *Shorter Subproofs do not necessarily a shorter total proof make*. This maxim dispels a myth that could easily be believed: Simply find ever shorter subproofs of the members of a conjunction, and progress will occur when the goal is that of finding a shorter or shortest proof of the entire conjunction. A simple example is in order.

Since the plum of this article concerns the Lukasiewicz shortest single axiom for the implicational fragment of two-valued (classical) logic, it (expressed as a clause) will be the focus of the example.

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

The following Tarski-Bernays 3-basis provides a target for proving that the preceding formula is in fact a complete axiom system.

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

For pedagogical reasons, I shall begin with a synthetic bit of data, turning to actual data once the point has been made patently clear. First, note that the first step of any condensed-detachment proof relying on the Lukasiewicz formula is forced, must be present in any proof in which the formula serves as the only hypothesis. Next, assume that one has separately attacked the problem of finding a subproof for each of the three target axioms. Then, assume that the three attacks have been unusually successful, yielding, respectively, proofs of lengths 10, 20 and 30 for the first through the third target axioms. Finally, assume that, other than the first deduced step of each proof, the three proofs are disjoint. In the hypothetical example under discussion, the proof of the conjunction is 10 + 20 + 30 - 2 (duplicate first steps) = 58. At the other end of one spectrum is the (hypothetical) example in which one is able to find a 40-step proof of the third Tarski-Bernays axiom such that it contains a 15-step subproof of the first target axiom and a 25-step proof of the second. The shorter subproofs do indeed not a shorter total proof make.

For the promised actual data—and a glimpse of the end of the story focusing on the abridgment of the abridgment—Otter did find a 35-step proof of the desired conjunction. The respective subproofs (for the three target axioms) have in order length 10, 23, and 30> The better proof (given later) was discovered by OTTER's application of the cramming strategy. It has length 32, containing subproofs of respective length 10, 28, and 30.

**3. The First Success**

This section illustrates the first use of the cramming strategy and (as promised) presents a most satisfying proof obtained with it. A comparison of the new proof with the Meredith-Prior proof and with the Lukasiewicz proof provides further insight into the nature of proof refinement with respect to length.

My attempt to reach the objective (posed by Fitelson) of finding a proof shorter than the Meredith-Prior proof began with focusing on their proof and to a much lesser degree on the Lukasiewicz prof for his shortest single axiom. Their 33-step proof was used (with the resonance strategy) to direct OTTER's search, and, occasionally, the Lukasiewicz 34-step proof was also used in this manner. If one uses Veroff's ints strategy rather than the resonance strategy, the program has a strong tendency to simply reproduce the proof in hand. A partial explanation rests with the fact that the hints strategy focuses on the specific formulas or equations supplied and, if desired, clauses that subsume or are subsumed by a hint. In contrast, because a resonator is treated as if its variables are indistinguishable, guidance from the resonance strategy encourages a program to focus on equivalence classes (rather than on subsumption) defined by a resonator. In other words, the resonance strategy compared with the hints strategy gives the program a different type of latitude.

That phase of my research eventually yielded a 33-step proof of the Tarski-Bernays system, a proof a few steps different from the Meredith-Prior 33-step proof. Next, to each 33-step proof, I applied the methodology whose main component is that of blocking proof steps one at a time to force the program to yield (if possible) a different proof. Sometimes a given proof has the most unexpected property of containing in a not rigorous sense a proper subset that is in fact also a proof of the theorem under study. In particular, sometimes a proper subset of the deduced steps of the proof in hand is such that, with different history and a different ordering, it itself is a shorter proof of the theorem to be proved. Failure was the result: No 32-step proof, or shorter, was found.

However, a small diamond was mined. Specifically, the new 33-step proof contained a 30-step proof of the third Tarskki-Bernays axiom (TB-3), that axiom that (intuitively) is the most difficult to deduce. The 30-step proof did not contain as a subproof either a proof of TB-1 or TB-2. The Meredith-Prior 33-step proof proves TB-3 in 31 steps, and the Lukasiewicz 34-step proof proves TB-3 in 32 steps. In all three cases, the new 33-step proof, the old 33-step proof, the 34-step proof, the proof of TB-3 fails to include a subproof of either TB-1 or TB-2.

