Chapter 13
What Do We Want a Foundation to Do?

Comparing Set-Theoretic, Category-Theoretic, and Univalent Approaches

Penelope Maddy

Abstract It's often said that set theory provides a foundation for classical mathematics because every classical mathematical object can be modeled as a set and every classical mathematical theorem can be proved from the axioms of set theory. This is obviously a remarkable mathematical fact, but it isn't obvious what makes it 'foundational'. This paper begins with a taxonomy of the jobs set theory does that might reasonably be regarded as foundational. It then moves on to category-theoretic and univalent foundations, exploring to what extent they do these same jobs, and to what extent they might do other jobs also reasonably regarded as foundational.

Mainstream orthodoxy holds that set theory provides a foundation for contemporary pure mathematics. Critics of this view argue that category theory, or more recently univalent foundations, is better suited to this role. Some observers of this controversy suggest that it might be resolved by a better understanding of what a foundation is. Despite considerable sympathy to this line of thought, I'm skeptical of the unspoken assumption that there's an underlying concept of a 'foundation' up for analysis, that this analysis would properly guide our assessment of the various candidates. In contrast, it seems to me that the considerations the combatants offer against opponents and for their preferred candidates, as well as the roles each candidate actually or potentially succeeds in playing, reveal quite a number of different jobs that mathematicians want done. What matters is these jobs we want our theories to do and how well they do them. Whether any or all of them, jobs or theories, deserves to be called 'foundational' is really beside the point.

The forces behind the rise of set-theoretic foundations (in the late nineteenth and early twentieth centuries) and its subsequent accomplishments (as of the early twenty-first) are explored and assessed in §13.1. §13.2 turns to the criticisms lodged against set-theoretic foundations as category theory developed (in the 1940s) and the

P. Maddy (✉)
University of California, Irvine, Irvine, CA, USA
e-mail: pjmaddy@uci.edu

S. Centrone et al. (eds.), *Reflections on the Foundations of Mathematics*,
Synthese Library 407, https://doi.org/10.1007/978-3-030-15655-8_13

subsequent case for category-theoretic foundations (beginning in the 1960s). The current proposal for univalent foundations is examined in the concluding §13.3.

13.1 Set-Theoretic Foundations[1]

It's commonplace to note that all standard mathematical objects and structures can be modeled as sets and all standard mathematical theorems proved from the axioms of set theory[2] – indeed, familiarity may well have dulled our sense of just how remarkable this fact is. For our purposes, though, let me draw attention to another commonplace: when called upon to characterize the foundational role of set theory, many observers are content merely to remind us that mathematics can be embedded in set theory in this way. But simply repeating that this is so leaves our present questions untouched: what's the point of this embedding?, what need does it serve?, what foundational job does it do?

To answer these questions, we should look back at the conditions under which set theory arose in the first place. Over the course of the nineteenth century, mathematics had expanded dramatically in an extraordinary variety of directions. This great torrent of new mathematics brought with it a pair of epistemological losses, as the subject outstripped available thinking on what ultimately justifies mathematical claims. Early efforts to make good on those losses eventually needed support of their own, and it was at this point that set theory entered the picture. A quick sketch of these developments should help explain what jobs set theory was at least partly designed to do.

Consider first the case of geometry. From the diagrams of Euclidean times to Kant's late eighteenth century theory of spatial intuition, geometry was generally understood as grounded in some variety of visualization or intuition. That changed in the nineteenth century with the introduction of 'points at infinity' (where parallel lines meet) and 'imaginary points' (with complex numbers as coordinates). There was no denying the fruitfulness of regarding geometry from this new perspective, but the imaginary points at which two disjoint circles 'intersect' can hardly be visualized or intuited! So this is the first epistemic loss: visualization and intuition were no longer adequate measures of correctness for the brave new geometry. What justification, then, could be given for admitting these new, invisible points, what guarantee that they wouldn't severely compromise the whole subject? Geometers were understandably queasy about this expansion of the proper domain of their inquiry.

[1]Many themes of this section are explored in more detail, with sources, in §I of [2017].

