The Reflection Theorem
Formalizing Meta-Theoretic Reasoning

Lawrence C. Paulson
Computer Laboratory
UNIVERSITY OF CAMBRIDGE
Lecture Overview

• Motivation for the Reflection Theorem
• Proving the Theorem in Isabelle
• Applying the Reflection Theorem
Why Do Proofs By Machine?

• **Claim**: too many been done already!
  – Gödel’s incompleteness theorem (Shankar)
  – thousands of Mizar proofs

• **Reply**: many types of reasoning are hard to formalize.
  – Algebraic structures (e.g. group theory)
  – Meta-level reasoning (e.g. about own proof)
Idea of the Reflection Theorem

\[ M = \bigcup_{\alpha \in \text{ON}} M_\alpha \]

\( \square \) has the same meaning in \( M \) as in \( M_\square \), for arbitrarily large \( \square \)
The Reflection Theorem

Define the class \( M = \bigcup_{\alpha \in \text{ON}} M_\alpha \)

where \( \{M_\alpha\} \) is a monotonic and continuous family of sets.

For each formula \( \phi(x_1, \ldots, x_n) \) there are arbitrarily large ordinals \( \square \) such that \( \square \) holds in \( M \) iff \( \square \) holds in \( M_\square \).
Why is it Hard to Formalize?

• “\( f \) holds in \( M \)” is not definable in ZF!
  – Because \( M \) is a proper class
  – Tarski: the nondefinability of truth

• \( f \) could take any number of arguments

• There is a different proof for each \( f \)
  – Reflection is a *meta-theorem*
  – … and not a *theorem scheme*
Must Define Truth Syntactically

\[(x = y)^M \iff x = y\]
\[(x \in y)^M \iff x \in y\]
\[(\phi \land \psi)^M \iff \phi^M \land \psi^M\]
\[\neg \phi^M \iff \neg (\phi^M)\]
\[(\exists x \phi)^M \iff \exists x (x \in M \land \phi^M)\]

The relativization of $\square$ to $M$
Lecture Overview

• Motivation for the Reflection Theorem
• **Proving the Theorem in Isabelle**
• Applying the Reflection Theorem
Isabelle/ZF

• Same code base as Isabelle/HOL
• Higher-order metalogic, ideal for
  – Theorem schemes
  – Classes
  – Class functions
• Develops set theory from the Zermelo-Fraenkel axioms to transfinite cardinals
Proving the Reflection Theorem in Isabelle/ZF

• Use a clean proof from Mostowski, 1969
  – closed unbounded classes of ordinals
  – normal functions (continuous, increasing)

• One lemma for each logical connective

• Isabelle automatically uses the lemmas to prove instances of the theorem
Closed/Unbounded Classes

• *Closed* means closed under unions (limits) of ordinals

• If $M, N$ are C.U. then so is $M \sqcup N$

• The fixedpoints of a continuous, increasing function form a C.U. class

• E.g. the many solutions of $\mathcal{N}_\alpha = \alpha$
Essence of Proof

• “Skolemize” each \( \exists \) quantifier, obtaining a normal function, \( F \)
  
  – The fixedpoints of \( F \) give the desired class of ordinals

• For \( \forall \) \( \exists \) simply intersect the classes

• Negation and atomic cases are trivial
Lecture Overview

• Motivation for the Reflection Theorem
• Proving the Theorem in Isabelle
• Applying the Reflection Theorem
The Axiom of Choice and the Generalized Continuum Hypothesis

• Gödel (1940) proved them consistent with set theory
  – A deep and important theorem
  – Addressed Hilbert’s First Problem

• Modern treatments (in ZF) require the Reflection theorem
Sketch of Gödel’s Proof

• Define the *constructible universe*, $L$
  – $L_{a+1}$ adds subsets that can be defined from existing elements (in $L_a$) by a formula
  – $L$ contains only sets that must exist

• Show that $L$ satisfies the ZF axioms
  – Comprehension uses Reflection Theorem:
    $\square$ holds in $L$ iff $\square$ holds in some $L_a$

• Show that $L$ satisfies AC and GCH
Showing That $L$ “Thinks” All Sets are Constructible

• Amounts to showing that the construction of $L$ is idempotent

• Relies on the concept of absoluteness:
  – $f$ is absolute if it’s preserved in all models
  – Not absolute: powersets, function spaces, transfinite cardinals

• Requires analysing $L$’s definition down to the last detail
Applying Reflection to $L$

- Define a ZF datatype of FOL formulas
- Define a vocabulary for Reflection
  - No function symbols; purely relational!
  - All concepts from the empty set to “constructible”
  - Repeat for the formula datatype
- For each instance of Comprehension, prove an instance of Reflection (automatically)
- Giant terms describe the classes of ordinals
Finish the Consistency Proof?

• Gödel, 1940: *if a contradiction from AC and GCH could be derived, it could be transformed into a contradiction from the axioms of set theory alone.*

• Theorem statement lies outside the language of set theory!

• It is an even better example of meta-theoretic reasoning.