Perhaps—and I thought almost certainly not the case—a means could be devised to take the 30-step proof of TB-3 and *cram* many, many of its steps into proofs of TB-1 and TB-2. To be of interest, the cramming must have the property that the only needed steps to complete the desired subproofs (ot 1 and 2) would be, respectively, and and 2 themselves. Stated a slightly different way, the cramming must have the property that 30 steps (proving TB-3) and TB-1 and TB-2 must (in effect) form a 32-step proof. To be the case, the 30-step proof must be such that one pair of its formulas yields with condensed detachment TB-1 and another pair yields TB-2.

To test whether I had indeed struck gold, OTTER offers just what is needed. Indeed, one places the thirty formulas in the set of support and instructs the program to conduct a level-saturation search. If the hard-to-accept is in fact true, then two level-1 proofs will be found, respectively, of TB-1 and TB-2. Success: Condensed detachment applied to the ninth step of the 30-step proof with itself yields TB-1; TB-2 is obtained by applying condensed detachment to the twenty-ninth and twenty-seventh steps. OTTER had indeed discovered a proof shorter than the Meredith-Prior abridgment, a proof of length 32.

Although that proof could be presented by appending to the 30-step proof the two cited applications of condensed detachment, more in the spirit of automated reasoning and full automation, perhaps an approach could be devised that would yield the 32-step proof directly. More to the pointd for use in other applications—perhaps a strategy could be formulated that would yield the 32-step proof from the 30-step proof, would appropriately cram in a manner that sufficed.

One obvious way to proceed calls for the inclusion of thirty-two resonators corresponding to the desired 32-step proof, each assigned a small value, say, 2. Then one assigns a small value to max_weight, not much larger than that assigned to the resonators. The clear intent is to force the program to focus on the thirty-two steps and not much else. If the assigned value to max_weight is 4, OTTER finds a 35-step proof; if the value is 2, OTTER finds a 33-step proof. Conclusion: the resonance strategy permits more than is desired. This observation holds in this case with or without the inclusion of the option ancestor_subsumption, a procedure offered by OTTER to automatically attempt to seek shorter proofs by comparing derivation lengths to the same conclusion.

However, as noted before, the hints strategy in general gives less latitude; its use is more likely to have a program focus on the proof in hand. Indeed, when the thirty-two formulas of concern were included as hints and not as resonators and a small max_weight was assigned, OTTER discovered on its own so-to-speak a 32-step proof that deduces the Tarski-Bernays three-axiom system from the Lukasiewicz shortest single axiom for the implicational fragment of two-valued sentential calculus. For a fine point that might eventually provide needed insight in the future for some researcher, an assignment of the value 4 to max_weight did permit OTTER to retain the clause P(i(x,x)) when either the resonance strategy or the hints strategy was employed as described. This weight-4 clause was not used in any proof when the hints strategy was relied upon; it was present in proofs when resonance was used instead.

*The* 32-*Step proof*

```
----- Otter 3.0.5b, March 1998 -----
The process was started by wos on soot.mcs.anl.gov, Wed May 24 16:24:24 2000
The command was "otter". The process ID is 5369.

-----> EMPTY CLAUSE at   0.23 sec ----> 82 [hyper,34,53,77,79] $ANS(TARSKI_BERNAYS).

Length of proof is 32. Level of proof is 29.

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

33 [] -P(i(x,y)) | -P(x) | P(y).
34 [] -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).
35 [] P(i(i(i(x,y),z),i(i(z,x),i(u,x)))).
44 [hyper,33,35,35] P(i(i(i(i(x,y),i(z,y)),i(y,u)),i(v,i(y,u)))).
45 [hyper,33,35,44] P(i(i(i(x,i(y,z)),i(i(u,y),i(v,y))),i(w,i(i(u,y),i(v,y))))).
46 [hyper,33,45,35] P(i(x,i(i(i(y,z),y),i(u,y)))).
47 [hyper,33,46,46] P(i(i(i(x,y),x),i(z,x))).
48 [hyper,33,35,47] P(i(i(i(x,y),i(y,z)),i(u,i(y,z)))).
49 [hyper,33,35,48] P(i(i(i(x,i(y,z)),i(u,y)),i(v,i(u,y)))).
50 [hyper,33,35,49] P(i(i(i(x,i(y,z)),i(u,i(z,v))),i(w,i(u,i(z,v))))).
51 [hyper,33,50,35] P(i(x,i(i(i(y,z),u),i(z,u)))).
```

