SLAC-PUB-579
(MISC)

AXIOM SYSTEMS IN AUTOMATIC THEOREM PROVING*

by

George Robinson
Stanford Linear Accelerator Center
Stanford, California, U.S.A.

and

Lawrence Wos
Argonne National Laboratory
Argonne, Illinois, U.S.A.

Presented at the IRIA Symposium on

Automatic Demonstration at Versailles, France
December 16-21, 1968

*Work performed under the auspices of the U. S. Atomic Energy Commission.

In judging the suitability of axiom systems for automatic theorem proving in a particular domain, two considerations are of prime importance: *logical strength* and *proof-search efficiency*. The axiom system should be just strong enough to yield all the theorems of the theory and only the theorems of the theory. We shall employ the concept of *adequacy* (defined in the following section) to study this property. Given two axiom systems each of appropriate strength for the theory in question, one would prefer to use the system that would, when used by the theorem-proving programs in question, discover proofs in the shortest time (or using the smallest amount of memory or other computer resources). Examples showing how the choice of axiom system can have important effects on proof-search will be given in the section on efficiency.

We shall be dealing with a language equipped with denumerably many individual variables, individual constants, n-adic function constants, and n-adic predicate constants. An individual constant or individual variable is a *term*, as is an n-adic function constant followed by n terms. A *ground term* is one involving no variables. An *atomic formula* (or *atom*) is an n-adic predicate letter followed by n-terms. A *literal* is an atomic formula or the negation thereof. A *clause* is the disjunction of finitely many literals; it is a *ground clause* if no variables occur. It is often profitable to identify clauses, particularly ground clauses, with the set whose members are the literals occurring in the clause.

If some or all of the variables of a clause are systematically replaced by terms, the resulting clause is an *instance* of the original clause.

A set of ground literals (literals involving no variables) *satisfies* a ground clause if it has a non-empty intersection with the clause; it *satisfies* any clause (with respect to some herbrand universe, H_S) if it satisfies all ground instances (over H_S) of that clause. If a set of literals contains the negation of each literal of a ground clause, it *condemns* the ground clause; if it condemns some ground instance of any clause, it *condemns* the clause. The clause $\square$ (read "the empty clause" or just "nil") having no literals can be satisfied by no set of literals and is (vacuously) condemned by every such set.

The *herbrand universe*, H_S, of a set S of clauses is the set of all ground terms that can be formed from the functions and individual constants occurring in S, with the constant a being supplied in the case that S contains no indidual constants. An *herbrand atom* for S is an atomic formula composed of a predicate letter occurring in S and the appropriate number of terms from H_S. An *interpretation* I of S over H_S is a set of (ground) literals such that for each herbrand atom L for S, exactly one of L or ∿L is in I. A *model* of S (over H_S) is an interpretation of S over H_S that satisfies S.

If S and T are sets of clauses, the S *implies* T (S ⊨ T) if no model of S condemns T. Then each clause C in T is a *logical consequence* of S (is *implied by* S, S ⊨ C).

This paper concerns itself principally with sets of clauses. There appears to be no reason, in principle, why the discussion should not go over to more general sentences, such as those obtained by using logical operators for conjunction, conditional, biconditional, and more general application of negation and quantification. (In effect, with clauses one considers only universal quantification over each clause.) In the discussion of eliminability, the biconditional ≡ is in fact used, since the statement of this property appears to become inordinately complex in terms of clauses alone. We assume some conventional rules of function, inference, and semantics for these additional connectives.

For clauses, three rules of inference are of prime importance in automatic theorem proving: *factoring, paramodulation,* and *resolution.*

Definition (Paramodulation): Let A and B be clauses such that a literal Rst (or Rts) occurs in A and a term u occurs in (a particular position in) B. Further assume that s and u have a most general common instance $s' = s\sigma = u\tau$ where σ and τ are the most general substitutions such that $s\sigma = u\tau$. Where $\hat{B}$ is obtained by replacing by $t\sigma$ the occurrence of $u\tau$ in the position in $B\tau$ corresponding to the particular position of the occurrence of u in B, infer the clause $C = \hat{B} \cup (A - \{Rst\})\sigma$ (or $C = \hat{B} \cup (A - \{Rts\})\sigma$). C is called a *paramodulant* of A and B and is said to be *inferred by paramodulation from* A *on* Rst (or Rts) *into* B *on (the occurrence in the particular position in* B *of)* u. The literal Rst (or Rts) is called the *literal of paramodulation.* [4]

Definition: For any literal l, /l/ is that atom such that either l = /l/ or l = ∼/l/.