[2]Items like the category of all groups or the category of all categories are exceptions. There is no set of all groups or set of all categories for the same reason that there's no set of all sets: sets are formed in a transfinite series of stages, and there's no stage at which all of them (or all of them that are groups or all of them that are categories) are available to be collected. The category-theoretic cases are explored in the next section.

The second epistemic loss came with the rise of pure mathematics during this same period.[3] Up through the eighteenth century, there was no distinction between pure and applied; mathematics was considered the study of the mathematical structure literally present in the physical world. Galileo, Newton, Euler, Fourier, and others took the goal of natural science to be the isolation of purely mathematical laws governing the behavior of observable phenomena (e.g., planetary motions or the distribution of heat in a body) without appeal to hidden causes (e.g., Cartesian vortices or caloric fluids). This strategy was a tremendously successful at the time, encouraging precise mathematization and eschewing dubious mechanical explanations. The ongoing clash between mathematization of observable behavior and causal explanation re-emerged in the late nineteenth century in the study of thermodynamics: descendants of the purely descriptive, mathematical tradition, proposed the experientially exception-less second law, that entropy can only increase, while descendants of the causal, explanatory tradition developed the kinetic theory, according to which a decrease in entropy is just highly unlikely. In the early years of the twentieth century, the tables of history turned: kinetic theory with its atoms and molecules in random motion was experimentally confirmed. This meant that the laws of classical thermodynamics were revealed to be merely probable, and more generally, that the many hard-won differential equations of the eighteenth and nineteenth centuries were highly effective approximations, smoothed-out versions of a more complex, discrete microstructure.

By the end of the nineteenth and beginning of the twentieth centuries, as pure mathematics proliferated and applied mathematics lost its claim to literal truth, it became clear that mathematics isn't actually in the business of discerning the precise formal structure of the physical world. Rather, it provides an array of abstract models for the scientist to choose from for any particular descriptive job. Most of these mathematical descriptions involve explicit idealizations and approximations, and applied mathematicians expend considerable effort on explaining how and why these non-literal representations are nonetheless effective, often within a limited range of cases. Some such descriptions are effective despite our inability to explain exactly what worldly features they're tracking (e.g., in quantum mechanics). These are crucial morals for the philosophy of science, but our concern here is the epistemic loss suffered by mathematics itself: in the new, pure mathematics, there was no place for physical interpretation or physical insight to guide developments. In Euler's day, a scientist's feel for the physical situation could help shape the mathematics, could constrain it to effective paths even when rigor was in short supply. Now that mathematicians had declared their independence – their freedom to pursue whatever paths caught their purely mathematical interest – physical results and physical intuition could no longer serve to support or justify mathematical work. Without this guidance, how were mathematicians to tell which among the proliferation of new, purely mathematical abstract structures under were trustworthy, legitimate, worthwhile?

[3] For more on this development, see [2008] or chapter 1 of [2011].

In the face of these epistemic set-backs, two types of remedies were offered. The first, in response to queasiness about the new, un-intuitable geometric points, came in the mid-nineteenth century, when Karl von Staudt managed to build surrogates for the suspicious entities out of straightforwardly acceptable materials. So, for example, a point at infinity where two given parallel lines meet can be identified with the collection ('pencil') of all lines parallel to those two, and this point at infinity taken to be on a given line if the line is in that collection.[4] In this way, a previously suspicious mathematical development is domesticated by building it up from previously justified mathematics. This method was widely used, but eventually a new question has to be faced: which means of building new from old are trustworthy, and why?

