A calculational proof of Andrews's challenge

David Gries[1]
Computer Science, Cornell University
Ithaca, NY 14853[2]

August 1996

At the Marktoberdorf summer school in August 1996, Larry Paulson lectured on his mechanical theorem prover, Isabelle; Natarajan Shankar lectured on his mechanical theorem prover, PVS; and I lectured on calculational logic. Both Paulson and Shankar suggested that I try the calculational approach on Andrew's challenge, which is one of several theorems used to benchmark mechanical theorem provers. Andrew's challenge is to prove the following theorem.[3]

(1) $((\exists x \forall y \mid: p.x \equiv p.y) \;\equiv\; ((\exists x \mid: q.x) \;\equiv\; (\forall y \mid: p.y))) \;\equiv$
$((\exists x \forall y \mid: q.x \equiv q.y) \;\equiv\; ((\exists x \mid: p.x) \;\equiv\; (\forall y \mid: q.y)))$

In proving Andrew's challenge using the calculational approach, I use theorems given in the text [1] (or in its as-yet-unpublished second edition). The Appendix contains theorems used here that may be unfamiliar to the reader.

Now, $\equiv$ is both associative and symmetric, so we can rewrite Andrew's challenge as

$P \;\equiv\; Q$

where P and Q are defined by the following.

$P:\; (\exists x \forall y \mid: p.x \equiv p.y) \;\equiv\; (\exists x \mid: p.x) \;\equiv\; (\forall y \mid: p.y)$
$Q:\; (\exists x \forall y \mid: q.x \equiv q.y) \;\equiv\; (\exists x \mid: q.x) \;\equiv\; (\forall y \mid: q.y)$

where it is assumed that this formula is closed (so $p.x$ and $q.x$ contain no free variables other than x).

This form gives the impression that perhaps P is valid (or invalid), regardless of p. If this is the case, then Q is also valid (or invalid). Hence, we try to prove P.

We don't have many theorems that deal with $\equiv$ as they appear in P, so we try to prove P by mutual implication, proving instead

(2) $((\exists x \forall y \mid: p.x \equiv p.y) \;\equiv\; (\exists x \mid: p.x)) \;\Leftarrow\; (\forall y \mid: p.y)$ and
(3) $((\exists x \forall y \mid: p.x \equiv p.y) \;\equiv\; (\exists x \mid: p.x)) \;\Rightarrow\; (\forall y \mid: p.y)$

[1] Supported by NSF grants CDA-9214957 and CCR-9503319.
[2] http://www.cs.cornell.edu/Info/People/gries/gries.html gries@cs.cornell.edu
[3] We use the notation $(\forall x \mid: (\exists y \mid: P))$ may be abbreviated as $(\forall x \exists y \mid: P)$. Also, we use $\equiv$ for equality over the bools and $=$ for equality over any type (including the bools). Our precedences are, beginning with the tightest, $\neg$, $=$, $\vee$ and $\wedge$, $\Rightarrow$ and $\Leftarrow$, $\equiv$. Finally, in order to eliminate parentheses, we write $p.x$ instead of $p(x)$ for application of function p to variable x.

Proof of (2). Assume $(\forall y \mid : p.y)$

$$
\begin{array}{ll}
 & (\exists x \forall y \mid : p.x \equiv p.y) \;\equiv\; (\exists x \mid : p.x) \\
= & \quad \langle \text{Assumption, instantiated with } y := x \text{ and with } y := y\,, \\
 & \quad\;\; \text{so } p.x \;\equiv\; \mathit{true} \text{ and } p.y \;\equiv\; \mathit{true}\,\rangle \\
 & (\exists x \forall y \mid : \mathit{true} \equiv \mathit{true}) \;\equiv\; (\exists x \mid : \mathit{true}) \\
= & \quad \langle \text{Identity of } \equiv \text{ (5); } (\forall y \mid : \mathit{true}) \;\equiv\; \mathit{true}\,\rangle \\
 & (\exists x \mid : \mathit{true}) \;\equiv\; (\exists x \mid : \mathit{true}) \quad \text{—Reflexivity of } \equiv \text{ (6)}
\end{array}
$$

□

Proof of (3).