52 [hyper,33,51,51] P(i(i(i(x,y),z),i(y,z))).
53 [hyper,33,52,52] P(i(x,i(y,x))).
55 [hyper,33,52,35] P(i(x,i(i(x,y),i(z,y)))).
56 [hyper,33,35,55] P(i(i(i(i(i(x,y),z),i(u,z)),x),i(v,x))).
57 [hyper,33,35,56] P(i(i(i(x,y),i(i(i(y,z),u),i(v,u))),i(w,i(i(i(y,z),u),i(v,u))))).
58 [hyper,33,35,57] P(i(i(i(x,i(i(i(y,z),u),i(v,u))),i(w,y)),i(v6,i(w,y)))).
59 [hyper,33,35,58] P(i(i(i(x,i(y,z)),i(u,i(i(i(z,v),w),i(v6,w)))),i(v7,i(u,i(i(i(z,v),w),i(v6,w)))))).
60 [hyper,33,59,35] P(i(x,i(i(i(y,z),i(u,v)),i(i(i(z,w),v),i(u,v))))).
61 [hyper,33,60,60] P(i(i(i(x,y),i(z,u)),i(i(i(y,v),u),i(z,u)))).
62 [hyper,33,61,35] P(i(i(i(x,y),i(z,u)),i(i(x,u),i(z,u)))).
63 [hyper,33,62,55] P(i(i(x,i(y,z)),i(i(i(x,u),z),i(y,z)))).
64 [hyper,33,63,35] P(i(i(i(i(i(x,y),z),u),i(v,x)),i(i(z,x),i(v,x)))).
65 [hyper,33,35,64] P(i(i(i(i(x,y),i(z,y)),i(i(i(y,u),x),v)),i(w,i(i(i(y,u),x),v)))).
66 [hyper,33,65,65] P(i(x,i(i(i(i(y,z),u),i(v,z)),i(i(i(z,w),v),i(y,z))))).
67 [hyper,33,66,66] P(i(i(i(i(x,y),z),i(u,y)),i(i(i(y,v),u),i(x,y)))).
68 [hyper,33,61,67] P(i(i(i(i(x,y),z),i(u,y)),i(i(i(y,v),x),i(u,y)))).
69 [hyper,33,67,68] P(i(i(i(i(x,y),z),i(i(y,u),v)),i(i(v,y),i(x,y)))).
70 [hyper,33,68,62] P(i(i(i(i(x,y),z),u),i(i(u,y),i(x,y)))).
71 [hyper,33,69,64] P(i(i(i(x,y),z),i(i(i(y,u),z),z))).
73 [hyper,33,64,71] P(i(i(x,y),i(i(i(x,z),y),y))).
75 [hyper,33,73,73] P(i(i(i(i(x,y),z),i(i(i(x,u),y),y)),i(i(i(x,u),y),y))).
76 [hyper,33,70,75] P(i(i(i(i(i(x,y),z),z),u),i(i(x,z),u))).
77 [hyper,33,75,71] P(i(i(i(x,y),x),x)).
79 [hyper,33,76,70] P(i(i(x,y),i(i(y,z),i(x,z)))).

As promised, comparison and analysis is in order. Crucial (apparently) is the 30-step proof of TB-3. It contains two steps not in the Meredith-Prior 31-step proof of TB-3 and three not in the 32-step Lukasiewicz proof of TB-3, the latter two proofs being subproofs, respectively, of the cited 33-step and 34-step proofs. In the context of shorter subproofs do not necessarily a shorter total proof make, the 34-step prof contains subproofs of lengths 10, 25, and 32 respectively of TB-1, TB-2, and TB-3. The Meredith-Prior 33-step proof contain subproofs of lengths 10, 26, and 31. The given 32-step proof contains subproofs of lengths 10, 28, and 30. When resonance was used in the obvious attempt to produce the desired 32-step proof, the resulting 35-step proof contained subproofs of lengths 10, 23, and 30. The 31-step proof and the 30-step proof of TB-3 begin with the same twenty formulas, which might suggest that Meredith and Prior were proceeding along a most profitable path, a path that culminates (if appropriate action is taken) with the given 32-step proof. What won the game was (1) the finding of a 30-step proof deducing TB-3 and (2) cramming many of its steps into needed proofs of TB-1 and TB-2, constrained by a limit of 32 on length.