The second remedy, in response to the proliferation of abstract structures, came later in the nineteenth century, with Hilbert's axiomatic method: each proposed structure should be rigorously axiomatized; if that axiomatization is coherent, the structure is legitimate. Though careful isolation of appropriate axioms might reveal unnoticed incoherence in some cases, for most others a new worry is immediate: how do we tell which axiomatizations are coherent?[5] A second concern is less obvious, but also serious. Suppose I've devised an axiom system for the natural numbers, another for the real numbers, another for analysis, another for point-set topology, another for computable functions, and so on – and I prove theorems in each of these. Can I use the theorems I've proved in one area to prove theorems in another? John Burgess illustrates how centrally modern mathematics replies on the ability to move easily between its branches:

> There is the interconnectedness of the different branches of mathematics, a phenomenon evident since the seventeenth century in the use of coordinate methods, but vastly expanded in the nineteenth century. With the group concept, an idea originating in algebra is applied to geometry. With 'functional analysis', ideas originating in geometry or topology are applied to analysis, as functions come to be considered 'points' in an auxiliary space, and operations like differentiation and integration come to be considered 'transformations' of that space. (Footnote: One reason one needs to allow in pathological functions like the Riemann-Weierstrass examples is in order to achieve a certain 'completeness', analogous to the completeness of the real number-line, in the 'space' of functions.) And so on across the whole of mathematics.
>
> Interconnectedness implies that *it will no longer be sufficient to put each individual branch of mathematics separately on a rigorous basis.* (Burgess 2015, pp. 59–60, emphasis in the original)

Today it's hard to see how Wiles could have proved Fermat's Last Theorem if he'd been confined to one or another of the individual axiom systems!

[4] Readers of Frege (1884) will recognize this as 'the direction of a line' and recall how it serves as Frege's model for identifying a natural number with a collection of equinumerous collections.

[5] This was before the development of formal languages and deductive systems, before a clear understanding of consistency, satisfiability, and of course, before Gödel's completeness and incompleteness theorems.

The epistemic and methodological questions raised by these nineteenth-century losses and their partial remedies permeated the climate in which set theory first arose. Though much of the initial motivation for introducing sets was purely mathematical – Cantor, for example, was just trying to generalize his theorem on the uniqueness of trigonometric representations – it soon became clear that the various building methods of von Staudt and the rest were all essentially set-theoretic in character, all derivable from a few elementary set-theoretic operations (like taking subsets, intersections and unions, cross-products, and power sets). Indeed it turned out that all the various items previously axiomatized in separate systems (natural numbers, real numbers,[6] analysis, etc.) could be constructed set-theoretically – the beginnings, in other words, of the famous embedding of mathematics in set theory.

In this way, set theory made progress on our first two questions: the trustworthy building methods are those of set theory; the coherent axiom systems are those that can be modeled by sets. Of course this is cold comfort unless we know that set theory itself is reliable – a particularly dubious proposition at the time, given both the paradoxes and wide-spread debates over fundamentals (the well-ordering principle, the axiom of choice, the continuum hypothesis, etc.). Working in the Hilbertian tradition, Zermelo set out to axiomatize the subject, successfully isolating the basic assumptions underlying the informal practice while forestalling the known routes to paradox. Though he hoped to include a consistency proof in his original presentation, Hilbert encouraged him to publish the axioms first and follow with the consistency proof when it was ready. Years later it became clear what good advice this was, when Gödel showed that only an stronger system could prove the consistency of Zermelo's axioms (assuming they are consistent).

So, much as we might like to have an iron-clad guarantee of the consistency of set theory, and thus of the trustworthiness of the methods embedded therein, this is a forlorn hope; all we get is the assurance that the embedded methods are no more dangerous than Zermelo's set theory. Most estimates of that danger have decreased substantially over the intervening century, with the development of a compelling intuitive picture of the universe of sets (the iterative conception), a finely-articulated model within set theory (the constructible universe), and a vast, intricate and far-reaching mathematical theory with no sign of contradiction. Meanwhile, various levels of consistency strength have been delineated and explored – from relatively weak subsystems of second-order arithmetic to ever-larger large cardinal axioms –

