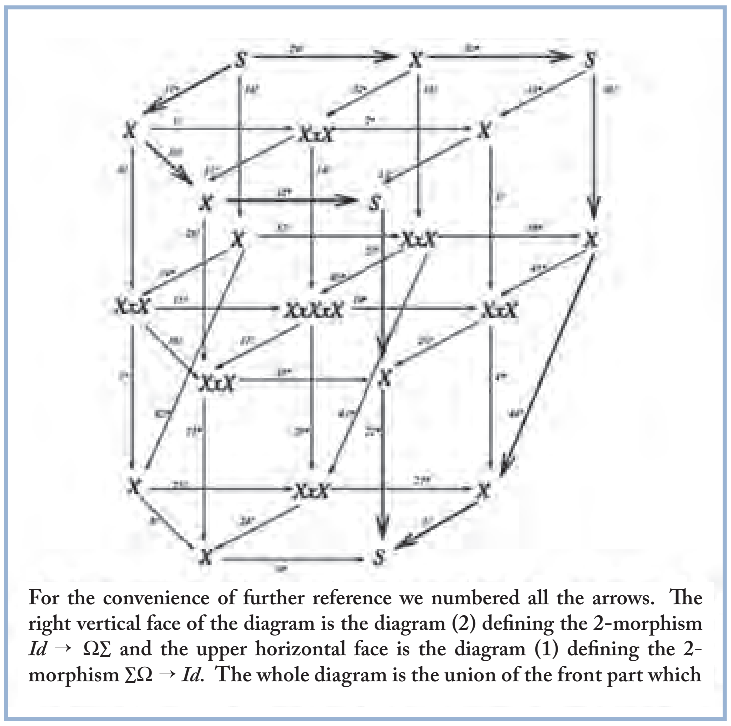
The Origins and Motivations of Univalent Foundations

Professor Voevodsky's Personal Mission to Develop Computer Proof Verification to Avoid Mathematical Mistakes

BY VLADIMIR VOEVODSKY

In January 1984, Alexander Grothendieck submitted to the French National Centre for Scientific Research his proposal "Esquisse d'un Programme." Soon copies of this text started circulating among mathematicians. A few months later, as a first-year undergraduate at Moscow University, I was given a copy of it by George Shabat, my first scientific adviser. After learning some French with the sole purpose of being able to read this text, I started to work on some of the ideas outlined there.



For the convenience of further reference we numbered all the arrows. The right vertical face of the diagram is the diagram (2) defining the 2-morphism $Id \to \Omega\Sigma$ and the upper horizontal face is the diagram (1) defining the 2-morphism $\Sigma\Omega \to Id$. The whole diagram is the union of the front part which

This three-dimensional diagram is an example of the kind of "formulas" that Voevodsky would have to use to support his arguments about 2-theories.

In 1988 or 1989, I met Michael Kapranov who was equally fascinated by the perspectives of developing mathematics of new "higher-dimensional" objects inspired by the theory of categories and 2-categories.

The first paper that we published together was called "∞-Groupoids as a Model for a Homotopy Category." In it, we claimed to provide a rigorous mathematical formulation and a proof of Grothendieck's idea connecting two classes of mathematical objects: ∞-groupoids and homotopy types.

Later we decided that we could apply similar ideas to another top mathematical problem of that time: to construct motivic cohomology, conjectured to exist in a 1987 paper by Alexander Beilinson, Robert MacPherson (now Professor in the School of Mathematics), and Vadim Schechtman.

In the summer of 1990, Kapranov arranged for me to be accepted to graduate school at Harvard without applying. After a few months, while he was at Cornell and I was at Harvard, our mathematical paths diverged. I concentrated my efforts on motivic cohomology and later on motivic homotopy theory. My notes dated March 29, 1991, start with the question "What is a homotopy theory for algebraic varieties or schemes?"

The field of motivic cohomology was considered at that time to be highly speculative and lacking firm foundation. The groundbreaking 1986 paper "Algebraic Cycles and Higher K-theory" by Spencer Bloch was soon after publication found by Andrei Suslin to contain a mistake in the proof of Lemma 1.1. The proof could not be fixed, and almost all of the claims of the paper were left unsubstantiated.