## 4. Incarnations of Cramming

The cramming strategy admits different incarnations, depending on the intent. For example, the intent may be to block the use of all or almost all new conclusions that fail to match an included resonator or hint. Instead, the intent may be to encourage the program to consider new conclusions, especially those that do not match an included resonator or hint. Then, within the total proof of the conjunction under attack, there is the question of which member's proof to use to cram, which of the subproofs of the total proof to choose to cram as many of its steps into proofs of the remaining members.

The first formulated incarnation of the cramming strategy has one choose from among the subproofs of the members of a conjunction that proof which is the longest. Its proof steps are added as so-to-speak lemmas in the set of support, and a level-saturation approach is conducted. The steps, as resonators or as hints, are (in effect) treated as being simple, short in nature, which can be done with an appropriate assigned value. Then max_weight is assigned the same small value or a value not much larger. The lemma-adjunction phase coupled with level saturation is intended to find the fewest additional steps needed to complete proofs of the members not directly in focus (through the use of resonance or hints). The small max_weight is chosen to prevent the program from straying.

In the case featured in the preceding section, the object was to show (if possible) that just two steps, in addition to the thirty that correspond to a 30-step proof of TB-3, sufficed. For resonators, thirty-two were chosen, one for each of TB-1 and TB-2 and thirty for the proof of TB-3. By using level saturation and (as lemmas) the deduced steps of the proof of TB-3, the plan was to force the program to apply condensed detachment to *all* pairs of the steps of the TB-3 proof together with the Lukasiewicz shortest single axiom. The small max_weight blocks the consideration of virtually all other pairs, which has (among others) the effect of avoiding the examination of possibly shorter proofs of TB-1 and TB-2. Without the actions taken, the program might indeed either fail to apply condensed detachment to some pair of formulas of the TB-3 proof or, more likely, reject the results because of having already in hand a shorter proof of, say, TB-2. As noted, the game would be won if two perfect pairs could be found from among the steps of the thirty-step proof of TB-1 (possibly including the Lukasiewicz axiom), where it was not required that the elements of a pair be distinct. A perfect pair, in this case, is one that yields with condensed detachment either TB-1 or TB-2.

The last phase of the first formulated incarnation of cramming consists of having the program focus on the selected subproof together with the new deduced steps its use yielded. Hints served well for having the program concentrate almost exclusively on the designated steps. This original incarnation was formulated specifically to thwart the program from seeking shorter proofs for individual members of the conjunction in focus.

As predictable, the given incarnation may fail to yield the steps needed to complete proofs of the remaining members of the conjunction under study. After all, little room is being given for steps other than those related to the included resonators or included hints. Further, the needed steps may number eleven or more. In such a case—and in a certain sense contrary to the original spirit of cramming—a second incarnation was born. One proceeds as in the first incarnation, but assigns a rather generous value to max_weight. For example, if the resonators or hints are treated as having a weight of 2, one might assign the value 16, or 20, or even 28 to max_weight. The intent is to find a few steps, some of which are definitely not in the set of subproofs in hand, such that a new total proof (of the conjunction) is available. The resonance strategy appears to be of more use in this context than does the hints strategy.

In this second incarnation, one begins by relying on the deduced steps of the longest subproof of the set of subproofs within the proof in hand of the conjunction to be proved. However, rather than including the remaining target elements of the conjunction as resonators or hints, one simply has the program find its own way to their respective proofs. This second incarnation, ordinarily accompanied by the use of ancestor subsumption, encourages the program to seek shorter proofs of the members other than that driving the reasoning. Of course, the length is not true length, for proof steps of the driving member are present as lemmas in the set of support. Indeed, the intent is to cram many of those steps into the needed proofs of the other members.

A fairly good indicator of the likelihood of success rests with a rather small difference, say five or less, between the level of the subproof driving the attack and the level of the total proof in hand. If the difference is much greater, then the level-saturation search might indeed be impractical, especially with a max_weight of 24 or larger. If one has not had the opportunity to run many experiments, the essentially exponential growth of level saturation is often an unpleasant surprise.