Definition (Resolution): If A and B are clauses with literals k and l respectively, such that k and l are opposite in sign (i.e., exactly one of them is an atom) but /k/ and /l/ have most general common instance m, and if σ and τ are the most general substitutions with $m = /k/\sigma = /l/\tau$, then infer from A and B the clause $C = (A - \{k\})\sigma \cup (B - \{l\})\tau$. C is called a *resolvent* of A and B and is inferred by *resolution.*

Definition (Factoring): If A is a clause with literals k and l such that k and l have a most general common instance m, and if σ is the most general substitution with $k\sigma = l\sigma = m$, then infer the clause $A' = (A - \{k\})\sigma$ from A. A' is called an *immediate factor* of A. The *factors* of A are given by: A is a factor of A, and an immediate factor of a factor of A is a factor of A.

Given a set S of sentences and implicitly understood rules of formation and semantics, we shall be interested in the set W(S) of all well-formed sentences over the vocabulary of S (with a denumerable supply of individual variables added if necessary) and the set $V(S) = \{A | A \in W(S), S \models A\}$. A theory T will be thought of as being defined by the set W(T) of well-formed sentences of the theory and the set V(T) of valid formulas of the theory even when we have no particular set of axioms in mind for T.

Adequacy of an Axiom System

A set E of sentences is a *non-creative** extension of a set E' if $V(E) \cap W(E') = V(E')$; it is an *eliminable* extension if for every $C \in W(E)$ there is a $C' \in W(E')$ such that $E \models C \equiv C'$.

In Figure 1 two (redundant) sets of axioms are given.** A1-8 were obtained by writing in clause form a set of axioms given by Abraham Robinson*** for group theory in terms of a single binary relation

*The concepts of non-creativity and eliminability as used here are closely related to the two criteria for definitions given in [7], page 154; hence the choice of terminology.

**{A1,...,A8}, {A6,...,A10}, and {A6,A7,A8,A9,A11,A12} are equivalent sets. {B1,B2,B5,B6,B7,B8,B9,B11,B13}, {B6,B7,B9,B11,B13,B14}, and {B6,B7,B11,B12,B13,B15,B16} are equivalent sets and each of them implies the three sets of A-axioms. Since A3 appears to be a more natural way of stating associativity one might ask whether A10 might not be replaced by A3 in the set A6-10. This question is answered negatively by the following counter-example:

Consider a domain of three elements {0,1,2}. Let f map to the usual cyclic group operation on this domain, let g map to the corresponding group inverse operation, and let e map to 0, but let R map to the equivalence relation K such that $0 \neq 1 = 2$; namely $K = \{(0,0), (1,1), (1,2), (2,1), (2,2)\}$. A3, A6, A7, A8, and A9 are all obviously satisfied, but A5 is falsified by the choice of $x = y = z = 1$ and $u = 2$, since $(1,1) \in K$, $(1,2) \in K$, and $(f(11),f(12)) = (2,0) \notin K$.

