The Relative Consistency of the Axiom of Choice — Mechanized Using Isabelle/ZF

(Extended Abstract)

Lawrence C. Paulson

Computer Laboratory, University of Cambridge, England
LP15@cam.ac.uk

Gödel [3] published a monograph in 1940 proving a highly significant theorem, namely that the axiom of choice (AC) and the generalized continuum hypothesis (GCH) are consistent with respect to the other axioms of set theory. This theorem addresses the first of Hilbert's famous list of unsolved problems in mathematics. I have mechanized this work [8] using Isabelle/ZF [5, 6]. Obviously, the theorem's significance makes it a tempting challenge; the proof also has numerous interesting features. It is not a single formal assertion, as most theorems are. Gödel [3, p. 33] states it as follows, using Σ to denote the axioms for set theory:

> What we shall prove is that, if a contradiction from the axiom of choice and the generalized continuum hypothesis were derived in Σ, it could be transformed into a contradiction obtained from the axioms of Σ alone.

Gödel presents no other statement of this theorem. Neither does he introduce a theory of syntax suitable for reasoning about transformations on proofs, surely because he considers it to be unnecessary.

Gödel's work consists of several different results which, taken collectively, express the relative consistency of the axiom the choice. The concluding inference takes place at the meta-level and is not formalized. Standard proofs use meta-level reasoning extensively. Gödel writes [3, p. 34],

> However, the only purpose of these general metamathematical considerations is to show how the proofs for theorems of a certain kind can be accomplished by a general method. And, since applications to only a finite number of instances are necessary . . ., the general metamathematical considerations could be left out entirely, if one took the trouble to carry out the proofs separately for any instance.

I decided to take the trouble, with the help of a mechanical theorem prover.

In brief, the proof goes as follows. We define a class model, called **L**, for the axioms of set theory.[1] **L** can be seen as containing just the sets that must exist because they can be defined by formulas. Since **L** is a proper class and

[1] A class in ZF is simply a first-order formula in one variable. We typically endow classes with set notation, e.g. writing $a \in \mathbf{L}$ rather than $\mathbf{L}(a)$, but they exist only in the metalanguage.

not a set, we need to be careful about the notion of satisfaction. We cannot talk within ZF about a formula being satisfied by a class model. Instead we transform the formula, restricting each quantifier to range over $\mathbf{L}$ instead of ranging over all sets. For example, we transform $\forall x\, \phi(x)$ into $\forall x\, [x \in \mathbf{L} \to \phi^{\mathbf{L}}(x)]$, where $\phi^{\mathbf{L}}(x)$ is the result of recursively transforming ϕ. This transformation is called *relativization.* If the relativized formula is a theorem, then we say that the original formula is true in $\mathbf{L}$. We must prove that $\mathbf{L}$ satisfies (in that sense) all the axioms of set theory, and we must further prove that $\mathbf{L}$ satisfies the axiom of choice. Although we continue to work in first-order logic and ZF, relativizing all quantifiers to $\mathbf{L}$ has the effect of augmenting our axiom system with the axiom of choice.

- Provided we work entirely with formulas relativized to $\mathbf{L}$, we can prove all the consequences of the axioms of set theory including the axiom of choice.
- Because relativization is merely a syntactic transformation within first-order logic, every proof in $\mathbf{L}$ is also a proof in the original set theory, which lacks the axiom of choice.
- The relativization of **false** is **false**.

Thus, if we prove **false** using the axiom of choice, then we have also found a contradiction in the original set theory. This is a strong form of relative consistency. Gödel specifically notes that a contradiction in basic set theory "could actually be constructed" [3, p. 87] from a contradiction in $\mathbf{L}$. We merely have to express this contradiction using formulas relativized to $\mathbf{L}$. However, to show that a proof exists using relativized formulas seems to require a small amount of proof theory [8].

The main steps of the proof are as follows:

1. Define the class $\mathbf{L}$.
2. Prove that $\mathbf{L}$ satisfies the axioms of set theory. For ZF, the main difficulty is the axiom scheme of comprehension, also known as separation.
3. Prove that $\mathbf{L}$ satisfies the assertion "every set belongs to $\mathbf{L}$," which is traditionally written $\mathbf{V} = \mathbf{L}$.
4. Prove that $\mathbf{V} = \mathbf{L}$ implies AC.

Set-theoretic notation complicates the formalization. We are accustomed to writing unions, intersections, etc, with variable binding as in $\bigcup_{x \in A} B(x)$. But formally, the language of set theory consists of first-order logic plus the membership relation and equality. It has no terms other than individual variables. Before we can relativize an expression $E(x)$, we must translate it into a pure formula $\phi(x, y)$ such that $\phi(x, y) \leftrightarrow y = E(x)$. We must even translate the complicated expressions generated by Isabelle/ZF as it processes recursive definitions of sets and functions. In mathematical textbooks, relativization is done implicitly: all you have to do is put the superscript $\mathbf{L}$ on a term or formula. For example, the claim that $\mathbf{L}$ satisfies $\mathbf{V} = \mathbf{L}$ is trivially expressed by $(\mathbf{V} = \mathbf{L})^{\mathbf{L}}$. In the Isabelle/ZF proof, I have had to write out each relativized expression explicitly for each concept used in the construction of $\mathbf{L}$, in order to express $(\mathbf{V} = \mathbf{L})^{\mathbf{L}}$.