The following symbolic example illustrate the second incarnation of the cramming strategy. Consider the case in which one is asked to find a shorter proof than that in hand showing that some given formula is a single axiom. Assume one has in hand a proof of some known axiom system consisting of *A*, *B*, *C, and D*. Within that proof, by assumption, the longest proof is that of *B*. With the second incarnation of the cramming strategy, one places in the initial set of support, in addition to the single axiom, the deduced steps of the proof of *B*. One places in list(usable) (with OTTER) the disjunction produced by negating the conjunction of *A* through *D*. The deduced steps of the proof of *B* are used as resonators, each assigned a small value, and the max_weight is assigned the value 28. Level saturation is the choice for the search strategy.

The intent is that after some time proofs of each of *A*, *C*, and *D* will be found such that the lengths of the proofs suggests an advance has occurred with respect to proof length of the conjunction. If so, almost certainly, the proofs will depend on the use of proof steps from the proof of *B*. The socalled new proofs are then placed in a succeeding run as resonators with a small assigned value, say, 1. The proof steps of *B* are placed next, each assigned a value, say, 2. Level saturation is replaced with the ratio strategy. The max_weight is typically assigned the value 2.

If all goes as planned, a shorter proof is found. Quite likely the length of any or all of the subproofs of *A*, *C*, and *D* will be longer than their correspondents in the proof that prompted the use of the cramming strategy. The new and shorter proof will contain the cited proof of *B*, and an analysis of the proofs of the other members will show that a sometimes rather large number of the proof steps of the proof of *B*have been crammed into proofs of the other members.

A third incarnation of the cramming strategy has one focus on the next to the longest proof among the proofs of the members of the conjunction, provided that at least three items need to be proved. Its proof steps are (as in the other two incarnations) placed in the set of support to act as lemmas, and a level-saturation approach is taken. For some success to occur, usually, the proof length of the selected member best not be more than fifteen less than the length of the total proof. In addition, the proof length of the chosen member ordinarily had best be at least twelve. The suggested lengths are merely guesses based on experimentation. The idea is (1) not to have too far to travel to complete a proof of the most difficult-to-prove member (based on proof length and (2) to give the program a fair amount to work with in the form of lemmas.

A pleasing feature of the cramming strategy is its performance when iterated. I have, for example, taken a proof of a theorem, used the cramming strategy (in one of its incarnations) to produce a shorter proof, and then again used the cramming strategy on this new and shorter proof to gain even more ground. Occasionally, three or more such iterations have yielded one shorter proof after another.

## 5. Evaluating Cramming with Additional Successes

One naturally wonders whether the strategy just happened to succeed when the Lukasiewicz shortest single axiom was the focus. Any new strategy demands application toother areas to gain some insight into its generality and power. Here we touch on some of those applications.

For an example taken from recent research, the right group calculus serves nicely. J. Kalman originally offered a five-axiom system (the following) for this area of logic, and McCune showed that the second of the five is in fact sufficient to axiomatize the area.

```
P(d(z,d(z,d(d(x,d(y,y)),x)))).                          % K1
P(d(u,d(u,d(d(z,y),d(d(z,x),d(y,x)))))).                % K2
P(d(v,d(v,d(d(u,d(z,y)),d(u,d(d(z,x),d(y,x))))))).      % K3
P(d(d(d(u,d(v,y)),d(z,d(v,x))),d(u,d(z,d(y,x))))).      % K4
P(d(d(v,d(z,d(u,d(y,x)))),d(d(v,d(x,u)),d(d(z,d(x,y)),x)))).   % K5
```

When I chose to test the cramming strategy by studying this logic, McCune first told me that I should try to find which of the five axioms sufficed, rather than simply beginning with his success. In other words, I did not at the time recall that the second of the Kalman axioms can be used to deduce the remaining four. As it turned out, I began by focusing on the third axiom.

Early in my study of this third formula, I asked McCune about its status with regard to being a single axiom and learned that that question was in fact open. When I answered the question in the affirmative, a natural extension of the research was to focus on seeking a shorter proof showing that the conjunction of K_1, K_2, K_4, and K5 could be proved from K_3. What a fine test of the cramming strategy!

One can treat the problem of showing *K*-3 to be a single axiom as the limiting case of an axiom-dependence problem. The object is to prove that each of the other four (of the five Kalman axioms) is dependent on *k*-3 alone. A promising approach that appealed to me was to pick off the four axioms one at a time; start, for example, by trying to prove that *K*-1 is depend on the remaining four. If successful, use the resonators of the proof and seek to add to the list of dependent axioms, say, by trying to next prove *K*-2 dependent on 3 through 5. Eventually, only *K*-3 remained, with *K*-5 having been the most difficult to prove dependent. The reason may rest with its length.

