

*Problem Corner:*

## Reasoning about Equality★

E. L. LUSK and R. A. OVERBEEK  
*Argonne National Laboratory, Argonne, IL 60439, U.S.A.*

(Received: 4 March 1985)

**Abstract.** This note contains a set of six theorems that can be used to assess the ability of a theorem-proving system to reason about equality. The six theorems are graduated in terms of difficulty: they range from fairly trivial to quite difficult. They do not cover all aspects of equality reasoning, but they have proved useful to us in developing our system.

### Introduction

In this note we present some examples of reasoning about equality using an automated reasoning system. The two primary techniques used are the inference rule of paramodulation and the term rewriting technique called demodulation. Both of these were introduced in the late 1960s [8, 3, 7]. In addition the paper by Knuth and Bendix in 1970 [1] started a long chain of research on manipulation of a set of rewrite rules by paramodulation into what is called a ‘complete set of reductions.’ Here we will not discuss equality reasoning in that particular setting, although it is quite important, preferring to postpone that topic to another column. We include here the complete details of the proofs, so that someone experimenting with his own system can track precisely at least one path to each proof.

There are many variations on demodulation and paramodulation, and we will consider here only some the most straightforward ones. *Demodulation* is the rewriting of terms. Equality clauses represent candidates for promotion to the status of rewrite rules. Our particular system [2] makes all equality clauses demodulators, using a weighting scheme to decide whether the demodulator should be used as a left-to-right rewrite rule, a right-to-left rewrite rule, or a ‘lexically dependent’ rewrite rule. Lexically dependent demodulators are described in detail in Chapter 8 of [6]. Demodulation is defined as follows: Given a clause  $C$  containing a term  $t$ , and a unit equality clause of the form  $\alpha = \beta$ , where  $t$  is an instance of  $\alpha$  ( $t = \alpha\sigma$ ), replace  $C$  by  $C'$ , obtained from  $C$  by replacing all occurrences of  $t$  by  $t'$  where  $t' = \beta\sigma$ . For example, given  $P(f(a, a))$  and demodulator  $f(x, x) = g(x)$ , the demodulation process causes us to add  $P(g(a))$  and delete  $P(f(a, a))$ . *Paramodulation* is a more general mechanism. Given a clause  $C$  containing a term  $t$  and a clause  $D$  containing an equality literal  $\alpha = \beta$ , where  $t$  unifies with  $\alpha$  with substitution  $\sigma$ ,

★ This work was supported in part by National Science Foundation grant MCS82-07496 and in part by the Applied Mathematical Sciences Research Program (KC-04-02) of the Office of Energy Research of the U.S. Department of Energy under Contract W-31-109-Eng-38.

we derive clause  $C'$ , which is  $C\sigma$  with  $t$  replaced by  $\beta\sigma$  (and other literals from  $D\sigma$  added). For example,

$$\begin{array}{ll} D: \text{sum}(0, x) = x & \alpha = \text{sum}(0, x); \beta = x \\ C: \text{sum}(y, \text{minus}(y)) = 0 & t = \text{sum}(y, \text{minus}(y)) \\ C': \text{minus}(0) = 0 & \sigma: (y \leftarrow 0, x \leftarrow \text{minus}(0)) \end{array}$$

Paramodulation is different from demodulation in that the unification goes both ways, the equality literal need not be in a unit clause, and both parents are ordinarily kept.

This paper contains six problems that we have used as benchmarks during the development of our theorem-proving systems. We have arranged the problems into an order that reflects their relative difficulty. They are as follows:

*Problem 1:* In a group, if  $x^2 = e$  for all  $x$  in the group, then the group is Abelian (for all  $x$  and  $y$ ,  $xy = yx$ ).

*Problem 2:* In a group,  $(x^{-1})^{-1} = x$  for all  $x$  in the group.

*Problem 3:* In a ring, if  $x^2 = x$  for all  $x$  in the ring, then  $xy = yx$  for all  $x, y$  in the ring.

*Problem 4:* In a group, if  $x^3 = e$  for all  $x$  in the group, then the commutator  $h(h(x, y), y) = e$  for all  $x$  and  $y$ . The commutator  $h(x, y)$  is defined as  $xyx^{-1}y^{-1}$ .

*Problem 5:* In a ternary Boolean algebra with the third axiom removed, it is true that  $f(x, g(x), y) = y$  for all  $x$  and  $y$ .

*Problem 6:* In a ring, if  $x^3 = x$  for all  $x$  in the ring, then  $xy = yx$  for all  $x$  and  $y$  in the ring.

The first problem is a classic in the theorem proving literature. It is normally used as an initial test to verify that an equality reasoning component is functioning properly.

The second problem is also quite simple, and should be easily solved by any system that includes substitution and simplification capabilities.

The third problem introduces the axioms for a ring. It is of moderate difficulty.

The fourth problem, also a classic, is substantially more difficult than the first two problems. It was included in one of the papers that introduced paramodulation [3].

The fifth problem involves ternary Boolean algebras, a rather obscure area in mathematics. Our system attained a proof of this problem by using 'noncomplexifying paramodulation'. In this restriction of paramodulation, variables that occur both in the *into* term and outside the *into* term can be instantiated only to other variables or to constants (variables in the *from* term can be arbitrarily instantiated). We have found noncomplexifying paramodulation useful in other proofs, as well; however, no comprehensive study has been made of its general utility.

The sixth problem is truly difficult for existing systems. Two researchers have reported on approaches that resulted in proofs [5, 4].

**Problem 1**

*Problem 1:* In a group, if  $x^2 = e$  for all  $x$  in the group, then the group is Abelian (for all  $x$  and  $y$ ,  $xy = yx$ )

Axioms for a group:

- |   |                                 |                                  |
|---|---------------------------------|----------------------------------|
| 1 | $f(e, x) = x$                   | $e$ is a left identity           |
| 2 | $f(x, e) = x$                   | $e$ is a right identity          |
| 3 | $f(g(x), x) = e$                | there exists a left inverse      |
| 4 | $f(x, g(x)) = e$                | there exists a right inverse     |
| 5 | $f(f(x, y), z) = f(x, f(y, z))$ | associativity                    |
| 6 | $x = x$                         | reflexivity of equality          |
| 7 | $f(x, x) = e$                   | $x * x = e$ (special hypothesis) |
| 8 | $\neg(f(a, b) = f(b, a))$       | denial of the theorem            |

*Proof:*

- |    |                           |           |
|----|---------------------------|-----------|
| 9  | $f(x, f(y, f(x, y))) = e$ | 7 5 5 5 5 |
| 10 | $x = f(y, f(y, x))$       | 7 5 1     |
| 11 | $f(x, f(y, x)) = y$       | 9 10 2    |
| 12 | $f(x, y) = f(y, x)$       | 11 10     |
| 13 | null                      | 12 8      |

*Details of Proof:*

$$9 \quad f(x, f(y, f(x, y))) = e$$

The clause was deduced by performing the following operations:

Paramodulate into clause 7 using clause 5  
*into-term:*  $f(x, x)$  *from-term:*  $f(x, f(y, z))$   
 result:  $f(f(f(y, z), y), z) = e$