Proving that **L** satisfies $\mathbf{V} = \mathbf{L}$ is a key part of the proof, and despite first appearances, it is not trivial. It amounts to saying that the construction of **L** is idempotent. In other words, if starting in **L** we repeat the construction of **L**, then it will yield the whole of **L** and not some subclass of it. The underlying concept is called *absoluteness*, which expresses that a given notion or expression is the same in every transitive model of set theory.[2] Most constructions are absolute. For example, $A \subseteq B$ can only mean that each element of A also belongs to B. The empty set, obviously, can only be a set containing no elements. If A and B are sets then their union can only be the set containing precisely the elements of those sets. Wellorderings and ordinals are absolute. Powersets however are not absolute, for there could be many subsets even of the natural numbers that cannot be shown to exist; they could exist in some models and not in others.

Skolem's paradox [4, p. 141] is a striking illustration that cardinality is not absolute. Set theorists naturally assume that models exist of the ZF axioms, from which it follows by the downward Löwenheim-Skolem theorem that there exists a countable model M of ZF. The "paradox" is that this countable model "thinks" that it contains arbitrarily large cardinals. More precisely, if α is an uncountable cardinal according to M, then obviously α must be really be countable because $\alpha \subseteq M$. The point is that none of the bijections between α and ω belong to M; although the property of being a bijection is absolute, the property of being the set of all functions from α to ω is not. Neither is the property of being a cardinal.

Papers on formal verification often describe the work as "straightforward but tedious." The idempotence proof meets this description in the extreme. It has been necessary to relativize all the concepts of set theory, from the empty set to ordinals, recursive functions, etc. Then I had to prove that each of these concepts was absolute. In essence, this amounts to examining each definition to ensure that it uses only absolute constructions. Powersets are not absolute, but they appear surprisingly often, and then an alternative definition must be found and proved equivalent to the original one. The treatment of recursion was particularly difficult. I had to prove much of the foundations of recursion again from first principles. Having done this, we cannot merely note that all functions defined by recursion are absolute, as textbooks do. We must take each recursive definition, take it apart piece by piece, prove absoluteness for the pieces and feed those results into a theorem that will yield absoluteness for that particular function. I have done all of this with respect to an arbitrary transitive class model **M**, and later instantiated the proofs to **L**.

Further tedium arises from the need to internalize the notion of formula. A recursive datatype of formulas is defined, since it is needed to define **L**. Most of the relativized formulas mentioned in the previous paragraph have to be translated a second time into this datatype of formulas. Fortunately, some of the translations are done automatically.

Once the idempotence proof is done, we are justified in assuming $\mathbf{V} = \mathbf{L}$. I have separately proved that $\mathbf{V} = \mathbf{L}$ implies the axiom of choice. This proof is straightforward both in concept and in execution. By transfinite induction,

[2] **M** is *transitive* if $x \in \mathbf{M}$ implies $x \subseteq \mathbf{M}$.

each level of the construction of **L** is well-ordered. The wellordering comes in an obvious way from the countability of the set of formulas. Gödel went on to prove that $\mathbf{V} = \mathbf{L}$ implies the generalized continuum hypothesis. Although I omitted this step, it can probably be done with an acceptable amount of effort.

My formalization has two limitations. First, I am not able to prove that **L** satisfies the axiom scheme of comprehension. Although Isabelle/ZF handles schematic proofs easily, the proof of comprehension for the formula ϕ requires an instance of the reflection theorem for ϕ. Each instance of reflection [7] involves recursion over the structure of ϕ. Each instance of comprehension therefore has a different proof and must be proved separately. At the meta-level, of course, all of these proofs are instances of one algorithm, and they are generated by nearly identical proof scripts. Reasoning at the meta-level, we can see that all instances of the reflection theorem are available and that they imply all instances of the axiom of comprehension. But these meta-level inferences cannot be formalized in my framework. The inability to prove comprehension once and for all added further tedium to the project: in the absoluteness proofs, I had to keep track of each instance of comprehension that I used. Then, in order to instantiate these proofs to **L**, I had to prove that each of those instances held in **L**. There are about 35 such instances.