When the four proofs were found, their steps as resonators led to the completion of a proof of the conjunction of 1, 2, 4, and 5 from 3. The proof of *K*-5 did not contain as a subproof a proof of any of

1, 2, and 4.

The proof of *K*-5 is generally longer than that of 1, 2, or 4. Therefore, for the cramming strategy, said proof offers more meat, more for the program to key upon as lemmas adjoined to the set of support for a level-saturation run. The goal is to find proofs of 1, 2, and 4 in which (no doubt) steps of *K*-5 will be present, crammed into their proofs. When an iterative cramming attack was initiated, I had in hand a proof of length 39 of the conjunction of 1, 2, 4, and 5 and a 29-step proof of 5. The first phase proved 1, 2, and 4 in a total of eight steps, in addition to the twenty-nine from a proof of *K*-5. Therefore, most likely cramming led to finding a proof of length strictly less than 39 for the target conjunction. It turned out that a proof of length 36 was yielded.

Occasionally with recourse to other methodologies to find a slightly shorter proof at various stages, the final iteration with cramming led to OTTER's discovery of a 27-step proof. The last iteration relied upon an 18-step proof of *K*-5, finding nine steps that together prove 1, 2, and 4. Currently, I know of no shorter proof that deduces the conjunction of 1, 2, 4, and 5 for *K*-3. In summary, the open question concerning the axiomatic status of *K*-3 was settled, and it was settled quite well in the sense of finding through cramming a rather short proof.

Next in order was the application of the cramming strategy to *K*-2. McCune (as noted) had proved it to be a single axiom, but not by directly deducing the conjunction of 1, 3, 4, and 5. Again, iteration was the choice, with a succession of proofs of *K*-5 playing the starring role. The last phase of the iteration yielded a 29-step proof of the desired conjunction, relying upon a 21-step proof of *K*-5 and n nine steps proving 1, 3, and 4 through cramming steps of the *K*-5 proof into proofs of 1, 3, and 4.

Finally, in order was the open question of the axiomatic status of *K*-4; 1 and 5 are insufficient. That question was settled in the affirmative. A 49-step proof was the first I found. Through cramming with iteration, OTTER eventually discovered a 31-step proof of the conjunction of 1, 2, 3, and 5. Rather than a succession of proofs of *K*-5 playing the key role, those for *K*-3 did; they were longer and more difficult in general to find.

In contrast to the preceding example that rested on the second incarnation of the craming strategy and to the prize of this article (the 32-step proof given earlier) that was discovered with the first incarnation, an example of a success with the third incarnation is provided by focusing on two-valued sentential calculus studied in terms of the Sheffer stroke. Various single axioms have been found for this area of logic, each of length twenty-three (not counting a predicate symbol. My colleague Fitelson has most profitably attacked this area and found many other such axioms including finding the following new single axiom.

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

To prove a candidate formula to be a single axiom in the Sheffer stroke, one can aim at proving the following Nicod 1-basis.

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

Instead, one can seek a proof of the following Lukasiewicz 3-basis.

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

Fitelson asked for a short proof for his axiom, and I considered both targets, with a bit more emphasis on the Lukasiewicz 3-basis. I soon found a 47-step proof and made some advances to the point of having in hand a 39-step proof, obtained by cramming Within that proof is a 28-step proof of the third member of the Lukasiewicz axiom that in turn contains a proof of the first member. The second member was proved in thirty-six applications of condensed detachment. Because little room existed between the longer of the three subproofs and the total proof, I turned to the third incarnation of the cramming strategy, focusing on the 28-step proof as lemmas with level saturation.

The problem offered substantial resistance. Indeed, after fourteen CPU-hours, an 11-step proof of the second member was found, which was of much use. However, after twenty-nine CPU-hours, a 9-step proof of the desired type was found. As a sign of the robustness of OTTER and also of the difficulty of the problem, the cited proof was found after 25999 (given) clauses had been chosen to direct the search.