Demodulate into the result using clause 5  
*into-term:*  $f(f(y, z), y)$  *from-term:*  $f(f(x, y), z)$   
 result:  $f(f(y, f(z, y)), z) = e$

Demodulate into the result using clause 5  
*into-term:*  $f(f(y, f(z, y)), z)$  *from-term:*  $f(f(x, y), z)$   
 result:  $f(y, f(f(z, y), z)) = e$

Demodulate into the result using clause 5  
*into-term:*  $f(f(z, y), z)$  *from-term:*  $f(f(x, y), z)$

$$10 \quad x = f(y, f(y, x))$$

The clause was deduced by performing the following operations:

Paramodulate from clause 7 into clause 5  
*from-term:*  $f(x, x)$  *into-term:*  $f(x, y)$  in  $f(f(x, y), z)$   
 result:  $f(e, z) = f(x, f(x, z))$

Demodulate into the result using clause 1

*into-term:*  $f(e, z)$  *from-term:*  $f(e, x)$

$$11 \quad f(x, f(y, x)) = y$$

The clause was deduced by performing the following operations:

Paramodulate from clause 9 into clause 10

*from-term:*  $f(x, f(y, f(x, y)))$  *into-term:*  $f(y, x)$

result:  $f(y, f(x, y)) = f(x, e)$

Demodulate into the result using clause 2

*into-term:*  $f(x, e)$  *from-term:*  $f(x, e)$

$$12 \quad f(x, y) = f(y, x)$$

The clause was deduced by performing the following operations:

Paramodulate from clause 11 into clause 10

*from-term:*  $f(x, f(y, x))$  *into-term:*  $f(y, x)$

## Problem 2

*Problem 2:* In a group,  $(x^{-1})^{-1} = x$  for all  $x$  in the group.

Axioms for a group:

- |   |                                 |                              |
|---|---------------------------------|------------------------------|
| 1 | $f(e, x) = x$                   | $e$ is a left identity       |
| 2 | $f(x, e) = x$                   | $e$ is a right identity      |
| 3 | $f(g(x), x) = e$                | there exists a left inverse  |
| 4 | $f(x, g(x)) = e$                | there exists a right inverse |
| 5 | $f(f(x, y), z) = f(x, f(y, z))$ | associativity                |
| 6 | $x = x$                         | reflexivity of equality      |
| 7 | $\neg (g(g(a)) = a)$            | denial of the theorem        |

*Proof:*

- |    |                        |       |
|----|------------------------|-------|
| 8  | $z = f(x, f(g(x), z))$ | 5 4 1 |
| 9  | $g(g(x)) = x$          | 8 4 2 |
| 10 | null                   | 9 7   |

*Details of Proof:*

$$8 \quad z = f(x, f(g(x), z))$$

The clause was deduced by performing the following operations:

Paramodulate into clause 5 using clause 4

*into-term:*  $f(x, y)$  *from-term:*  $f(x, g(x))$

result:  $f(e, z) = f(x, f(g(x), z))$

Demodulate into the result using clause 1

*into-term:*  $f(e, z)$  *from-term:*  $f(e, x)$

$$9 \quad g(g(x)) = x$$

The clause was deduced by performing the following operations:

- Paramdoulate from clause 4 into clause 8  
*from-term:*  $f(x, g(x))$  *into-term:*  $f(g(x), z)$   
*result:*  $g(g(x)) = f(x, e)$
- Demodulate into the result using clause 2  
*into-term:*  $f(x, e)$  *from-term:*  $f(x, e)$

**Problem 3**

*Problem 3:* In a ring, if  $x^2 = x$  for all  $x$  in the ring, then  $xy = yx$  for all  $x, y$  in the ring.

Axioms for a ring:

- |    |                                       |                                      |
|----|---------------------------------------|--------------------------------------|
| 1  | $j(0, x) = x$                         | 0 is a left identity for sum         |
| 2  | $j(x, 0) = x$                         | 0 is a right identity for sum        |
| 3  | $j(g(x), x) = 0$                      | there exists a left inverse for sum  |
| 4  | $j(x, g(x)) = 0$                      | there exists a right inverse for sum |
| 5  | $j(j(x, y), z) = j(x, j(y, z))$       | associativity of addition            |
| 6  | $x = x$                               | reflexivity of equality              |
| 7  | $j(x, y) = j(y, x)$                   | commutativity of addition            |
| 8  | $f(f(x, y), z) = f(x, f(y, z))$       | associativity of multiplication      |
| 9  | $f(x, j(y, z)) = j(f(x, y), f(x, z))$ | distributivity axioms                |
| 10 | $f(j(y, z), x) = j(f(y, x), f(z, x))$ |                                      |
| 11 | $f(x, x) = x$                         | $x * x = x$ (special hypothesis)     |
| 12 | $-(f(a, b) = f(b, a))$                | denial of the theorem                |

*Proof:*

- |    |   |                     |
|----|---|---------------------|
| 13 | $f(x, j(x, y)) = j(x, f(x, y))$                   | 11 9                |
| 14 | $f(x, j(x, x)) = j(x, x)$                         | 11 13               |
| 15 | $j(x, y) = j(f(x, j(x, j(x, y))), f(y, j(x, y)))$ | 11 10               |
| 16 | $j(x, x) = j(j(x, x), j(x, x))$                   | 14 15 14            |
| 17 | $j(j(x, x), y) = j(j(x, x), j(j(x, x), y))$       | 16 5                |
| 18 | $0 = j(x, x)$                                     | 17 4 4 2            |
| 19 | $j(x, j(y, z)) = j(y, j(x, z))$                   | 7 5 5               |
| 20 | $j(x, y) = j(x, j(y, j(f(x, y), f(y, x))))$       | 15 9 11 9 11 7 5 19 |
| 21 | $y = j(g(x), j(x, y))$                            | 3 5 1               |
| 22 | $j(x, j(f(y, x), f(x, y))) = x$                   | 21 20 21            |
| 23 | $j(f(x, y), f(y, x)) = 0$                         | 21 22 3             |
| 24 | $x = j(y, j(y, x))$                               | 5 18 1              |
| 25 | $f(x, y) = f(y, x)$                               | 24 23 2             |
| 26 | null  | 25 12               |

*Details of Proof:*

$$13 \quad f(x, j(x, y)) = j(x, f(x, y))$$

The clause was deduced by performing the following operations:

Paramodulate into clause 9 using clause 11

$$\text{into-term: } f(x, y) \quad \text{from-term: } f(x, x)$$

$$14 \quad f(x, j(x, x)) = j(x, x)$$

The clause was deduced by performing the following operations:

Paramodulate into clause 13 using clause 11

$$\text{into-term: } f(x, y) \quad \text{from-term: } f(x, x)$$

$$15 \quad j(x, y) = j(f(x, j(x, y)), f(y, j(x, y)))$$

The clause was deduced by performing the following operations:

Paramodulate into clause 10 using clause 11

$$\text{into-term: } f(j(y, z), x) \quad \text{from-term: } f(x, x)$$

$$16 \quad j(x, x) = j(j(x, x), j(x, x))$$

The clause was deduced by performing the following operations:

Paramodulate into clause 15 using clause 14

$$\text{into-term: } f(x, j(x, y)) \quad \text{from-term: } f(x, j(x, x))$$

$$\text{result: } j(x, x) = j(j(x, x), f(x, j(x, x)))$$

Demodulate into the result using clause 14

$$\text{into-term: } f(x, j(x, x)) \quad \text{from-term: } f(x, j(x, x))$$

$$17 \quad j(j(x, x), y) = j(j(x, x), j(j(x, x), y))$$

The clause was deduced by performing the following operations:

Paramodulate into clause 5 using clause 16

$$\text{into-term: } j(x, y) \quad \text{from-term: } j(j(x, x), j(x, x))$$

$$18 \quad 0 = j(x, x)$$

The clause was deduced by performing the following operations:

Paramodulate into clause 17 using clause 4

$$\text{into-term: } j(j(x, x), y) \quad \text{from-term: } j(x, g(x))$$

$$\text{result: } 0 = j(j(x, x), j(j(x, x), g(j(x, x))))$$

Demodulate into the result using clause 4

$$\text{into-term: } j(j(x, x), g(j(x, x))) \quad \text{from-term: } j(x, g(x))$$

$$\text{result: } 0 = j(j(x, x), 0)$$

Demodulate into the result using clause 2

$$\text{into-term: } j(j(x, x), 0) \quad \text{from-term: } j(x, 0)$$

$$19 \quad j(x, j(y, z)) = j(y, j(x, z))$$

The clause was deduced by performing the following operations:

Paramodulate into clause 5 using clause 7

*into-term:*  $j(j(x, y), z)$  *from-term:*  $j(x, y)$   
*result:*  $j(j(y, x), z) = j(x, j(y, z))$

Demodulate into the result using clause 5

*into-term:*  $j(j(y, x), z)$  *from-term:*  $j(j(x, y), z)$   
 20  $j(x, y) = j(x, j(y, j(f(x, y), f(y, x))))$

The clause was deduced by performing the following operations:

Paramodulate into clause 15 using clause 9

*into-term:*  $f(x, j(x, y))$  *from-term:*  $f(x, j(y, z))$   
*result:*  $j(x, y) = j(j(f(x, x), f(x, y)), f(y, j(x, y)))$

Demodulate into the result using clause 11

*into-term:*  $f(x, x)$  *from-term:*  $f(x, x)$   
*result:*  $j(x, y) = j(j(x, f(x, y)), f(y, j(x, y)))$

Demodulate into the result using clause 9

*into-term:*  $f(y, j(x, y))$  *from-term:*  $f(x, j(y, z))$   
*result:*  $j(x, y) = j(j(x, f(x, y)), j(f(y, x), f(y, y)))$

Demodulate into the result using clause 11

*into-term:*  $f(y, y)$  *from-term:*  $f(x, x)$   
*result:*  $j(x, y) = j(j(x, f(x, y)), j(f(y, x), y))$

Demodulate into the result using clause 7

*into-term:*  $j(f(y, x), y)$  *from-term:*  $j(x, y)$   
*result:*  $j(x, y) = j(j(x, f(x, y)), j(y, f(y, x)))$

Demodulate into the result using clause 5

*into-term:*  $j(j(x, f(x, y)), j(y, f(y, x)))$  *from-term:*  $j(j(x, y), z)$   
*result:*  $j(x, y) = j(x, j(f(x, y), j(y, f(y, x))))$

Demodulate into the result using clause 19

*into-term:*  $j(f(x, y), j(y, f(y, x)))$  *from-term:*  $j(x, j(y, z))$   
 21  $x = j(g(y), j(y, x))$

The clause was deduced by performing the following operations:

Paramodulate into clause 5 using clause 3

*into-term:*  $j(x, y)$  *from-term:*  $j(g(x), x)$   
*result:*  $j(0, z) = j(g(x), j(x, z))$

Demodulate into the result using clause 1

*into-term:*  $j(0, z)$  *from-term:*  $j(0, x)$

22  $j(x, j(f(y, x), f(x, y))) = x$

The clause was deduced by performing the following operations:

Paramodulate into clause 21 using clause 20

*into-term:*  $j(y, x)$  *from-term:*  $j(x, j(y, j(f(x, y), f(y, x))))$   
*result:*  $j(y, j(f(x, y), f(y, x))) = j(g(x), j(x, y))$

Demodulate into the result using clause 21

*into-term:*  $j(g(x), j(x, y))$  *from-term:*  $j(g(y), j(y, x))$

$$23 \quad j(f(x, y), f(y, x)) = 0$$

The clause was deduced by performing the following operations:

Paramodulate into clause 21 using clause 22

$$\text{into-term: } j(y, x) \quad \text{from-term: } j(x, j(f(y, x), f(x, y)))$$

$$\text{result: } j(f(y, x), f(x, y)) = j(g(x), x)$$

Demodulate into the result using clause 3

$$\text{into-term: } j(g(x), x) \quad \text{from-term: } j(g(x), x)$$

$$24 \quad x = j(y, j(y, x))$$

The clause was deduced by performing the following operations:

Paramodulate into clause 5 using clause 18

$$\text{into-term: } j(x, y) \quad \text{from-term: } j(x, x)$$

$$\text{result: } j(0, 2) = j(x, j(x, z))$$

Demodulate into the result using clause 1

$$\text{into-term: } j(0, z) \quad \text{from-term: } j(0, x)$$

$$25 \quad f(x, y) = f(y, x)$$

The clause was deduced by performing the following operations:

Paramodulate into clause 24 using clause 23

$$\text{into-term: } j(y, x) \quad \text{from-term: } j(f(x, y), f(y, x))$$

$$\text{result: } f(y, x) = j(f(x, y), 0)$$

Demodulate into the result using clause 2

$$\text{into-term: } j(f(x, y), 0) \quad \text{from-term: } j(x, 0)$$

## Problem 4

*Problem 4:* In a group, if  $x^3 = e$  for all  $x$  in the group, then the commutator  $h(h(x, y), y) = e$  for all  $x$  and  $y$ . The commutator  $h(x, y)$  is defined as  $xyx^{-1}y^{-1}$ .