$$
\begin{array}{ll}
 & (3) \\
= & \quad \langle \text{Contrapositive, } X \Rightarrow Y \;\equiv\; \neg Y \Rightarrow \neg X\,\rangle \\
 & \neg(\forall y \mid : p.y) \;\Rightarrow\; \neg((\exists x \forall y \mid : p.x \equiv p.y) \;\equiv\; (\exists x \mid : p.x)) \\
= & \quad \langle \text{De Morgan (12) on antecedent;} \\
 & \qquad \neg(X \equiv Y) \;\equiv\; X \;\equiv\; \neg Y \text{ and De Morgan (11) on the consequent} \rangle \\
 & (\exists y \mid : \neg p.y) \;\Rightarrow\; ((\exists x \forall y \mid : p.x \equiv p.y) \;\equiv\; (\forall x \mid : \neg p.x))
\end{array}
$$

By Metatheorem Witness (13), the last formula is a theorem iff the following one is.

$$\neg p.\hat{y} \;\Rightarrow\; ((\exists x \forall y \mid : p.x \equiv p.y) \;\equiv\; (\forall x \mid : \neg p.x))$$

We calculate:

Assume $\neg p.\hat{y}$, so also $p.\hat{y} \;\equiv\; \mathit{false}$

$$
\begin{array}{ll}
 & (\exists x \forall y \mid : p.x \equiv p.y) \\
= & \quad \langle \text{Lemma (4)} \quad \text{—heading to change } p.x \text{ to } p.\hat{y}\,\rangle \\
 & (\exists x \mid : (\forall y \mid : p.x \equiv p.y) \wedge p.x \equiv p.\hat{y}) \\
= & \quad \langle \text{Substitution (8)} \rangle \\
 & (\exists x \mid : (\forall y \mid : p.\hat{y} \equiv p.y) \wedge p.x \equiv p.\hat{y}) \\
= & \quad \langle \text{Lemma (4)} \rangle \\
 & (\exists x \forall y \mid : p.\hat{y} \equiv p.y) \\
= & \quad \langle \text{Assumption } p.\hat{y} \equiv \mathit{false}\,;\; \mathit{false} \;\equiv\; X \;\equiv\; \neg X\,\rangle \\
 & (\exists x \forall y \mid : \neg p.y) \\
= & \quad \langle \text{Provided } x \text{ doesn't occur free in } X\,,\; (\exists x \mid : X) \;\equiv\; X\,\rangle \\
 & (\forall y \mid : \neg p.y)
\end{array}
$$

□

(4) **Lemma.** $(\forall x \mid : S.x) \;\equiv\; (\forall x \mid : S.x) \wedge S.t$

Proof.

$$
\begin{array}{ll}
 & (\forall x \mid \mathit{true} : S.x) \\
= & \quad \langle \text{Zero of } \vee \text{ (7)} \rangle \\
 & (\forall x \mid \mathit{true} \vee x = t : S.x) \\
= & \quad \langle \text{Range split (10)} \rangle \\
 & (\forall x \mid \mathit{true} : S.x) \wedge (\forall x \mid x = t : S.x) \\
= & \quad \langle \text{One-point rule (9)} \rangle \\
 & (\forall x \mid \mathit{true} : S.x) \wedge S.t
\end{array}
$$

□

References

[1] Gries, D., and F.B. Schneider. *A Logical Approach to Discrete Math.* Springer Verlag, NY, 1993.

Appendix. Some of the theorems used in the proof

(5) **Identity of $\equiv$:** $true \equiv Q \equiv Q$

(6) **Reflexivity of $\equiv$:** $P \equiv P$

(7) **Zero of $\vee$:** $P \vee true \equiv true$

(8) **Substitution:** $X{=}Y \wedge E^V_X \equiv X{=}Y \wedge E^V_Y$

(9) **One-point rule:** Provided x does not occur free in E,

$(\forall x \mid x = E : P) = P[x := E]$

(10) **Range split:** $(\forall x \mid R \vee S : P) = (\forall x \mid R : P) \wedge (\forall x \mid S : P)$

(11) **De Morgan:** $\neg(\exists x \mid R : P) \equiv (\forall x \mid R : \neg P)$

(12) **De Morgan:** $\neg(\forall x \mid R : P) \equiv (\exists x \mid R : \neg P)$

(13) **Metatheorem Witness.** Suppose $\hat{x}$ does not occur free in P, Q, and R. Then

$(\exists x \mid R : P) \Rightarrow Q$ is a theorem iff
$(R \wedge P)[x := \hat{x}] \Rightarrow Q$ is a theorem.