SENTENTIAL REFLECTION

by
Harvey M. Friedman
Ohio State University
Princeton University
friedman@math.ohio-state.edu
http://www.math.ohio-state.edu/~friedman/

January 5, 2003

Abstract. We present two forms of "sentential reflection", which are shown to be mutually interpretable with Z_2 and ZFC, respectively.

1. Introduction.

We use "class" as a neutral term, without commitment to the developed notions of "set" and "class" that have become standard in set theory and mathematical logic. We use $\in$ for membership.

This framework supports interpretations of sentential reflection that may differ from conventional set theory or class theory. However, we do not pursue this direction here.

This framework is intended to accommodate objects that are not classes. Such nonclasses are treated as classes with no elements. Thus we are careful not to assume extensionality. In fact, we will not assume any form of extensionality.

All of our formal theories of classes are in the language $L(\in)$, which is the usual classical first order predicate calculus with only the binary relation symbol $\in$ (no equality).

We use "category of classes" or just "category" as a neutral term, not specifically related to category theory. They are given by a formula of $L(\in)$ with a distinguished free variable, with parameters allowed.

The first version of sentential reflection that we consider is $SR(\in)$, which informally asserts the following.

if a given sentence of $L(\in)$ holds in a given category
then it holds in a subclass.

Here “holds in” means that it holds if the quantifiers are relativized over the category or class.

THEOREM 1.1. SR(□) is mutually interpretable with Z2.

See [Si99] for copious material on Z_2, which is the standard two sorted first order system for “second order arithmetic”.

Z_2 is much weaker in interpretation power than ZFC. In particular, ZFC is not interpretable in SR(□). So in order to achieve mutually interpretability with ZFC, we strengthen the choice of language.

One way of strengthening SR(□) is to strengthen the notion of “subclass”.

We say that the class x is an inclusion subclass of the category K if and only if

x is a subclass of K, and any element of K
that is a subclass of an element of x lies in x.

This can be restated by a single clause as follows.

any subclass of any element of x lies in K
if and only if it lies in x.

SRIS(□) informally asserts the following.

if a given sentence of L(□) holds in a given category
then it holds in an inclusion subclass.

Here SRIS(□) is read “sentential reflection by inclusion subclass for L(□)”.

THEOREM 1.2. SRIS(□) is provable in ZF and interprets ZFC.

For the formal presentations of SR(□) and SRIS(□), let □ be a formula of L(□), and let □ be a sentence of L(□) with no variables in common with □. Let y be a variable in □. Let □[□,y] be the result of replacing all quantifiers (Qz) in □ by

(Qz|□[y/z])

and expanding the result to a formula of L(□).

Informally, □[□,y] is the (formula expressing the) result of relativizing all quantifiers in □ to {y: □(y)}. In particular, □[y □ x,y] is the (formula expressing the) result of relativizing all quantifiers in □ to (the elements of) x.

SR(□) is the formal system in L(□) whose nonlogical axioms are

□[□,y] □ (□x)((□y □ x)(□) □ □[y □ x,y])

where x,y are distinct variables, □ is a formula of L(□) in which x is not free, □ is a sentence of L(□) with no variables in common with □, and y is a variable in □.

SRIS(□) is the formal system in L(□) whose nonlogical axioms are

□[□,y] □ (□x)((□y □ z □ x)(□ □ y □ x) □ □[y □ x,y])

where x,y,z are distinct variables, □ is a formula of L(□) in which x,z are not free, □ is a sentence of L(□) with no variables in common with □, and y is a variable in □.

2. SR(□). To be completed.

3. SRIS(□). To be completed.