The results of this long run strongly suggested that a 37-step proof of the 3-basis could be produced, which in fact is what occurred. One of the aspects of cramming that makes it so effective is its consideration of formulas whose complexity might otherwise cause the program to ignore them for participating in the reasoning. As evidence, in the cited experiment, three of the steps of the 37-step proof do not appear in the 39-step proof, each of weight twenty-four. for the curious, the shortest proof found in my experiments with the Fitelson axiom that completes with the deduction of the Lukasiewicz basis has length 34. If Nicod's 1-basis is used as the target instead, I can offer a 31-step proof.

Yet one more example of the effects of cramming much of one proof into another is in order. The focus is Meredith's second single axiom (the following) for the <C,o> calculus.

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

The target in this study was (what appears to be) the shorest basis for this area of logic, a 2-basis from Lukasiewicz, the following.

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

The first member of this 2-basis is an old friend (from Section 2), namely, Lukasiewicz's shortest single axiom for the implicational fragment of two-valued logic.

Through the use of various methodologies, I had found a 61-step proof of the 2-basis from the Meredith axiom. Within that proof, one finds a 54-step proof that deduces the first member of the 2-basis and a 10-step proof of the second member. With cramming and the focus on the 54-step proof as temporary lemmas, OTTER discovered a 56-step proof of the 2-basis. The newer proof is obtained by (in effect) trading the cited 10-step proof for a 35-step proof of the second member of the 2-basis, a proof of course sharing all but two of its steps with the cited 54-step proof. In other words, thirty-three of the fifty-four steps of the proof of the first member were crammed into a 35-step proof of the second member. So much cramming occurred in fact that the second needed proof was nearly absorbed.

Regarding another example of success with the original incarnation, my study of *C*5 serves well. That area of logic is the implicational fragment of *S*5, a modal logic that was formulated in part to capture the more commonly accepted notion of implication. Meredith offered the following single axiom for *C*5.

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

After various experiments, my colleague Fitelson and I had found a 28-step proof, completing with the deduction of a Meredith 3-basis. That proof contains a 24-step subproof of one of the members whic, when used with cramming suggested the a 27-step proof of the 3-basis exited. With the max_weight assigned a value of 4 and the inclusion of resonators for the 24-step proof together with three obtained with cramming, the first incarnation did yield a 27-step proof. Again a piquant trade was (in effect) made: An 8-step proof of the first member of the 3-basis was traded for a 25-step proof.

At a later point in the study, I had obtained a 21-step proof that was compact, meaning that the longest subproof of the three subproofs of the members of the 3-basis was in fact a proof of the entire basis. Put another way, I had found a proof of one of the members of the 3-basis such that it contained as subproofs proofs of the other two members. Still intent on more progress, I turned to the third incarnation, focusing on a 12-step proof of the first member as lemmas for cramming. That action produced eight steps of proof such that a 20-step proof was found, but, instead of deducing the 3-basis, a new and shorter 2-basis (due to Ernst) was the target for completion. In particular, cramming had succeeded in coping with a compact proof by using the proof steps obtained from cramming with one target to reach in a shorter distance a different target axiom system.

## 6. Highlights and Summary

The cramming strategy attempts to take a proof **P** of a member of a conjunction and find proofs of the remaining members by forcing or cramming many of the steps of **P** into the needed proofs. The objective is to discover a proof of the entire conjunction that is shorter than that in hand. The proof **P** is a subproof of the proof in hand of the conjunction, often the longest of the subproofs of the members of the conjunction. The cramming is by means of a type of lemma adjunction coupled with level saturation. Specifically, the steps of **P** are added to the hypothesis or hypotheses already present in the set of support. The introduction in this article of the cramming strategy marks a somewhat different use of level saturation.

One of the piquant properties of the strategy is its encouragement of trading shorter proofs for longer proofs when the target is a member of a conjunction, where the objective is to find a shorter proof of the whole. The relevance to the newly discovered Hilbert twenty-fourth problem rests with the notion that shorter proofs are in general simpler proofs, although no such guarantee exists. Of course, proofs ca be shortened (as occurs in literature) by simply omitting steps and treating them as implicitly present. Explicitly is required that all steps of a proof be present and that one or more inference rules be specified.

If the use of the cramming strategy leads to advances in the direction of proof refinement with respect to length, they can usually be traced to the introduction of new proof steps. (this observation may seem patently obvious until one notes that a shorter proof can sometimes be found such that all of its steps are among the formulas or equations of the longer proof.) Such new steps often arise from the application of an inference rule to a set of hypotheses that might ordinarily be skipped because of complexity considerations. for example, condensed detachment might be applied to a pair of formulas one or both of which have a length that causes the program to fail to consider the pair.