Axioms for a group:

- |   |                                       |                                      |
|---|---------------------------------------|--------------------------------------|
| 1 | $f(e, x) = x$                         | $e$ is a left identity               |
| 2 | $f(x, e) = x$                         | $e$ is a right identity              |
| 3 | $f(g(x), x) = e$                      | there exists a left inverse          |
| 4 | $f(x, g(x)) = e$                      | there exists a right inverse         |
| 5 | $f(f(x, y), z) = f(x, f(y, z))$       | associativity                        |
| 6 | $x = x$                               | reflexivity of equality              |
| 7 | $h(x, y) = f(x, f(y, f(g(x), g(y))))$ | definition of commutator             |
| 8 | $f(x, f(x, x)) = e$                   | $x * x * x = e$ (special hypothesis) |
| 9 | $-(h(h(a, b), b) = e)$                | denial of the theorem                |

*Proof:*

$$10 \quad g(e) = e$$

3 2

11	$-(f(a, f(b, f(g(a), f(g(b), f(b, f(g(f(a, f(b, f(g(a), g(b))))), g(b))))), g(b)))))) = e$	9 7 7 5 5 5
12	$x = f(y, f(g(y), x))$	5 4 1
13	$x = f(g(y), f(y, x))$	5 3 1
14	$e = f(x, f(y, g(f(x, y))))$	5 4
15	$-(f(a, f(b, f(g(a), f(g(f(a, f(b, f(g(a), g(b))))), g(b))))), g(b)))) = e$	13 11
16	$g(g(x)) = x$	12 4 2
17	$f(g(x), g(x)) = x$	12 8 2
18	$f(x, x) = g(x)$	17 16 16
19	$f(x, f(y, f(x, y))) = g(f(x, y))$	18 5
20	$f(g(x), y) = f(x, f(x, y))$	18 5
21	$f(x, f(g(y), x)) = f(y, g(f(g(y), x)))$	19 12
22	$f(x, f(y, x)) = f(g(y), g(f(y, x)))$	19 13
23	$f(x, f(g(y), f(x, z))) = f(y, f(g(f(g(y), x)), z))$	21 5 5 5
24	$f(x, f(y, f(x, z))) = f(g(y), f(g(f(y, x)), z))$	22 5 5 5
25	$-(f(a, f(g(b), f(g(a), f(g(b), f(a, f(b, f(g(a), b))))), g(b)))) = e$	24 15 5 5 18 16 5 5 20
26	$f(x, g(f(y, x))) = g(y)$	14 22 10 2 14 2
27	$g(f(x, y)) = f(g(y), g(x))$	26 13
28	$f(x, f(y, f(x, y))) = f(g(y), g(x))$	27 19
29	$f(x, f(g(y), f(x, z))) = f(y, f(g(x), f(y, z)))$	27 23 16 5
30	$-(e = e)$	29 25 20 28 16 16 12 8
31	null	30 6

*Details of Proof:*

$$10 \quad g(e) = e$$

The clause was deduced by performing the following operations:

Paramodulate into clause 3 using clause 2

*into-term:*  $f(g(x), x)$     *from-term:*  $f(x, e)$

$$11 \quad -(f(a, f(b, f(g(a), f(g(b), f(b, f(g(f(a, f(b, f(g(a), g(b))))), g(b))))), g(b)))) = e$$

The clause was deduced by performing the following operations:

Paramodulate into clause 9 from clause 7

*from-term:*  $h(x, y)$     *into-term:*  $h(a, b)$

result:  $-(h(f(a, f(b, f(g(a), g(b))))), b) = e$

Demodulate into the result using clause 7

*into-term:*  $h(f(a, f(b, f(g(a), g(b))))), b)$     *from-term:*  $h(x, y)$

result:  $-(f(f(a, f(b, f(g(a), g(b))))), f(b, f(g(f(a, f(b, f(g(a), g(b))))), g(b)))) = e$

Demodulate into the result using clause 5

*into-term*:  $f(f(a, f(b, f(g(a), g(b))))), f(b, f(g(f(a, f(b, f(g(a), g(b))))), g(b))))$

*from-term*:  $f(f(x, f(y, z)))$

result:  $-(f(a, f(b, f(g(a), g(b))), f(b, f(g(f(a, f(b, f(g(a), g(b))))), g(b)))) = e$

Demodulate into the result using clause 5

*into-term*:  $f(f(b, f(g(a), g(b))), f(b, f(g(f(a, f(b, f(g(a), g(b))))), g(b))))$

*from-term*:  $f(f(x, f(y, z)))$

result:  $-(f(a, f(b, f(f(g(a), g(b))), f(b, f(g(f(a, f(b, f(g(a), g(b))))), g(b)))) = e$

Demodulate into the result using clause 5

*into-term*:  $f(f(g(a), g(b)), f(b, f(g(f(a, f(b, f(g(a), g(b))))), g(b))))$

*from-term*:  $f(f(x, f(y, z)))$

12  $x = f(y, f(g(y), x))$

The clause was deduced by performing the following operations:

Paramodulate into clause 5 from clause 4

*into-term*:  $f(x, y)$  *from-term*:  $f(x, g(x))$

result:  $f(e, z) = f(x, f(g(x), z))$

Demodulate into the result using clause 1

*into-term*:  $f(e, z)$  *from-term*:  $f(e, x)$

13  $x = f(g(y), f(y, x))$

The clause was deduced by performing the following operations:

Paramodulate into clause 5 from clause 3

*into-term*:  $f(x, y)$  *from-term*:  $f(g(x), x)$

result:  $f(e, z) = f(g(x), f(x, z))$

Demodulate into the result using clause 1

*into-term*:  $f(e, z)$  *from-term*:  $f(e, x)$

14  $e = f(x, f(y, g(f(x, y))))$

The clause was deduced by performing the following operations:

Paramodulate into clause 5 from clause 4

*into-term*:  $f(f(x, y), z)$  *from-term*:  $f(x, g(x))$

result:  $e = f(x, f(y, g(f(x, y))))$

15  $-(f(a, f(b, f(g(a), f(g(f(a, f(b, f(g(a), g(b))))), g(b)))) = e$

The clause was deduced by performing the following operations:

Paramodulate from clause 13 into clause 11

*from-term*:  $f(g(y), f(y, x))$

*into-term*:  $f(g(b), f(b, f(g(f(a, f(b, f(g(a), g(b))))), g(b))))$

$$16 \quad g(g(x)) = x$$

The clause was deduced by performing the following operations:

Paramodulate into clause 12 from clause 4

$$\text{into-term: } f(g(y), x) \quad \text{from-term: } f(x, g(x))$$

$$\text{result: } g(g(x)) = f(x, e)$$

Demodulate into the result using clause 2

$$\text{into-term: } f(x, e) \quad \text{from-term: } f(x, e)$$

$$17 \quad f(g(x), g(x)) = x$$

The clause was deduced by performing the following operations:

Paramodulate into clause 12 from clause 8

$$\text{into-term: } f(g(y), x) \quad \text{from-term: } f(x, f(x, x))$$

$$\text{result: } f(g(y), g(y)) = f(y, e)$$

Demodulate into the result using clause 2

$$\text{into-term: } f(y, e) \quad \text{from-term: } f(x, e)$$

$$18 \quad f(x, x) = g(x)$$

The clause was deduced by performing the following operations:

Paramodulate into clause 17 from clause 16

$$\text{into-term: } g(x) \quad \text{from-term: } g(g(x))$$

$$\text{result: } f(y, g(g(y))) = g(y)$$

Demodulate into the result using clause 16

$$\text{into-term: } g(g(y)) \quad \text{from-term: } g(g(x))$$

$$19 \quad f(x, f(y, f(x, y))) = g(f(x, y))$$

The clause was deduced by performing the following operations:

Paramodulate into clause 18 from clause 5

$$\text{into-term: } f(x, x) \quad \text{from-term: } f(f(x, y), z)$$

$$20 \quad f(g(x), y) = f(x, f(x, y))$$

The clause was deduced by performing the following operations:

Paramodulate from clause 18 into clause 5

$$\text{from-term: } f(x, x) \quad \text{into-term: } f(x, y)$$

$$21 \quad f(x, f(g(y), x)) = f(y, g(f(g(y), x)))$$

The clause was deduced by performing the following operations:

Paramodulate from clause 19 into clause 12

$$\text{from-term: } f(x, f(y, f(x, y))) \quad \text{into-term: } f(g(y), x)$$

$$22 \quad f(x, f(y, x)) = f(g(y), g(f(y, x)))$$

The clause was deduced by performing the following operations:

Paramodulate from clause 19 into clause 13

$$\text{from-term: } f(x, f(y, f(x, y))) \quad \text{into-term: } f(y, x)$$

$$23 \quad f(x, f(g(y), f(x, z))) = f(y, f(g(f(g(y), x)), z))$$

The clause was deduced by performing the following operations:

Paramodulate from clause 21 into clause 5

$$\begin{aligned} \text{from-term: } & f(y, g(f(g(y), x))) \quad \text{into-term: } f(x, y) \\ \text{result: } & f(f(x, f(g(y), x)), z) = f(y, f(g(f(g(y), x)), z)) \end{aligned}$$

Demodulate into the result using clause 5

$$\begin{aligned} \text{into-term: } & f(f(x, f(g(y), x)), z) \quad \text{from-term: } f(f(x, y), z) \\ \text{result: } & f(x, f(f(g(y), x), z)) = f(y, f(g(f(g(y), x)), z)) \end{aligned}$$

Demodulate into the result using clause 5

$$\text{into-term: } f(f(g(y), x), z) \quad \text{from-term: } f(f(x, y), z)$$

$$24 \quad f(x, f(y, f(x, z))) = f(g(y), f(g(f(y), x)), z)$$

The clause was deduced by performing the following operations:

Paramodulate from clause 22 into clause 5

$$\begin{aligned} \text{from-term: } & f(g(y), g(f(y, x))) \quad \text{into-term: } f(x, y) \\ \text{result: } & f(f(x, f(y, x)), z) = f(g(y), f(g(f(y), x)), z) \end{aligned}$$

Demodulate into the result using clause 5

$$\begin{aligned} \text{into-term: } & f(f(x, f(y, x)), z) \quad \text{from-term: } f(f(x, y), z) \\ \text{result: } & f(x, f(f(y, x), z)) = f(g(y), f(g(f(y), x)), z) \end{aligned}$$

Demodulate into the result using clause 5

$$\text{into-term: } f(f(y, x), z) \quad \text{from-term: } f(f(x, y), z)$$

$$25 \quad - (f(a, f(g(b), f(g(a), f(g(b), f(a, f(b, f(g(a), b))))))) = e$$

The clause was deduced by performing the following operations:

Paramodulate from clause 24 into clause 15

$$\begin{aligned} \text{from-term: } & f(g(y), f(g(f(y), x)), z) \\ \text{into-term: } & f(g(a), f(g(f(a, f(b, f(g(a), g(b))))), g(b))) \\ \text{result: } & - (f(a, f(b, f(f(b, f(g(a), g(b))), f(a, f(b, f(g(a), g(b))), \\ & \quad g(b)))))) = e \end{aligned}$$

Demodulate into the result using clause 5

$$\begin{aligned} \text{into-term: } & f(f(b, f(g(a), g(b))), g(b)) \quad \text{from-term: } f(f(x, y), z) \\ \text{result: } & - (f(a, f(b, f(f(b, f(g(a), g(b))), f(a, f(b, f(f(g(a), g(b)), \\ & \quad g(b)))))) = e \end{aligned}$$

Demodulate into the result using clause 5

$$\begin{aligned} \text{into-term: } & f(f(g(a), g(b)), g(b)) \quad \text{from-term: } f(f(x, y), z) \\ \text{result: } & - (f(a, f(b, f(f(b, f(g(a), g(b))), f(a, f(b, f(g(a), f(g(b), \\ & \quad g(b)))))) = e \end{aligned}$$

Demodulate into the result using clause 18

$$\begin{aligned} \text{into-term: } & f(g(b), g(b)) \quad \text{from-term: } f(x, x) \\ \text{result: } & - (f(a, f(b, f(f(b, f(g(a), g(b))), f(a, f(b, f(g(a), \\ & \quad g(g(b)))))) = e \end{aligned}$$

Demodulate into the result using clause 16

$$\text{into-term: } g(g(b)) \quad \text{from-term: } g(g(x))$$

$$\text{result: } - (f(a, f(b, f(f(b, f(g(a), g(b))), f(a, f(b, f(g(a), b)))))) = e)$$

Demodulate into the result using clause 5

$$\text{into-term: } f(f(b, f(g(a), g(b))), f(a, f(b, f(g(a), b))))$$

$$\text{from-term: } f(f(x, y), z)$$

$$\text{result: } - (f(a, f(b, f(b, f(f(g(a), g(b))), f(a, f(b, f(g(a), b)))))) = e)$$

Demodulate into the result using clause 5

$$\text{into-term: } f(f(g(a), g(b)), f(a, f(b, f(g(a), b))))$$

$$\text{from-term: } f(f(x, y), z)$$

$$\text{result: } - (f(a, f(b, f(b, f(g(a), f(g(b), f(a, f(b, f(g(a), b)))))) = e)$$

Demodulate into the result using clause 20

$$\text{into-term: } f(b, f(b, f(g(a), f(g(b), f(a, f(b, f(g(a), b))))))$$

$$\text{from-term: } f(x, f(x, y))$$

$$26 \quad f(x, g(f(y, x))) = g(y)$$

The clause was deduced by performing the following operations:

Paramodulate from clause 14 into clause 22

$$\text{from-term: } f(x, f(y, g(f(x, y)))) \quad \text{into-term: } f(y, x)$$

$$\text{result: } f(f(y, g(f(x, y))), f(x, f(y, g(f(x, y)))) = f(g(x), g(e))$$

Demodulate into the result using clause 10

$$\text{from-term: } g(e) \quad \text{into-term: } g(e)$$

$$\text{result: } f(f(y, g(f(x, y))), f(x, f(y, g(f(x, y)))) = f(g(x), e)$$

Demodulate into the result using clause 2

$$\text{from-term: } f(x, e) \quad \text{into-term: } f(g(x), e)$$

$$\text{result: } f(f(y, g(f(x, y))), f(x, f(y, g(f(x, y)))) = g(x)$$

Demodulate into the result using clause 14

$$\text{from-term: } f(x, f(y, g(f(x, y))))$$

$$\text{into-term: } f(x, f(y, g(f(x, y))))$$

$$\text{result: } f(f(y, g(f(x, y))), e) = g(x)$$

Demodulate into the result using clause 2

$$\text{from-term: } f(x, e) \quad \text{into-term: } f(f(y, g(f(x, y))), e)$$

$$27 \quad g(f(x, y)) = f(g(y), g(x))$$

The clause was deduced by performing the following operations:

Paramodulate from clause 26 into clause 13

$$\text{from-term: } f(x, g(f(y, x))) \quad \text{into-term: } f(y, x)$$

$$28 \quad f(x, f(y, f(x, y))) = f(g(y), g(x))$$

The clause was deduced by performing the following operations:

Paramodulate from clause 27 into clause 19

$$\text{from-term: } g(f(x, y)) \quad \text{into-term: } g(f(x, y))$$

$$29 \quad f(x, f(g(y), f(x, z))) = f(y, f(g(x), f(y, z)))$$

The clause was deduced by performing the following operations:

Paramodulate from clause 27 into clause 23

$$\begin{aligned} \text{from-term: } & g(f(x, y)) \quad \text{into-term: } g(f(g(y), x)) \\ \text{result: } & f(x, f(g(y), f(x, z))) = f(y, f(f(g(x), g(g(y))), z)) \end{aligned}$$

Demodulate into the result using clause 16

$$\begin{aligned} \text{from-term: } & g(g(x)) \quad \text{into-term: } g(g(y)) \\ \text{result: } & f(x, f(g(y), f(x, z))) = f(y, f(f(g(x), y), z)) \end{aligned}$$

Demodulate into the result using clause 5

$$\text{into-term: } f(f(g(x), y), z) \quad \text{from-term: } f(f(x, y), z)$$

$$30 - (e = e)$$

The clause was deduced by performing the following operations:

Paramodulate from clause 29 into clause 25

$$\begin{aligned} \text{from-term: } & f(x, f(g(y), f(x, z))) \\ \text{into-term: } & f(g(b), f(g(a), f(g(b), f(a, f(b, f(g(a), b)))))) \\ \text{result: } & - (f(a, f(a, f(g(g(b))), f(a, f(a, f(b, f(g(a), b)))))) = e \end{aligned}$$

Demodulate into the result using clause 20

$$\begin{aligned} \text{from-term: } & f(x, f(x, y)) \quad \text{into-term: } f(a, f(a, f(b, f(g(a), b)))) \\ \text{result: } & - (f(a, f(a, f(g(g(b))), f(g(a), f(b, f(g(a), b)))))) = e \end{aligned}$$

Demodulate into the result using clause 28

$$\begin{aligned} \text{from-term: } & f(x, f(y, f(x, y))) \quad \text{into-term: } f(g(a), f(b, f(g(a), b))) \\ \text{result: } & - (f(a, f(a, f(g(g(b))), f(g(b), g(g(a)))))) = e \end{aligned}$$

Demodulate into the result using clause 16

$$\begin{aligned} \text{from-term: } & g(g(x)) \quad \text{into-term: } g(g(a)) \\ \text{result: } & - (f(a, f(a, f(g(g(b))), f(g(b), a))) = e \end{aligned}$$

Demodulate into the result using clause 16

$$\begin{aligned} \text{from-term: } & g(g(x)) \quad \text{into-term: } g(g(b)) \\ \text{result: } & - (f(a, f(a, f(b, f(g(b), a)))) = e \end{aligned}$$

Demodulate into the result using clause 12

$$\begin{aligned} \text{from-term: } & f(y, f(g(y), x)) \quad \text{into-term: } f(b, f(g(b), a)) \\ \text{result: } & - (f(a, f(a, a)) = e \end{aligned}$$

Demodulate into the result using clause 8

$$\text{from-term: } f(x, f(x, x)) \quad \text{into-term: } f(a, f(a, a))$$

## Problem 5

*Problem 5:* In a ternary Boolean algebra with the third axiom removed, it is true that  $f(x, g(x), y) = y$  for all  $x$  and  $y$ .

Ternary Boolean algebras were defined in by A. A. Grau in 1947 (see 'Ternary Boolean algebra', *Bull. Amer. Math. Soc.* **53**, (6) June 1947, pp. 567–572). Later Chinthayamma published work on independent axioms for ternary Boolean algebras (see 'Sets of independent axioms for a ternary Boolean algebra', *Not. Amer. Math. Soc.* **16**, (4) June 1969, p. 654).

The function  $f$  can be thought of as a 3-place produce, and the function  $g$  may be thought of as inverse.

Axioms for a ternary Boolean algebra:

- |   |  |                                    |
|---|--|------------------------------------|
| 1 | $f(f(v, w, x), y, f(v, w, z)) = f(v, w, f(x, y, z))$ | ax. 1 of a ternary Boolean algebra |
| 2 | $f(y, x, x) = x$                                     | ax. 2 of a ternary Boolean algebra |
| 3 | $f(x, y, g(y)) = x$                                  | ax. 3 of a ternary Boolean algebra |
| 4 | $f(x, x, y) = x$                                     | ax. 4 of a ternary Boolean algebra |
| 5 | $f(g(y), y, x) = x$                                  | ax. 5 of a ternary Boolean algebra |
| 6 | $x = x$  | reflexivity of equality            |

Now remove axiom 3 and add the following lemma:

- |   |                  |                                    |
|---|------------------|------------------------------------|
| 7 | $f(x, y, x) = x$ | lemma provable from 1, 2, 5, and 6 |
|---|------------------|------------------------------------|

The denial of the theorem is as follows:

- |   |                            |                       |
|---|----------------------------|-----------------------|
| 8 | $\neg (f(a, g(a), b) = b)$ | denial of the theorem |
|---|----------------------------|-----------------------|

*Proof:*

- |    |  |                |
|----|--|----------------|
| 9  | $f(f(v, w, x), x, v) = f(v, w, x)$           | 1 7 4          |
| 10 | $f(f(y, w1, z), y, z) = f(y, w1, z)$         | 1 9 2 4 9      |
| 11 | $f(f(v, w, g(y)), y, v) = v$                 | 1 7 5 7        |
| 12 | $f(x, y, f(v, x, y)) = f(v, x, y)$           | 1 2 2          |
| 13 | $f(f(v, w, x), x, f(v1, v, w)) = f(v, w, x)$ | 1 12 4         |
| 14 | $f(y, g(y), z) = z$                          | 1 13 2 10 11 2 |
| 15 | null   | 8 15           |

*Details of Proof:*

- |   |                                    |
|---|------------------------------------|
| 9 | $f(f(v, w, x), x, v) = f(v, w, x)$ |
|---|------------------------------------|

The clause was deduced by performing the following operations:

- Paramodulate into clause 1 using clause 7  
*into-term:*  $f(v, w, z)$  *from-term:*  $f(x, y, x)$   
 result:  $f(f(v, w, x), y, v) = f(v, w, f(x, y, v))$   
 Paramodulate into the result using clause 4  
*into-term:*  $f(x, y, v)$  *from-term:*  $f(x, x, y)$   
 10  $f(f(y, w1, z), y, z) = f(y, w1, z)$

The clause was deduced by performing the following operations:

- Paramodulate into clause 1 using clause 9  
*into-term:*  $f(v, w, x)$  *from-term:*  $f(f(v, w, x), x, v)$   
 result:  $f(f(v1, w1, x1), y, f(f(v1, w1, x1), x1, z)) =$   
 $f(f(v1, w1, x1), x1, f(v1, y, z))$   
 Paramodulate into the result using clause 2  
*into-term:*  $f(f(v1, w1, x1), x1, z)$  *from-term:*  $f(y, x, x)$

result:  $f(f(v1, w1, z), y, z) = f(f(v1, w1, z), z, f(v1, y, z))$

Paramodulate into the result using clause 4

*into-term:*  $f(v1, y, z)$  *from-term:*  $f(x, x, y)$

result:  $f(f(y, w1, z), y, z) = f(f(y, w1, z), z, y)$

Demodulate the result using clause 9

*into-term:*  $f(f(y, w1, z), z, y)$  *from-term:*  $f(f(v, w, x), x, v)$

11  $f(f(v, w, g(y)), y, v) = v$

The clause was deduced by performing the following operations:

Paramodulate into clause 1 using clause 7

*into-term:*  $f(v, w, z)$  *from-term:*  $f(x, y, x)$

result:  $f(f(v, w, x), y, v) = f(v, w, f(x, y, v))$

Paramodulate into the result using clause 5

*into-term:*  $f(x, y, v)$  *from-term:*  $f(g(y), y, x)$

result:  $f(f(v, w, g(y)), y, v) = f(v, w, v)$

Paramodulate into the result using clause 7

*into-term:*  $f(v, w, v)$  *from-term:*  $f(x, y, x)$

12  $f(x, y, f(v, x, y)) = f(v, x, y)$

The clause was deduced by performing the following operations:

Paramodulate into clause 1 using clause 2

*into-term:*  $f(v, w, x)$  *from-term:*  $f(y, x, x)$

result:  $f(x, y, f(v, x, z)) = f(v, x, f(x, y, z))$

Paramodulate into the result using clause 2

*into-term:*  $f(x, y, z)$  *from-term:*  $f(y, x, x)$

13  $f(f(v, w, x), x, f(v1, v, w)) = f(v, w, x)$

The clause was deduced by performing the following operations:

Paramodulate into clause 1 using clause 12

*into-term:*  $f(v, w, z)$  *from-term:*  $f(x, y, f(v, x, y))$

result:  $f(f(v, w, x), y, f(v1, v, w)) =$

$f(v, w, f(x, y, f(v1, v, w)))$

Paramodulate into the result using clause 4

*into-term:*  $f(x, y, f(v1, v, w))$  *from-term:*  $f(x, x, y)$

14  $f(y, g(y), z) = z$

The clause was deduced by performing the following operations:

Paramodulate into clause 1 using clause 13

*into-term:*  $f(v, w, x)$  *from-term:*  $f(f(v, w, x), x, f(v1, v, w))$

result:  $f(f(v1, w1, w), y, f(f(v1, w1, w), w, z)) =$

$f(f(v1, w1, w), w, f(f(v2, v1, w1), y, z))$

Paramodulate into the result using clause 2

*into-term:*  $f(f(v1, w1, w), w, z)$  *from-term:*  $f(y, x, x)$

$$\text{result: } f(f(v1, w1, z), y, z) = f(f(v1, w1, z), z, f(f(v2, v1, w1), y, z))$$

Paramodulate into the result using clause 10

$$\text{into-term: } f(f(v1, w1, z), y, z) \quad \text{from-term: } f(f(y, w1, z), y, z)$$

$$\text{result: } f(y, w1, z) = f(f(y, w1, z), z, f(f(v2, y, w1), y, z))$$

Paramodulate into the result using clause 11

$$\text{into-term: } f(f(v2, y, w1), y, z) \quad \text{from-term: } f(f(v, w, g(y)), y, v)$$

$$\text{result: } f(y, g(y), z) = f(f(y, g(y), z), z, z)$$

Paramodulate into the result using clause 2

$$\text{into-term: } f(f(y, g(y), z), z, z) \quad \text{from-term: } f(y, x, x)$$

### Problem 6

*Problem 6:* In a ring, if  $x^3 = x$  for all  $x$  in the ring, then  $xy = yx$  for all  $x$  and  $y$  in the ring.

Axioms for a ring:

- |    |                                       |                                      |
|----|---------------------------------------|--------------------------------------|
| 1  | $j(0, x) = x$                         | 0 is a left identity for sum         |
| 2  | $j(x, 0) = x$                         | 0 is a right identity sum            |
| 3  | $j(g(x), x) = 0$                      | there exists a left inverse sum      |
| 4  | $j(x, g(x)) = 0$                      | there exists a right inverse sum     |
| 5  | $j(j(x, y), z) = j(x, j(y, z))$       | associativity of addition            |
| 6  | $x = x$                               | reflexivity of equality              |
| 7  | $j(x, y) = j(y, x)$                   | commutativity of addition            |
| 8  | $f(f(x, y), z) = f(x, f(y, z))$       | associativity of multiplication      |
| 9  | $f(x, j(y, z)) = j(f(x, y), f(x, z))$ | distributivity axioms                |
| 10 | $f(j(y, z), x) = j(f(y, x), f(z, x))$ |                                      |
| 11 | $f(x, f(x, x)) = x$                   | $x * x * x = x$ (special hypothesis) |
| 12 | $\neg (f(a, b) = f(b, a))$            | denial of the theorem                |

The proof of this theorem is complex enough to prohibit the type of presentation that we have used for the preceding theorems. We include a proof in the form that a human mathematician might write it. Since the problem is a standard one for graduate courses in algebra, there are commonly available proofs. However, the one that we supply seems somewhat unusual. It was given to us by Steve Winker.

*Proof:*

(1) Note that  $(x^2)^2 = x^2$ , since  $x^3 = x$ .

(2) First, we prove that  $x^2y^2x^2 = y^2x^2y^2$

If we multiply out  $(x^2 - y^2)^3$ , and simplify the result, we find that

$$(x^2 - y^2)^3 = x^2 - x^2y^2x^2 + y^2x^2y^2 - y^2.$$

However, since  $(x^2 - y^2)^3 = (x^2 - y^2)$  by the hypothesis of the theorem,  $x^2y^2x^2 = y^2x^2y^2$ .

(3) Now we can show that squares commute; that is, for any  $x, y$

$$x^2y^2 = y^2x^2.$$

Start with the equation

$$x^2y^2x^2 = y^2x^2y^2.$$

If we right multiply each side with  $(x^2y^2x^2y^2)$  and simplify the result, we derive

$$x^2y^2 = y^2x^2y^2.$$

On the other hand, if we left multiply each side by  $(y^2x^2y^2x^2)$  and simplify we derive

$$y^2x^2 = y^2x^2y^2.$$

By transitivity, we arrive at the desired lemma:  $x^2y^2 = y^2x^2$ .