[6]It's worth noting that Dedekind's set-theoretic construction of the reals was different in character from von Staudt's construction of imaginary points. Von Staudt was faced with a practice in good working order, but with questionable posits. Dedekind was faced with a defective practice (basic theorems of the calculus couldn't be proved). So von Staudt's challenge was to remove queasiness about the posits by domesticating them, while Dedekind's was to produce a more precise replacement that would both conform to previous practice and extend it (proving those basic theorems). Thus Dedekind's construction had a different, plausibly 'foundational' function (called **Elucidation** in [2017]). As both category-theoretic and univalent foundations are content to relegate **Elucidation** to ETCS, a weak category-theoretic theory of collections (see [2017], §II, and UFP (2013), p. 8, respectively), it won't figure in the comparative analysis here.

yielding a hierarchy that's now routinely used to calibrate the level of danger a proposed theory presents. Presumably the ability to assess these risks is something mathematicians value. Insofar as we're inclined to regard **Risk Assessment** as a 'foundational' virtue, this is one foundational job that contemporary set theory does quite well.

Finally, the other question raised by the axiomatic response to the loss of older forms of justification (intuition/visualization, physical interpretation/insight) concerned the interrelations between the various branches of pure mathematics: if each branch is characterized by its own separate list of axioms, how can work in one branch be brought to bear in another?

> To guarantee that rigor is not compromised in the process of transferring material from one branch of mathematics to another, it is essential that the starting points of the branches being connected should ... be compatible. ... The only obvious way to ensure compatibility of the starting points ... is ultimately to derive all branches from a common, unified starting point. (Burgess 2015, pp. 61–62)

This 'common, unified starting point' emerges when the various branches are all embedded in a single theory of sets, when all theorems are treated as theorems in the same system. In this way, set theory provides a **Generous Arena** where all of modern mathematics takes place side-by-side and a **Shared Standard** of what counts as a legitimate construction or proof. These are the striking achievements of the well-known embedding of mathematics in set theory. Insofar as they fairly count as 'foundational', set theory is playing two more crucial foundational roles.

Let's pause a moment to notice that everything claimed so far on set theory's behalf has been at the level of straightforward mathematical benefits: the embedding of mathematics in set theory allows us to assess the risk of our theories, to bring results and techniques one branch of mathematics to bear on concepts and problems in another, and to agree on standards of construction and proof. Some observers, especially philosophers, have been tempted to draw – in addition – strong metaphysical or epistemological morals: we've discovered that all mathematical entities were really sets all along, or that our knowledge of mathematics is reducible to our knowledge of sets.[7] These further claims might rightly be called 'foundational', too, but they're also controversial, to say the least. For mathematical purposes, the metaphysical claim is beside the point: it doesn't matter whether we say the von Neumann ordinals are the numbers or the von Neumann ordinals can serve as fully effective mathematical surrogates for the numbers. As for the epistemological claim, it's just false: however it is that we know the things we know in the various, far-flung branches of mathematics, it isn't by deriving them from the axioms of set theory. Most of the time, it's our conviction that the mathematics is correct that makes us think there must be a formal proof from those axioms!

While dubious philosophical claims like these are unlikely to affect practice, other intrusions of irrelevant ontological thinking might come uncomfortably close.

[7]These are the spurious foundational virtues called **Metaphysical Insight** and **Epistemic Source** in [2017].

Quite generally, if we take the claim that set theory determines the ontology of mathematics too seriously, we might be tempted to think of it as 'the final court of appeal', charged with passing stern judgement on new mathematical avenues. In fact, I think this gets the situation backwards: casting set theory as the **Generous Arena** isn't intended to limit mathematics; rather it places a heavy responsibility on set theory to be as generous as possible in the types of structure whose existence it implies.[8] This admonition to maximize is one of the most fundamental and powerful methodological principles guiding the development of set theory. If we imagine, in our overly philosophical mood, that set theory has some kind of special access to the metaphysical facts about what abstracta exist, then we might be tempted to put the onus on mathematical practice to conform to the dictates of set theory, to raise that special access above informed judgements of mathematical advantage. I trust we can all agree that this would be a grave mistake.