***Reference [3], p. 26.

A1	$\overline{R}xy\ Ryx$	(sym)	B1	$\overline{R}xy\ Ryx$	
A2	$\overline{R}xy\ \overline{R}yz\ Rxz$	(trans)	B2	$\overline{R}xy\ \overline{R}yz\ Rxz$	
A3	$Rf(f(xy)z)f(xf(yz))$	(assoc.)			
A4	$\overline{R}xy\ Rg(x)g(y)$	(g-subst.)			
A5	$\overline{R}xz\ \overline{R}yu\ Rf(xy)f(zu)$	(f-subst.)	B5	$\overline{R}xt\ \overline{R}yu\ \overline{R}zw\ \overline{P}xyz\ Ptuw$	
A6	$Rf(ex)x$	(l.ident.)	B6	$Pexx$	
A7	$Rf(g(x)x)e$	(l.inv.)	B7	$Pg(x)xe$	
A8	Rxx	(reflex.)	B8	Rxx	
A9	$\overline{R}xz\ \overline{R}yz\ Rxy$		B9	$\overline{P}xyz\ \overline{P}xyu\ Rzu$	(uniq.of prod.)
A10	$\overline{R}f(xy)u\ \overline{R}f(yz)t\ Rf(uz)f(xt)$				
A11	$\overline{R}f(xy)u\ \overline{R}f(yz)t\ \overline{R}f(uz)w\ Rf(xt)w$	(assoc.)	B11	$\overline{P}xyu\ \overline{P}yzt\ \overline{P}uzw\ Pxtw$	
A12	$\overline{R}f(xy)u\ \overline{R}f(yz)t\ \overline{R}f(xt)w\ Rf(uz)w$		B12	$\overline{P}xyu\ \overline{P}yzt\ \overline{P}xtw\ Puzw$	
			B13	$Pxyf(xy)$	(closure)
			B14	$\overline{R}zu\ \overline{P}xyz\ Pxyu$	(P_3-subst)
			B15	$\overline{R}xy\ Pexy$	
			B16	$\overline{P}exy\ Rxy$	
			B17	$\overline{P}xyz\ \overline{P}xyu\ Pezu$	
			B18	$\overline{P}ezu\ \overline{P}xyz\ Pxyu$	
			B19	$\overline{R}f(xy)z\ Pxyz$	
			B20	$\overline{P}xyz\ Rf(xy)z$	
A21	$Rf(xe)x$	(r.ident.)	B21	$Pxex$	
A22	$Rf(xg(x))e$	(r.inv.)	B22	$Pxg(x)e$	

Figure 1.

R and two functions f and g. B1-2, B5-9, B11, and B13 were obtained from another set of group theory axioms in the same book**** by replacement (in a set of sentences not originally involving function symbols) of existentially-quantified variables by Skolem function symbols.

Consider the set S4 = S1 ∪ {B15,B16}, where S1 = {B6,B7,B11,B12,B13}. Thus defined, S4 is obviously a non-creative, eliminable extension of S1.

If a theory T is a non-creative, eliminable extension of a set S of axioms, then the set S is of appropriate logical strength for a study of the theory T by means of automatic theorem-proving. That is, since every sentence C in W(T) can be mapped into a sentence Cμ in W(S) in such a way that Cμ is in V(S) iff C is in V(T), we need only apply a proof procedure to V(S). But confining the choice of axiom sets to those which have T as a non-creative, eliminable extension is unnecessarily restrictive. It forbids, for example, using a system such as S1 to study one such as A1-A8, when the former is much more efficient in proof search with some types of inference apparatus than the latter. This problem is not avoided by allowing the use of a set S which is itself a non-creative, eliminable extension of the theory. It appears appropriate for automatic theorem proving to make only a requirement concerning the existence of an appropriate (and effective) mapping μ. We shall do this by defining a set of sentences S to be *adequate* for a theory T iff there exists a uniform means of transforming the atomic formulae of W(T) into atomic formulae of W(S) such that the mapping μ induced on all sentences of W(T) maps the

****op. cit., p. 229.

sentences C in W(T) into sentences $C\mu$ of W(S) such that $C\mu \in V(S)$ iff $C \in V(T)$.*****

Whenever T is a non-creative, (effectively) eliminable extension of S, S is adequate for T. By *effectively eliminable* we mean that there is an effective means of determining the C', given the C. In that case, let $C\mu = C'$ for $C \in W(T) - W(S)$ and let $C\mu = C$ for $C \in W(S)$. Then for $C \in W(S)$, $C\mu = C \in V(S)$ iff $C \in V(T)$ since $V(T) \cap W(S) = V(S)$. For $C \in V(T) - V(S)$, we have $C\mu \in V(T)$ (since $T \models C \equiv C\mu$) and hence $C\mu \in V(S)$. For $C \in W(T) - V(T)$, we have $C\mu \notin V(S)$, since if $C\mu \in V(S)$ we would have $C\mu \in V(T)$ and hence $C \in V(T)$ (since $T \models C \equiv C\mu$) contrary to hypothesis.

On the other hand, S can be adequate for T but fail to be a non-creative extension of S as illustrated by the trivial example $T = \{P\}$, $S = \{\overline{P}\}$ (μ is the mapping from P to $\overline{P}$, and from $\overline{P}$ to P).

If we let S2 = {B9,B14} and consider $S3 = S1 \cup S2$ we find that S1 is adequate for S3.****** To see that this is the case, let τ map each

*****It may be possible to generalize the concept of adequacy by partially relaxing the restriction that μ be induced by a transformation of atomic formulae. The restriction is quite natural for automatic theorem proving, since it requires, in effect, that $C\mu$ be a clause if C is a clause. The restriction apparently cannot be completely removed to allow arbitrary effective μ without admitting, for example, certain pathological mappings by which any set of tautologies in a sufficiently rich vocabulary could be shown adequate for any finitely-axiomatizable theory.