A new proof, which replaced one paragraph from the original paper by thirty pages of complex arguments, was not made public until 1993, and it took many more years for it to be accepted as correct. Interestingly, this new proof was based on an older result of Mark Spivakovsky, who, at about the same time, announced a proof of the resolution of singularities conjecture. Spivakovsky's proof of resolution of singularities was believed to be correct for several years before being found to contain a mistake. The conjecture remains open.

I DIDN'T HAVE THE TOOLS TO EXPLORE THE AREAS WHERE CURIOSITY WAS LEADING ME AND THE AREAS THAT I CONSIDERED TO BE OF VALUE AND OF INTEREST AND OF BEAUTY.

The approach to motivic cohomology that I developed with Andrei Suslin and Eric Friedlander circumvented Bloch's lemma by relying instead on my paper "Cohomological Theory of Presheaves with Transfers," which was written when I was a Member at the Institute in 1992–93. In 1999–2000, again at the IAS, I was giving a series of lectures, and Pierre Deligne (Professor in the School of Mathematics) was taking notes and checking every step of my arguments. Only then did I discover that the proof of a key lemma in my paper contained a mistake and that the lemma, as stated, could not be salvaged. Fortunately, I was able to prove a weaker and more complicated lemma, which turned out to be sufficient for all applications. A corrected sequence of arguments was published in 2006.

This story got me scared. Starting from 1993, multiple groups of mathematicians studied my paper at seminars and used it in their work and none of them noticed the mistake. And it clearly was not an accident. A technical argument by a trusted author, which is hard to check and looks similar to arguments known to be correct, is hardly ever checked in detail.

But this is not the only problem that allows mistakes in mathematical texts to persist. In October 1998, Carlos Simpson submitted to the arXiv preprint server a paper called "Homotopy Types of Strict 3-groupoids." It claimed to provide an argument that implied that the main result of the "∞-groupoids" paper, which Kapranov and I had published in 1989, cannot be true. However, Kapranov and I had considered a similar critique ourselves and had convinced each other that it did not apply. I was sure that we were right until the fall of 2013 (!!).

I can see two factors that contributed to this outrageous situation: Simpson claimed to have constructed a counterexample, but he was not able to show where the mistake was in our paper. Because of this, it was not clear whether we made a mistake somewhere in our paper or he made a mistake somewhere in his counterexample. Mathematical research currently relies on a complex system of mutual trust based on reputations. By the time Simpson's paper appeared, both Kapranov and I had strong reputations. Simpson's paper created doubts in our result, which led to it being unused by other researchers, but no one came forward and challenged us on it.

Around the time that I discovered the mistake in my motivic paper, I was working on a new development, which I called 2-theories. As I was working on these ideas, I was getting more and more uncertain about how to proceed. The mathematics of 2-theories is an example of precisely that kind of higher-dimensional mathematics that Kapranov and I had dreamed about in 1989. And I really enjoyed discovering new structures that were not direct extensions of structures in lower dimensions.

But to do the work at the level of rigor and precision I felt was necessary would take an enormous amount of effort and would produce a text that would be very hard to read. And who would ensure that I did not forget something and did not make a mistake, if even the mistakes in much more simple arguments take years to uncover? I think it was at this moment that I largely stopped doing what is called "curiosity-driven research" and started to think seriously about the future. I didn't have the tools to explore the areas where curiosity was leading me and the areas that I considered to be of value and of interest and of beauty.

So I started to look into what I could do to create such tools. And it soon became clear that the only long-term solution was somehow to make it possible for me to use computers to verify my abstract, logical, and mathematical constructions. The software for doing this has been in development since the sixties. At the time, when I started to look for a practical proof assistant around 2000, I could not find any. There were several groups developing such systems, but none of them was in any way appropriate for the kind of mathematics for which I needed a system.

When I first started to explore the possibility, computer proof verification was almost a forbidden subject among mathematicians. A conversation about the need for computer proof assistants would invariably drift to Gödel's incompleteness theorem (which has nothing to do with the actual problem) or to one or two cases of verification of already existing proofs, which were used only to demonstrate how impractical the whole idea was.