A more subtle danger in the same general direction arises from the fact that our embedding of mathematics in set theory is more like von Staudt's approach than Hilbert's: a surrogate for the mathematical item in question is constructed by set-theoretic means, as an item in V, the set-theoretic universe; it's not enough, as the Hilbertian would have it, that there's a model somewhere in V that thinks there is such an item. A simple example would be a proof of $1 = 0$ from the axioms of (first-order) Peano Arithmetic: PA + not-Con(PA) is consistent (assuming PA is), so it has a model that thinks there's a proof of $1 = 0$ from PA; but viewed set-theoretically, that model is benighted, the thing it takes for a proof of $1 = 0$ has non-standard length, isn't really a proof. For a more interesting example, consider a definable[9] well-ordering of the real numbers. There is such an ordering in Gödel's inner model, the constructible universe L, but if we add large cardinal axioms to our list, as many set theorists these days do, then that model is benighted: the thing it takes for a well-ordering of the reals only orders the reals present in L; in fact, there is no definable well-ordering of all the reals.

Speaking loosely, we might express this by saying that the inconsistency proof and the well-ordering exist on the Hilbertian standard, while on the von Staudtian set-theoretic standard, they don't. This way of talking is expressive and largely benign, but it can lead us astray if we forget that it's figurative, if we fall into taking it too literally. We need to bear in mind that the cash value of 'these things exist in V' is just 'the existence of (surrogates for) these things can be proved from the axioms of set theory' – a straightforward manifestation of set theory's role as **Shared Standard** of proof. To say that 'the universe of sets is the ontology of mathematics' amounts to claiming that the axioms of set theory imply the existence of (surrogates for) all the entities of classical mathematics – a simple affirmation of set theory's role as **Generous Arena**.

[8]The underlying methodological maxim here is to prefer non-restrictive, maximizing theories. [1997] concludes with an early attempt to formalize this notion. Various developments of this idea and alternatives to it have been suggested, but the problem remains open.

[9]That is, projectively definable.

The danger in taking figurative ontological talk too seriously is that it can lead to a sort of rigidity in practice. Consider that definable well-ordering of the reals. Suppose a pure mathematician has a clever and fruitful approach to a certain problem, or an applied mathematician has a way to effectively model some physical situation, by means of such an ordering. If we believe that set theory is the 'ontology' of mathematics, 'the final court of ontological appeal', we'll be tempted to say 'tough luck, it might be nice if there were such a thing, but there isn't'. But this seems wrong. Both mathematicians' activities can be carried out inside L – by which we mean, in set theory with $V = L$ as an additional axiom. Since that theory includes the standard axioms, it provides a fairly **Generous Arena** all by itself: the usual constructions and techniques are ready to hand; to speak in the figurative idiom, L is a pretty good place to do mathematics. The disadvantage is that results proved using $V = L$ can't automatically be exported to other areas of mathematics, and results from other areas that depend on large cardinals can't automatically be imported. But as long as these import/export restrictions are observed, as long as the use of axioms beyond the standard ones is carefully flagged, there's no reason to rule out these developments. The pure mathematician's work on her problem is simply part of the investigation of L, a particularly important part of V; the applied mathematician has determined that it's most effective to model his physical situation in L rather than V.

This leaves us with a tempered version of the von Staudian 'final court of ontological appeal': the axioms for our **Generous Arena**, which constitute our **Shared Standard** of proof, include the usual axioms – plus some others, beginning with large cardinals, that add to their generosity – but these can be temporarily adjusted for mathematical or scientific purposes with suitable import/export restrictions. Once we reject the idea that the choice of a fundamental theory to do these foundational jobs is a matter of determining the 'true mathematical ontology', once we focus instead on the literal mathematical content of our decisions, we come to see that we can and should allow some wiggle room for both pure and applied mathematicians to work in well-motivated variants of the fundamental theory. I won't attempt to explicate what counts as 'well-motivated' – this requires the sound judgment of insightful practitioners – but one clear qualifier is the existence of an attractive, well-understood model inside V,[10] as in the case of L and $V = L$.[11] Though this marks a slight adjustment to strict von-Staudism, it's still very far from full Hilbertism, where any consistent theory as good as any other – precious few such theories can deliver a pure mathematical theorem worth proving or an applied mathematical model amenable to actual use.[12]

[10] This is, the existence of such a model can be proved from the fundamental axioms.