(4) Now we can show that a square will commute with anything, that is, for any  $x$  and  $y$ ,  $xy^2 = y^2x$ :

$$\begin{aligned} (xy^2)^3 &= x(y^2xy^2x)y^2 = xy^2(y^2xy^2x) = xy^2xy^2x \quad (\text{since squares commute}) \\ (y^2x)^3 &= y^2(xy^2xy^2)x = (xy^2xy^2)y^2x = xy^2xy^2x. \end{aligned}$$

Thus,  $xy^2 = (xy^2)^3 = xy^2xy^2x = (y^2x)^3 = y^2x$

(5) Now we can finish the proof of the theorem: for any  $x$  and  $y$ ,  $xy = yx$ .

$$\begin{aligned} xy &= xyxyxy = xyxy^3xy = xy(xy)y^2xy = xyy^2xyxy = \\ xyyy(xy)^2 &= xy y(xy)^2y = xy^2(xy)^2y = x(xy)^2y^2y = \\ xxyxyyyy &= x^2yxy^2 = y^2yxx^2 = yx \end{aligned}$$

## Summary

In this note we have offered a set of six problems that represent a spectrum in terms of difficulty. The first two problems are relatively trivial. The third and fourth are fairly difficult, although there are several existing theorem-proving systems capable of deriving proofs in fairly short time periods. We have found the fifth problem quite challenging, although the proof is not terribly long. The sixth problem represents the most complex problem in equality for which a proof has been derived by an automated system [5, 4].

The reader should note that the proofs given are not always exactly those generated by automated reasoning systems. In some cases, we have shortened proofs derived by our system. We include proofs only as aids for those who wish to study exactly why their system might be failing to reach a complete proof.

## References

1. Knuth D. E. and Bendix, P. B. 'Simple word problems in universal algebras', pp. 263–297 in *Computational Problems in Abstract Algebra* (ed. Leech) Pergamon Press (1970).

2. Lusk, Ewing L. and Overbeek, Ross A. 'The automated reasoning system ITP', ANL-84-27, Argonne National Laboratory (April, 1984).
3. Robinson, G. and Wos, L. 'Paramodulation and theorem proving in first-order theories with equality', pp. 135-150 in *Machine Intelligence 4* (ed. B. Meltzer and D. Michie) Edinburgh University Press(1969).
4. Stickel, Mark E. 'A case study of theorem proving by the Knuth-Bendix method discovering that  $x^3 = x$  implies ring commutativity', in *Proceedings of the 7th International Conference on Automated Deduction* (ed. R. E. Shostak)
5. Veroff, R. 'Canonicalization and demodulation', Technical Report ANL-81-6, Argonne National Laboratory, Argonne, Illinois (February 1981).
6. Wos, Larry, Overbeek, Ross, Lusk, Ewing and Boyle, Jim, *Automated Reasoning: Introduction and Applications*, Prentice-Hall, Englewood Cliffs, New Jersey (1984).
7. Wos, L. Robinson, G. Carson, D. and Shalla, L. 'The concept of demodulation in theorem proving', *Journal of the ACM* 14, 698-704 (1967).
8. Wos L. and Robinson, G. A. 'Paramodulation and set of support', *Proceedings of the IRLA Symposium on Automatic Demonstration, Versailles, France*, Springer-Verlag pp. 276-310 (1968).

## Response

An abbreviated version of this discussion and example set appeared in the third issue of the AAR Newsletter. Dallas Lankford wrote in response to that article:

I read with some interest the discussion of graduated problems for testing equality reasoning by R. Overbeek and E. Lusk in the AAR Newsletter #3. Perhaps the subscribers might be interested in some additional information on these problems.

The earliest published computer proof of problem 1 that I have found is in Huet's 'Experiments with an interactive prover for logic with equality'. Case Western Reserve University, Jennings Computing Center, report 1106, 1972, pp. 41-42 and pp. 62-63. The computer proof required five rounds of resolution and paramodulation, used some equations as rewrite rules, generated seventeen clauses during the proof search, and required about twelve seconds of CPU. Mike Ballantyne and I subsequently derived a complete set for this problem using completion and commutative-associative completion. The complete set for problem 1 is:

1.  $[x^2] \rightarrow [e]$
2.  $[x^{-1}] \rightarrow [x]$
3.  $[x \cdot e] \rightarrow [x]$

where congruence classes are defined by the commutative and associative axioms. This and other computer experiments were presented at the 3rd CADE at MIT in August of 1977. The most impressive solution of problem 2 is contained in one of the computer experiments by Knuth and Bendix [1970] where the first complete set for free groups was derived. Their paper is found in *Computational Problems in Abstract Algebras*, Pergamon Press, 1970, and in Vol. 2 of *Automation of Reasoning*, Springer-Verlag, 1983. Problem 3 also has a solution by completion, apparently first observed by Hsiang in his July 1981 paper, 'Refutational theorem proving using term rewriting systems'. The complete set is

1.  $[x + 0] \rightarrow [x]$
2.  $[x + x] \rightarrow [0]$

3.  $[x \cdot 1] \rightarrow [x]$
4.  $[x \cdot x] \rightarrow [x]$
5.  $[x \cdot (y + z)] \rightarrow [(x \cdot y) + (x \cdot z)]$
6.  $[x \cdot 0] \rightarrow [0]$
7.  $[-x] \rightarrow [x]$

where congruence classes are defined by the commutative and associative axioms for addition and multiplication. The first computer proof of problem 4 that I know of is contained in Nevin's 1974 *JACM* paper, required 30 minutes of CPU time, and generated a search space of over 400 formulas. Subsequently a completion-based theorem prover implemented by Ballantyne and Lankford solved problem 4 in 30 seconds, and terminated with a search space of 11 formulas, cf. Bledsoe's 'None-resolution theorem proving' in *IJCAI-75* and *AI Journal* (1977). A variant of problem 5 is proved by Nevin's computer program, see his 1974 *JACM* paper. Concerning the difficulty of problem 6, it depends on how much information is given to the computer program. In Veroff's computer proof, much information was provided by the human in the form of considerable clausal information, and so the computer proof was relatively easy. By contrast, little information about the problem was given to his program by Stickel, and so the computer proof was quite difficult. Moreover, Stickel's program found a decision algorithm for free  $(x^3 = x)$ -rings, which is a much deeper result than just showing  $(x^3 = x)$ -rings are commutative. Because complete sets are decision algorithms, the computer completion proofs for problems 1, 2, and 3 also found decision algorithms for  $(x^2 = e)$ -groups, groups, and  $(x^2 = x)$ -rings (i.e. Boolean rings). Whether completion decision algorithms exist for  $(x^3 = e)$ -groups and ternary Boolean algebras (problems 4 and 5) appears to be currently unknown. Although it is open whether complete sets exist for problems 4 and 5, problem 4 is known to be decidable. Groups satisfying  $x^n = e$  are called Burnside groups, and have word problem decision algorithms for  $n = 2, 3, 4$ , and 6, and for odd  $n \geq 665$ , see Adian's *The Burnside Problem and Identities in Groups*, Springer-Verlag, 1979, pp. 250. It is unknown whether tractable computer implementations of the approach described by Adian can be developed. In my opinion, these are two very important open problems in applied equational logic.