Among the very few mathematicians who persisted in trying to advance the field of computer verification in mathematics during this time were Tom Hales and Carlos Simpson. Today, only a few years later, computer verification of proofs and of mathematical reasoning in general looks completely practical to many people who work on univalent foundations and homotopy type theory.

The primary challenge that needed to be addressed was that the foundations of mathematics were unprepared for the requirements of the task. Formulating mathematical reasoning in a language precise enough for a computer to follow meant using a foundational system of mathematics not as a standard of consistency to establish a few fundamental theorems, but as a tool that can be employed in everyday mathematical work. There were two main problems with the existing foundational systems, which made them inadequate. *Firstly*, existing foundations of mathematics were based on the languages of predicate logic and languages of

(Continued on page 9)

Vladimir Voevodsky, who joined the School of Mathematics as Professor in 2002, is known for his work in the homotopy theory of schemes, algebraic K-theory, and interrelations between algebraic geometry and algebraic topology. He made one of the most outstanding advances in algebraic geometry in the past few decades by developing new cohomology theories for algebraic varieties. Among the consequences of his work are the solutions of the Milnor and Bloch-Kato conjectures.

UNIVALENT FOUNDATIONS *(Continued from page 8)*

this class are too limited. *Secondly*, existing foundations could not be used to directly express statements about such objects as, for example, the ones in my work on 2-theories.

Still, it is extremely difficult to accept that mathematics is in need of a completely new foundation. Even many of the people who are directly connected with the advances in homotopy type theory are struggling with this idea. There is a good reason: the existing foundations of mathematics—ZFC and category theory—have been very successful. Overcoming the appeal of category theory as a candidate for new foundations of mathematics was for me personally the most challenging.

The story starts with ZFC: the Zermelo-Fraenkel theory with the axiom of choice. Since the first half of the twentieth century, mathematics has been presented as a science based on ZFC, and ZFC was introduced as a particular theory in predicate logic. Therefore, someone who wanted to get to the bottom of things in mathematics had a simple road to follow—learn what predicate logic is, then learn a particular theory called ZFC, then learn how to translate propositions about a few basic mathematical concepts into formulas of ZFC, and then learn to believe, through examples, that the rest of mathematics can be reduced to these few basic concepts.

From left to right: Voevodsky at a lunch seminar, pictured here with Jonathan Israel, Andrew W. Mellon Professor in the School of Historical Studies; delivering the lecture "What if Current Foundations of Mathematics are Inconsistent?" at the 80th anniversary celebration of the Schools of Mathematics and Natural Sciences in 2010; a lunchtime conversation this spring

FROM LEFT: DAN KOMODA, CLIFF MOORE, ANDREA KANE

This state of affairs was extremely beneficial for mathematics, and it is rightly credited for the great successes of abstract mathematics in the twentieth century. Historically, the first problems with ZFC could be seen in the decline of the great enterprise of early Bourbaki, which occurred because the main organizational ideas of mathematics of the second half of the twentieth century were based on category theory, and category theory could not be well presented in terms of ZFC. The successes of category theory inspired the idea that categories are "sets in the next dimension" and that the foundation of mathematics should be based on category theory or on its higher-dimensional analogues.

FORMULATING MATHEMATICAL REASONING IN A LANGUAGE PRECISE ENOUGH FOR A COMPUTER TO FOLLOW MEANT USING A FOUNDATIONAL SYSTEM OF MATHEMATICS AS A TOOL THAT CAN BE EMPLOYED IN EVERYDAY MATHEMATICAL WORK.

The greatest roadblock for me was the idea that categories are "sets in the next dimension." I clearly recall the feeling of a breakthrough that I experienced when I understood that this idea is wrong. Categories are not "sets in the next dimension." They are "partially ordered sets in the next dimension" and "sets in the next dimension" are groupoids.