******Henceforth we shall restrict the discussion to consider only clauses as sentences, i.e., W(S) will always be a set of clauses.

atomic formula of the form $R\alpha\beta$ into $Pe\alpha\beta$ while leaving one of the form $P\gamma\alpha\beta$ unchanged. Then let θ be the mapping on clauses induced by τ. First we show that θ maps V(S3) into V(S1). Suppose by way of contradiction that for some C, S3 $\models$ C but S1 $\not\models$ Cθ. Then there is a model M of S1 over H_{S1} that condemns some ground instance Dθ of Cθ, where D is a ground instance of C. Let $M^* = M \cup \{R\alpha\beta | Pe\alpha\beta \in M\} \cup \{\overline{R}\alpha\beta | Pe\alpha\beta \notin M\}$. Then M* is an interpretation of S3 that satisfies S1. If $R\alpha\beta$ and $P\gamma\delta\alpha$ are in M*, then $Pe\alpha\beta$ and $P\gamma\delta\alpha$ must be in M and, since $\overline{P}ezu\ \overline{P}xyz\ Pxyu$ is a theorem of S1, it follows that $P\gamma\delta\beta$ is in M and hence in M*. Thus M* satisfies B14. A similar argument shows that M* satisfies B9 as well, hence M* is a model of S3 and must satisfy D. Let $M^* \cap D = E$. Then $M \cap D\theta = E\theta \neq \emptyset$, contrary to the hypothesis that M condemned Dθ. It remains to show that θ maps only elements of V(S3) into V(S1): Since S3 $\models$ $\overline{P}exy$ Rxy and S3 $\models$ $\overline{R}xy$ Pexy, it follows that S3,Cθ $\models$ C for any C $\in$ W(S3). If S1 $\models$ Cθ, then S3 $\models$ Cθ, since S1 $\subseteq$ S3. Hence if S1 $\models$ Cθ, S3 $\models$ C.

To show that S3 is adequate for A1-8, let μ be the identity mapping on clauses and first note that S3 $\models$ Ai for i - 1,...,8. (Two of the examples in the section on proof search efficiency contain proofs that S3 $\vdash$ B21-22. Proofs that S3,B21-22 $\vdash$ A1-8 are given in Appendix A.) Hence if C $\in$ V(A1-8), then Cμ = C $\in$ V(S3). Now suppose that C $\in$ W(A1-8) - V(A1-8). Then there must be a model M of A1-8 over H_{A1-8} that condemns C. Let $M^* = M \cup \{P\alpha\beta\gamma | Rf(\alpha\beta)\gamma \in M\} \cup \{\overline{P}\alpha\beta\gamma | Rf(\alpha\beta)\gamma \notin M\}$. Now A1-8 $\cup$ {$\overline{R}$f(xy)z Pxyz, $\overline{P}$xyz Rf(xy)z} $\models$ S3. Hence M* must be a model of S3. Since M* condemns C, Cμ = C cannot be in V(S3).

In general, if S' is adequate for S and S is adequate for T, it follows that S' is adequate for T, since if μ' is the mapping of S into S' and μ the mapping of T into S, $\mu\mu'$ will do for the mapping of T into S'. In particular, since S1 is adequate for S3 and S3 adequate for A1-8, S1 must be adequate for A1-8 and in turn for any theories for which A1-8 may be adequate. One might wish to note that A1-8 is not a non-creative extension of S1.

Proof-search Efficiency

In the literature on automatic theorem-proving, considerable attention has been devoted to selection of an efficient proof search algorithm. [1], [2], [4], [6], [8], [10]. Nevertheless, at the present state of the art, the ease--even the feasibility--of automatic theorem proving in a given theory (e.g., group theory) is vitally affected by the choice of axioms and representations of theorem-candidates for the study of the theory. Indeed, the choice can be so important that it is difficult to find examples to compare search times for systems such as A1-8 with more efficient (with a given proof search algorithm) systems such as S3. With an unfortunate choice of axioms, the running times for quite simple theorems can be prohibitively long to obtain a numerical comparison.