In the context of cramming, the prize to this date in early 2001 is an abridgment of the Meredith-Prior abridgment of a Lukasiewicz proof. The proof concerns the Lukasiewicz shortest single axiom for the implicational fragment of two-valued sentential (or propositional) calculus. Added satisfaction is derived from the knowledge that Meredith in particular was quite concerned for proof length. Other successes are given in Section 5, some derived from each of the three incarnations of the cramming strategy offered in Section 4.

the prized 32-step proof (given in Section 3) was discovered by cramming so much of a 30-step proof into the proofs of two other members of a Tarskki-Bernays axiom system that only two additional steps were needed to prove the conjunction, that for *TB*-1 and *TB*-2. As it turned out, the 30-step proof (of *TB*-3) that starred in the success with cramming itself offered unique properties. Indeed, subsequent experimentation led to finding two other 30-step proof of *TB*-3, neither of which enable the completion of a 32-step proof. Indeed, the 30-step proof that appears as a subproof of the given 32-proof is such that (clearly) two of its pairs admit the deduction, respectively, of *TB*-1 and *TB*-2 with a single application of condensed detachment.

One has yet another wondrous example of the beauty of logic and mathematics, to the point of being startling. Indeed, as I myself was so vividly aware before running the crucial experiment, the probability was near zero of cramming exactly two steps and the 30-step proof in hand into a 32-step proof with the desired properties. Especially when one has access to a powerful reasoning program such as OTTER, one might be strongly tempted to emphasize the study of proofs and their properties rather than placing the emphasis almost exclusively on which assertions are in fact theorems. More generally, (as my Colleague Ted Ulrich has observed) an entire field might emerge devoted to the study of the space of proofs, in contrast to the study of the space of deducible conclusions. For but one question, does there exist a small set of resonators or hints (sa, less than 100) that provides powerful guidance for studying disparate fields of logic or of mathematics?

The cramming strategy opens a new window to the huge set of proofs for any given theorem, a set far lager than one might ordinarily imagine. Its use can focus on sets of hypotheses that an unaided researcher is prohibited from considering. The result might be the discovery of a proof some of whose steps would have proved most elusive. Cramming gives one a weapon for attacking the proof of a conjunction as a whole in a manner that avoids traps that take the form of encountering a very sort proof of ne of the members of the conjunction. How effective this strategy will prove in domains where equality dominates is yet to be determined. Also unknown at this time is its usefulness in the context of a target that consists of a single formula or equation. As for the case in which the axiom set under study does not concern a single axiom, I can at this time offer no experimental data, but conjecture all will go well.

An example of the profitable interplay of cramming with other methodologies designed to find shorter proofs is provided by focusing on the use of demodulation to block steps of a proof in hand. In particular, where difficulty concerns proof length alone, if blocking a step of a proof of a conjunction produces another proof of the same length as that in hand, but if such blocking produces a shorter proof of the most difficult member to prove, then one can sometimes profitably apply cramming to the new and shorter subproof. The increase in the distance between the length of the most difficult-to-prove member and the entire conjunction is encouraging. In a distantly related way, cramming is more likely to yield results of interest if the focus is on the longest subproof of the members of a conjunction rather than on the next to the longest. The explanation rests with a lager supply of so-to-speak lemmas for the program to rely upon.

As for the use of cramming in the context of answering open questions, my studies of the right-group calculus provide some insight. Specifically, imagine that one is attempting to deduce three formulas or equations from a set of hypotheses, and that one has in hand proofs of two of the three. One takes the steps of the longer of the two proofs, uses them as lemmas in the set of support, chooses a level-saturation approach, and attempts to cram some or all of those steps into a proof of the third member of the conjunction.

Cramming also admits an odd type of inversion. In particular, one can focus on the most difficult (in terms of length) to prove member to cram many of its steps into a proof of the next to the most difficult. Then one can turn instead to the next most difficult to prove in the context of the newer and shorter proof and attempt to cram many of its steps into a new shorter proof of the most difficult to prove member.

The introduction of cramming and the evidence of its effectiveness add more support to the assertion that an automated reasoning program can indeed play an indispensable role in research.