This new perspective on "groupoids" and "categories" took some adjustment for me because I remember it being emphasized by people I learned mathematics from that one of the things that made Grothendieck's approach to algebraic geometry so successful was that he broke with the old-schoolers and insisted on the importance of considering all morphisms and not only isomorphisms. (Groupoids are often made of set-level objects and their isomorphisms, while categories are often made of set-level objects and *all* morphisms.)

Univalent foundations, like ZFC-based foundations and unlike category theory, is a complete foundational system, but it is very different from ZFC. To provide a format for comparison, let me suppose that any foundation for mathematics adequate both for human reasoning and for computer verification should have the following three components.

The *first component* is a formal deduction system: a language and rules of manipulating sentences in this language that are purely formal, such that a record of such manipulations can be verified by a computer program. The *second component* is a structure that provides a meaning to the sentences of this language in terms of mental objects intuitively comprehensible to humans. The *third component* is a structure that enables humans to encode mathematical ideas in terms of the objects directly associated with the language.

In ZFC-based foundations, the first component has two "layers." The first layer is a general mechanism for building deduction systems, which is called predicate logic; the second layer is a particular deduction system called ZFC obtained by applying this mechanism to a set of operations and axioms. The second component in ZFC is based on the human ability to intuitively comprehend hierarchies. In fact, the axioms of ZFC can be seen as a collection of properties that all hierarchies satisfy, together with the axiom of infinity, which postulates the existence of an infinite hierarchy. The third component is a way to encode mathematical notions in terms of hierarchies that starts with rules for encoding mathematical properties of sets. That is why ZFC is often called a set theory.

The original formal deduction system of univalent foundations is called the calculus of inductive constructions, or CIC. It was developed by Thierry Coquand and Christine Pauline around 1988 and was based on a combination of ideas from the theory and practice of computer languages with ideas in constructive mathematics. The key names associated with these ideas are Nicolaas Govert de Bruijn, Per Martin-Löf and Jean-Yves Girard. The formal deduction system of the proof assistant Coq is a direct descendant of CIC.

The second component of univalent foundations, the structure that provides a direct meaning to the sentences of CIC, is based on univalent models. The objects directly associated with sentences of CIC by these models are called homotopy types. The world of homotopy types is stratified by what we call h-levels, with types of h-level 1 corresponding to logical propositions and types of h-level 2 corresponding to sets. Our intuition about types of higher levels comes mostly from their connection with multidimensional shapes, which was studied by ZFC-based mathematics for several decades.

The third component of univalent foundations, a way to encode general mathematical notions in terms of homotopy types, is based on the reversal of Grothendieck's idea from the late seventies considered in our "∞-groupoids" paper. Both mathematically and philosophically, this is the deepest and least understood part of the story.

I have been working on the ideas that led to the discovery of univalent models since 2005 and gave the first public presentation on this subject at Ludwig-Maximilians-Universität München in November 2009. While I have constructed my models independently, advances in this direction started to appear as early as 1995 and are associated with Martin Hofmann, Thomas Streicher, Steve Awodey, and Michael Warren.

In the spring of 2010, I suggested to the School of Mathematics that I would organize a special program on new foundations of mathematics in 2012–13, despite the fact that at the time it was not clear that the field would be ready for such a program.

And I now do my mathematics with a proof assistant. I have a lot of wishes in terms of getting this proof assistant to work better, but at least I don't have to go home and worry about having made a mistake in my work. I know that if I did something, I did it, and I don't have to come back to it nor do I have to worry about my arguments being too complicated or about how to convince others that my arguments are correct. I can just trust the computer. There are many people in computer science who are contributing to our program, but most mathematicians still don't believe that it is a good idea. And I think that is very wrong.

I would like to thank all of those who are trying to understand the ideas of univalent foundations, who are developing these ideas, and who are trying to communicate these ideas to others. ■

The special program on univalent foundations that Vladimir Voevodsky organized at the Institute in 2012–13 resulted in a group of two-dozen mathematicians writing a six-hundred-page book in less than six months. The book is available freely at http://homotopytypetheory.org/book/. This article was adapted from a lecture given by Voevodsky in March; the video may be viewed at https://video.ias.edu/voevodsky14/.