The figures for running times given in this section are for the PG1-PG5 series of theorem-proving programs developed at Argonne a number of years ago and more fully described in [8], [9], and [10]. PG5 has been singled out for use in most of the cases run because it provides the fairest basis for comparing diverse axiom systems, even though it is slower in some cases than some others in the PG1-PG5 family. In order to obtain numerical comparisons for efficiency of axiom systems in ordinary first-order theories with no special treatment of equality, the demodulation apparatus of PG5 was disabled. It proved to be infeasible to obtain, with the programs available, conclusive efficiency comparisons for first-order theories with equality, due to incompleteness difficulties introduced by the treatment of demodulation in those programs having special treatment of equality.

First we shall consider the example of proving that in a group a right inverse exists. As is usually the case in automatic theorem proving, the procedure is to deny the existence of a right inverse and proceed to refute the denial by appeal to the axioms. In the vocabulary of A1-8, the denial is $\overline{R}f(ay)e$; for S3 it is $\overline{P}aye$. PG5 obtained a refutation from S3 in less than one second, but could obtain no refutation from A1-8 in the 288 seconds allowed for running the case. Some insight into the difficulties may perhaps be obtained by examining refutations in the two systems. One refutation from A1-8 is as follows:

1.	$\overline{R}xy\ Ryx$	A1-sym
2.	$\overline{R}xy\ \overline{R}yz\ Rxz$	A2-trans
3.	$Rf(f(xy)z)f(xf(yz))$	A3-assoc
4.	$\overline{R}xz\ \overline{R}yu\ Rf(xy)f(zu)$	A5-f-subst
5.	$Rf(ex)x$	A6-1. ident
6.	$Rf(g(x)x)e$	A7-1. inv
7.	Rxx	A8-reflx
8.	$\overline{R}f(ay)e$	denial
9.	$Rf(xf(yz))f(f(xy)z)$	A3-A1$_1$
10.	$\overline{R}xz\ Rf(xf(ey))f(zy)$	A6-A5$_2$
11.	$Rf(xf(ey))f(xy)$	10$_1$-A8
12.	$\overline{R}wf(zf(ex))\ Rwf(zx)$	11-A2$_2$
13.	$Rf(f(ze)x)f(zx)$	A3-12$_1$
14.	$\overline{R}wf(g(x)x)\ Rf(we)$	A7-A2$_1$
15.	$Rf(f(g(x)e)x)e$	13-14$_1$
16.	$\overline{R}f(ax)y\ \overline{R}ye$	8-A2$_3$

17.	$\overline{R}f(ax)f(f(g(y)e)y)$	$15\text{-}16_2$
18.	$\overline{R}af(g(y)e\ \overline{R}xy$	$17\text{-}A5_3$
19.	$\overline{R}af(g(x)e$	$A8\text{-}18_2$
20.	$\overline{R}yu\ Rf(f(g(x)x)y)f(eu)$	$A7\text{-}A5_1$
21.	$Rf(f(g(x)x)y)f(ey)$	$A8\text{-}20_1$
22.	$\overline{R}wf(f(g(x)x)y)\ Rwf(ey)$	$21\text{-}A2_2$
23.	$Rf(g(x)f(xz))f(ez)$	$9\text{-}22_1$
24.	$\overline{R}wf(ez)\ Rwz$	$A6\text{-}A2_2$
25.	$Rf(g(x)f(xz))z$	$23\text{-}24_1$
26.	$Rzf(g(x)f(xz))$	$25\text{-}A1_1$
27.	$\overline{R}f(g(x)f(xz))w\ Rzw$	$26\text{-}A2_1$
28.	$\overline{R}f(g(x)f(xa))f(g(y)e)$	$19\text{-}27_2$
29.	$\overline{R}g(x)g(y)\ \overline{R}f(xa)e$	$28\text{-}A5_3$
30.	$\overline{R}f(ya)e$	$A8\text{-}29_1$
31.	$\square$	A7-30

Contrast that refutation with the following one obtained from S3:

1.	$Pexx$	B6-1.ident
2.	$Pg(x)xe$	B7-1.inverse
3.	$\overline{P}xyu\ \overline{P}yzt\ \overline{P}uzw\ Pxtw$	B11
4.	$\overline{P}xyu\ \overline{P}yzt\ \overline{P}xtw\ Puzw$	B12
5.	$\overline{P}aye$	denial
6.	$\overline{P}xya\ \overline{P}yzt\ \overline{P}xte$	$5\text{-}4_4$
7.	$\overline{P}xea\ \overline{P}xte$	$1\text{-}6_2$
8.	$\overline{P}g(t)ea$	$2\text{-}7_2$
9.	$\overline{P}g(t)yu\ \overline{P}yze\ \overline{P}uza$	$8\text{-}3_4$