[11] Another well-known example is the theory ZF + $V = L(R)$ + AD. Again, separating the 'mathematically worthy' from the unworthy no doubt requires keen mathematical discernment and well-informed good judgement.

[12] For successful application, it's not enough that our theory prove the existence of a suitable structure; it must exist in a context with enough mathematical tools to study and manipulate that structure. See [2011], pp. 90–96, for a related discussion.

pjmaddy@uci.edu

One last point. Returning once more to the historical development of set theory, Zermelo's axioms were soon supplemented with replacement and foundation, and his imprecise notion of 'definite property' was sharpened to 'formula in the first-order language of set theory'. This generated what we now know as the formal theory ZFC. At that point, the embedding of mathematics in set theory came to serve yet another purpose: once mathematics was successfully encoded in a list of formal sentences, meta-mathematical tools could be brought to bear to prove theorems about its general features. Among the greatest of these results were those of Gödel – classical mathematics, if consistent, can't prove its own consistency or the negation of the Continuum Hypothesis – and Cohen – or the Continuum Hypothesis itself. Here set theory provides a **Meta-mathematical Corral**, tracing the vast reaches of mathematics to a set of axioms so simple that they can then be studied formally with remarkable success. Perhaps this accomplishment, too, has some claim to the honorific 'foundational'.

So my suggestion is that we replace the claim that set theory is a (or 'the') foundation for mathematics with a handful of more precise observations: set theory provides **Risk Assessment** for mathematical theories, a **Generous Arena** where the branches of mathematics can be pursued in a unified setting with a **Shared Standard** of proof, and a **Meta-mathematical Corral** so that formal techniques can be applied to all of mathematics at once. I haven't offered any argument that these accomplishments must be understood to be 'foundational', but it seems to me consistent with the ordinary use of the term to so apply it. I take it for granted that these accomplishments are of obvious mathematical value, whatever we decide about the proper use of the term 'foundational'.

Let's now turn to two of set theory's purported rivals: first category-theoretic foundations, then univalent foundations.

13.2 Category-Theoretic Foundations[13]

By the end of the 1930s, ZFC had been codified in its first-order form and its role as **Generous Arena, Shared Standard, Meta-mathematical Corral**, and in **Risk Assessment** were widely accepted. Soon thereafter, mathematical pressures in abstract algebra gave rise to category theory, and category theorists began to criticize set theory as a 'foundation'. By the 1960s, category theory was being proposed as alternative to set theory that could overcome these weaknesses. A look at the objections raised and the solutions offered should help us determine what jobs the critics thought a 'foundation' was supposed to do.

So, what was wrong with set-theoretic foundations? The first objection is that category theory deals with unlimited categories, like the category of all groups or the category of all categories or the category of all mathematical X's, but nothing

[13]For more on many themes of this section, with sources, see §II of [2017].

is vast, its construction techniques wildly indiscriminate, so it includes hordes of useless structures and – this is the important point – no way of telling the mathematically promising structures from the rest. Furthermore, the set-theoretic surrogates have lots of extraneous structure, artifacts of the way they're constructed. Here the hope was to find a foundation that would guide mathematicians toward the important structures and characterize them strictly in terms of their mathematically essential features. Such a foundation would actually be useful to mainstream mathematicians in their day-to-day work, not remote, largely irrelevant, like set theory; it would provide **Essential Guidance**. Proponents held that this is precisely what category theory had done for algebraic geometry and algebraic topology.