My formalization has another limitation. The proof that **L** satisfies $\mathbf{V} = \mathbf{L}$ cannot be combined with the proof that $\mathbf{V} = \mathbf{L}$ implies the axiom of choice in order to conclude that **L** satisfies the axiom of choice. The reason is that the two instances of $\mathbf{V} = \mathbf{L}$ are formalized very differently: one is relativized and the other is not. These problems arise because my work builds on the existing Isabelle/ZF formalization of set theory, comprising some 20 000 lines of proof scripts, rather than creating an new mechanized proof system specifically for Gödel's proof. Using Isabelle/ZF allows much of the work to be undertaken in the style of textbook proofs, and it enjoys the property that every proof involving relativized formulas (including one of **false**) is also a proof in ZF.

We could remedy both limitations by tackling Gödel's proof in a quite different way, working entirely in the metatheory. Unfortunately, experience shows that a formalized metatheory is convenient only for proving metatheorical results and not for proving, e.g., specific theorems of set theory. The formalization of the theorem statement would have to be done with care: the obvious $\mathrm{Con(ZF)} \to \mathrm{Con(ZF} + (\mathbf{V}{=}\mathbf{L}))$ sacrifices a crucial aspect of Gödel's result, namely that a contradiction in $\mathrm{ZF} + (\mathbf{V}{=}\mathbf{L})$ can be *effectively transformed* into a contradiction in ZF. Thus it appears necessary to introduce both proof theory and a model of computation, imposing two unenlightening technical layers onto Gödel's construction. I leave such issues as a challenge for the theorem-proving community.

A few other researchers have undertaken mechanized proof in set theory. Quaife [9] has generated proofs of hundreds of elementary results using the Otter resolution theorem prover, starting with a machine-oriented formalization of Bernays-Gödel set theory. BG set theory differs from ZF in that it replaces the axiom scheme of comprehension by a finite set of primitives that can be used to express particular comprehensions; these primitives are difficult to use, but they allow the axiom system to be finite. Building on Quaife's work, Belinfante has implemented a Mathematica program for translating comprehensions into the BG formalism; he submits these to Otter and thereby has proved facts about the ordinals [2], for example. The Mizar system is based on Tarski-Grothendieck set theory. It is designed for formalizing mathematics [1] and not merely for formalizing set theory; however, much set theory has been formalized using Mizar, for example some elementary facts concerning large cardinal axioms [10].

In many respects, my formalization follows traditional ones. My development is largely based on Kunen [4]. My use of native set theory (as embodied in Isabelle/ZF) is very much in the spirit of those proofs, although it leads to the difficulties mentioned above. A byproduct of the work is a general theory of absoluteness for arbitrary class models of ZF. It could be used for other formal investigations of inner models.

Acknowledgements. Krzysztof Grąbczewski devoted much effort to an earlier, unsuccessful, attempt to formalize Gödel's proof. Isabelle is funded by the U.K.'s Engineering and Physical Sciences Research Council, grants GR/M75440, GR/R 01156/01, GR/S57198/01, etc. Isabelle is developed in large part by Prof. Tobias Nipkow's group at the Technical University of Munich.

References

1. Grzegorz Bancerek and Piotr Rudnicki. A compendium of continuous lattices in mizar. *Journal of Automated Reasoning*, 29(3-4):189–224, 2002.
2. Johan G. F. Belinfante. On computer-assisted proofs in ordinal number theory. *Journal of Automated Reasoning*, 22(3):341–378, March 1999.
3. Kurt Gödel. The consistency of the axiom of choice and of the generalized continuum hypothesis with the axioms of set theory. In S. Feferman et al., editors, *Kurt Gödel: Collected Works*, volume II, pages 33–101. Oxford University Press, 1990. First published in 1940 by Princeton University Press.
4. Kenneth Kunen. *Set Theory: An Introduction to Independence Proofs.* North-Holland, 1980.
5. Lawrence C. Paulson. Set theory for verification: I. From foundations to functions. *Journal of Automated Reasoning*, 11(3):353–389, 1993.
6. Lawrence C. Paulson. Set theory for verification: II. Induction and recursion. *Journal of Automated Reasoning*, 15(2):167–215, 1995.
7. Lawrence C. Paulson. The reflection theorem: A study in meta-theoretic reasoning. In Andrei Voronkov, editor, *Automated Deduction — CADE-18 International Conference*, LNAI 2392, pages 377–391. Springer, 2002.
8. Lawrence C. Paulson. The relative consistency of the axiom of choice — mechanized using Isabelle/ZF. *LMS Journal of Computation and Mathematics*, 6:198–248, 2003. `http://www.lms.ac.uk/jcm/6/lms2003-001/`.
9. Art Quaife. Automated deduction in von Neumann-Bernays-Gödel set theory. *Journal of Automated Reasoning*, 8(1):91–147, 1992.
10. Josef Urban. Basic facts about inaccessible and measurable cardinals. *Journal of Formalized Mathematics*, 12, 2000. Published online at `http://mizar.uwb.edu.pl/JFM/Vol12/card_fil.html`.