10.	$\overline{P}g(t)ye\ \overline{P}yae$	$1\text{-}9_3$
11.	$\overline{P}yae$	$2\text{-}10_1$
12.	□	11-2

For further contrast, we cite the following proof from A1-8 using the special equality mechanisms of paramodulation. This proof suggests that A1-8 (more precisely the subset {A3,A6,A7,A8}) is probably a better choice than S3 when such special mechanisms for equality are incorporated into the theorem-proving program.

1.	Rf(f(xy)z)f(xf(yz))	A3
2.	Rf(ex)x	A6
3.	Rf(g(x)x)e	A7
4.	$\overline{R}f(ay)e$	denial of theorem
5.	Rf(f(xg(z))z)f(xe)	A7(f(g(x)x))-A3(f(yz))
6.	Rf(f(ez)f(g(g(z))e)	A7(f(g(x)x))-5(f(xg(z))
7.	Rzf(g(g(z))e)	A6(f(ex))-6(f(ez))
8.	Rf(f(xe)z)f(xz)	A6(f(ex))-A3(f(yz))
9.	Rf(f(g(z)e)z)e	A7(f(g(x)x))-8(f(xz))
10.	Rf(zg(z)e)	7(f(g(g(z)e))-9(f(g(z)e))
11.	□	10-4

Similar results are obtained when the two systems S3 and A1-8 are used to refute the denial that in a group, the left identity element is also a right identity. A refutation from A1-A8 looks much like that for right inverse. No refutation is obtained by PG5 from this set after

288 seconds, while less than one second is required for a refutation from S3. This is not surprising since short refutations such as the following are available from S3.

1.	Pexx	B6
2.	Pg(x)xe	B7
3.	$\overline{P}xyu\ \overline{P}yzt\ \overline{P}uzw\ Pxtw$	B11
4.	$\overline{P}xyu\ \overline{P}yzt\ \overline{P}xtw\ Puzw$	B12
5.	$\overline{P}aea$	denial
6.	$\overline{P}xya\ \overline{P}yet\ \overline{P}xta$	$5\text{-}B12_4$
7.	$\overline{P}xea$	$B6\text{-}6_2$
8.	$\overline{P}xyu\ \overline{P}yze\ \overline{P}uza$	$7\text{-}B11_4$
9.	$\overline{P}xye\ \overline{P}yae$	$B6\text{-}8_3$
10.	$\overline{P}yae$	$B7\text{-}9_1$
11.	□	10–B7

If we are correct in our conjecture that it is advantageous to use additional free variables and, if necessary, additional literals in order to avoid long terms as arguments, one would expect that adding, say, A11 and A12 (or replacing A3 by A11-12) might improve the performance of that set. PG5 does in fact get a proof of right identity from A1-8,11-12 in less than two seconds, while it failed to find a proof from A1-8 alone in 288 seconds.

If we go to slightly more difficult (to prove from the basic axioms) theorems such as that if in a group the square of every element is the identity then the group is commutative, we find that even S3 is sorely taxed. This is one of a large class of theorems for which proof-search efficiency is greatly improved by the addition of the logically dependent axioms B21-22 for right identity and right inverse. This phenomenon--that inclusion of dependent axioms does not always detract from proof search efficiency but may be a positive benefit, possibly even a necessity--is one of the more important insights into axiom selection, at least with the type of search algorithms we employ in our programs. The denial of the theorem above is $Rf(xx)e \wedge Rf(ab)c \wedge \overline{R}f(ba)c$ for A1-8 and $Pxxe \wedge Pabc \wedge \overline{P}bac$ for S3. The proof from A1-8,A21-22 is again quite tedious:

1.	$\overline{R}xy\ Ryx$	A1
2.	$\overline{R}xy\ \overline{R}yz\ Rxz$	A2
3.	$Rf(f(xy)z)f(xf(yz))$	A3
4.	$\overline{R}xz\ \overline{R}yu\ Rf(xy)f(zu)$	A5
5.	$Rf(ex)x$	A6
6.	Rxx	A8
7.	$Rf(xe)x$	A21
8.	$Rf(xx)e$	denial of theorem
9.	$Rf(ab)c$	" " "
10.	$\overline{R}f(ba)c$	" " "
11.	$Ref(xx)$	$8\text{-}A1_1$
12.	$\overline{R}zu\ Rf(xz)f(xu)$	$A8\text{-}A5_1$
13.	$Rf(ue)f(uf(xx))$	$11\text{-}12_1$
14.	$Rf(xf(yz))f(f(xy)z)$	$A3\text{-}A1_1$
15.	$\overline{R}f(uf(xx))z\ Rf(ue)z$	$13\text{-}A2_1$
16.	$Rf(ue)f(f(ux)x)$	$14\text{-}15_1$

17.	Ruf(ue)	A21-A1$_1$
18.	$\bar{R}$f(ue)z Ruz	17-A2$_1$
19.	Ruf(f(ux)x)	16-18$_1$
20.	Rf(xu)f(xf(f(uw)w))	19-12$_1$
21.	$\bar{R}$uf(xf(yz)) Ruf(f(xy)z)	14-A2$_2$
22.	Rf(xu)f(f(xf(uw))w)	20-21$_1$
23.	$\bar{R}$zu Rf(zy)f(uy)	A8-A5$_2$
24.	Rf(f(zz)yf(ey)	8-23$_1$
25.	$\bar{R}$xf(f(zz)y) Rxf(ey)	24-A2$_2$
26.	Rf(f(uw)u)f(ew)	22-25$_1$
27.	$\bar{R}$xf(ez) Rxz	A6-A2$_2$
28.	Rf(f(uw)u)w	26-27$_1$
29.	$\bar{R}$xf(ab) Rxc	9-A2$_2$
30.	$\bar{R}$f(ba)f(ab)	10-29$_2$
31.	$\bar{R}$f(ab)f(ba)	30-A1$_2$
32.	$\bar{R}$f(ab)y $\bar{R}$yf(ba)	31-A2$_3$
33.	$\bar{R}$f(f(f(ab)z)z)f(ba)	19-32$_1$
34.	$\bar{R}$f(f(ab)a)b	33-23$_2$
35.	$\square$	34-28

The proof from S3 $\cup$ {B21,B22} is, as before, shorter (still shorter proofs than the one given below have been obtained by the computer):

1.	Pexx	B6
2.	Pxex	B21
3.	Pxg(x)e	B22
4.	$\overline{P}$xyu $\overline{P}$yzt $\overline{P}$uzw Pxtw	B11
5.	$\overline{P}$xyu $\overline{P}$yzt $\overline{P}$xtw Puzw	B12
6.	Pxxe	denial
7.	Pabc	"
8.	$\overline{P}$bac	"
9.	$\overline{P}$xye $\overline{P}$ywt Pytw	$1-4_3$
10.	$\overline{P}$ywt Pytw	$6-9_1$
11.	Pacb	$7-10_1$
12.	$\overline{P}$yzt $\overline{P}$ezw Pytw	$6-4_1$
13.	$\overline{P}$bza $\overline{P}$ezc	$8-12_3$
14.	$\overline{P}$wyu $\overline{P}$yze Puzw	$2-5_3$
15.	$\overline{P}$g(w)ze Pezw	$3-14_1$
16.	Peg(w)w	$6-15_1$
17.	$\overline{P}$bg(c)a	$16-13_2$
18.	$\overline{P}$wyu Pug(y)w	$3-14_2$
19.	Pbg(c)a	$11-18_1$
20.	□	17-19

References

1. Kowalski, R., and Hayes, P. "Semantic trees in automatic theorem proving," *Machine Intelligence IV,* ed. by D. Michie and B. Meltzer (1969).

2. Luckham, D. "Some tree-paring strategies for theorem-proving." *Machine Intelligence III,* ed. by D. Michie, Edinburgh Univ. Press, Edinburgh (1968).

3. Robinson, A. *Model theory and the mathematics of algebra.* North Holland, Amsterdam (1963).

4. Robinson, G., and Wos, L. "Paramodulation and theorem-proving in first-order theories with equality." *Machine Intelligence IV,* ed. by D. Michie and B. Meltzer (1969).

5. Robinson, J. "A machine-oriented logic based on the resolution principle." *J.ACM* 12 (1965), pp. 23-41.

6. Slagle, J. "Automatic theorem proving with renumerable and semantic resolution." *J. ACM* 14 (1967), pp. 687-697.

7. Suppes, P. *Introduction to Logic,* Prentice Hall, New York (1957).

8. Wos, L., Carson, D., and Robinson, G. "The unit preference strategy in theorem proving." *AFIPS Conf. Proc.* 26, Spartan Books, Washington D. C. (1964), pp. 615-621.

9. Wos, L., Robinson, G., and Carson, D. "Efficiency and completeness of the set of support strategy in theorem proving. *J. ACM* 12 (1965), pp. 536-541.

10. Wos, L., Robinson, G., Carson, D., and Shalla, L. "The concept of demodulation in theorem proving." *J. ACM* 14 (1967), pp. 698-709.

Appendix A

Proof that S3,B21-22 A1-8:

	1.	$\overline{P}$exu Rxu	$B6\text{-}B9_1$	(resolution of B6 against first literal of B9)
A8:	2.	Rxx	$B6\text{-}1_1$	(resolution of B6 against first literal of step 1)
	3.	$\overline{R}$zu $\overline{P}$xyz $\overline{P}$xyw Ruw	$B14_3\text{-}B9_1$	(resolution of third literal of B14 against first literal of B9)
	4.	$\overline{R}$zu $\overline{P}$xyz Ruz	$3_2\text{-}3$	(factoring step 3 on second and third literals)
A1:	5.	$\overline{R}$zu Ruz	$B6\text{-}4_2$	
	6.	$\overline{R}$yu Peyu	$B6\text{-}B14_2$	
	7.	$\overline{R}$yz $\overline{R}$zu Peyu	$B14_2\text{-}6_2$	
	8.	$\overline{R}$yz $\overline{R}$zu $\overline{P}$eyw Rwu	$B9_2\text{-}7_3$	
A2:	9.	$\overline{R}$yz $\overline{R}$zu Ryu	$B6\text{-}8_3$	
	10.	$\overline{P}$xyu Rf(xy)u	$B13\text{-}B9_1$	
	11.	$\overline{P}$yzt $\overline{P}$xtw Pf(xy)zw	$B13\text{-}B12_1$	
	12.	$\overline{P}$xf(yz)w Pf(xy)zw	$B13\text{-}11_1$	
	13.	Pf(xy)zf(xf(yz))	$B13\text{-}12_1$	
A3:	14.	Rf(f(xy)z)f(xf(yz))	$13\text{-}10_1$	
	15.	$\overline{R}$yu $\overline{P}$yzt $\overline{P}$etw Puzw	$B12_1\text{-}6_2$	
	16.	$\overline{R}$yu $\overline{P}$yzt Puzt	$B6\text{-}15_3$	
	17.	$\overline{R}$xu Pxeu	$B21\text{-}B14_2$	
	18.	$\overline{R}$xu $\overline{P}$yxz $\overline{P}$zew Pyuw	$B11_2\text{-}17_2$	
	19.	$\overline{R}$xu $\overline{P}$yxz Pyuz	$B21\text{-}18_3$	
	20.	$\overline{R}$uy $\overline{P}$yzt Puzt	$5_2\text{-}16_1$	
	21.	$\overline{R}$ux $\overline{P}$yxz Pyuz	$5_2\text{-}19_1$	
	22.	$\overline{R}$ux $\overline{P}$yxz $\overline{R}$wy Pwuz	$21_3\text{-}20_2$	
	23.	$\overline{R}$wy $\overline{R}$ux Pwuf(yx)	$B13\text{-}22_4$	

A5:	24.	$\overline{R}wy\ \overline{R}ux\ Pf(wu)f(yx)$	23_3-10_1
	25.	$\overline{P}xzt\ \overline{P}ezw\ Pg(x)tw$	B7-$B11_1$
	26.	$\overline{P}xzt\ Pg(x)tz$	B6-25_2
	27.	$\overline{R}tx\ \overline{P}tzu\ Pg(x)uz$	26_1-16_3
	28.	$\overline{R}tx\ Pg(x)eg(t)$	B22-27_2
	29.	$\overline{P}yeu\ Ruy$	B21-$B9_2$
A4:	30.	$\overline{R}tx\ Rg(t)g(x)$	29_1-28_2
A6:	31.	$Rf(ex)x$	B6-10_1
A7:	32.	$Rf(g(x)x)e$	B7-10_1