Now it could be that some over-zealous partisan of set-theoretic foundations at one time or another claimed that mathematics would be better off if all mathematicians thought like set theorists, but as far as I can tell, this was never one of the foundational jobs that set theory was seriously proposed to do. No reasonable observer would suggest that an algebraic geometer or algebraic topologist would do better to think in set-theoretic rather than category-theoretic terms. But it seems equally unreasonable to suggest that an analyst, or for that matter a set theorist, would do better to think in category-theoretic terms.[17] What's intriguing here is that proponents of category-theoretic 'foundations' would apparently agree. Mac Lane, for example, writes:

> Categories and functors are everywhere in topology and in parts of algebra, but they do not yet relate very well to most of analysis.
>
> We conclude that there is as yet no simple and adequate way of conceptually organizing all of Mathematics. (Mac Lane 1986, p. 407)

If a 'foundation' is to reveal the underlying essence, the conceptual core, omit all irrelevancies, and guide productive research, then it's unlikely that it can encompass all areas of mathematics. Faced with this tension between **Essential Guidance** and **Generous Arena**, Mac Lane seems willing to forego **Generous Arena**, and with it presumably **Shared Standard** and **Meta-Mathematical Corral**.

This preference is more-or-less explicit in the theory of categories that's proposed as our fundamental foundation. The 'Category of Categories as a Foundation' (CCAF) was introduced by Lawvere in the 1960s and subsequently improved by McLarty in the 1990s. CCAF is a actually a minimal background theory which is then supplemented as needed to guarantee the existence of particular categories for this or that area of mathematics. One such special category is 'The Elementary Theory of the Category of Sets' (ETCS), which codifies a relatively weak theory of collections (ZC with bounded separation). Collections in this sense are understood in a natural way in terms of their arrows rather than their elements, but to gain a category-theoretic set theory with sufficient strength for, say, **Risk Assessment**,

[17] See, e.g., the work of Mathias discussed in [2017].

more characteristically set-theoretic notions have to be translated in from outside.[18] A category for synthetic differential geometry is another example that could be added with a suitable axiom. As might be expected from the Hilbertian flavor of this approach, it isn't conducive to **Generous Arena**.

So despite the rhetoric – pitting category theory against set theory, proposing to replace set-theoretic foundations with category-theoretic foundations – the two schools are aimed at quite different goals. Set theory provides **Risk Assessment**, **Generous Arena**, **Shared Standard**, and **Meta-mathematical Corral**, and it apparently continues to do these jobs even in the context of category-theoretic foundations. What category theory offers is **Essential Guidance**, but only for those branches of mathematics of roughly algebraic character. I have no objection to calling this a 'foundational' achievement, so long as it isn't taken to supersede the other foundational goals explored here. What category theory has accomplished – however this achievement is labeled – is a way of thinking about a large part of mathematics, of organizing and understanding it, that's been immensely fruitful in practice. Proponents of set-theoretic foundations should have nothing but admiration for this achievement. It raises deep and important methodological questions about which 'ways of thinking' are effective for which areas of mathematics, about how they differ, about what makes them so effective where they are and ineffective where they aren't, and so on.

So, should we regard set theory's range of accomplishments for mathematics in general as more 'foundational' than category-theory's conceptual achievements across several important areas of the subject, or vice versa? I confess that this doesn't strike me as a productive debate. In contrast, a concerted study of the methodological questions raised by category theory's focus on providing a fruitful 'way of thinking' would almost certainly increase our fundamental understanding of mathematics itself. I vote for that.

13.3 Univalent Foundations

With these nineteenth and twentieth century developments in the background, the turn of the 21st brought a new critique of set-theoretic foundations and a new proposal for its replacement. Like set theory and category theory, this more recent effort also arose out of ongoing mathematical practice. The mathematics involved this time is homotopy theory, which, like category theory, has its roots in abstract algebra; proponents of the subject describe it as 'an outgrowth of algebraic topology and homological algebra, with relationships to higher category theory' (UFP 2013, p. 1). The program of univalent foundations involves using homotopy theory to interpret Martin-Löf's type theory, then adding the so-called 'Univalence Axiom' –

[18]Of course set theory also translates notions from outside when locating their surrogates, but set theory isn't claiming to provide **Essential Guidance**.

which has the effect, understood roughly, of identifying isomorphic structures.[19] The result is declared to be 'incompatible with conventional [presumably, set-theoretic and category-theoretic] foundations' (Awodey (2014), p. 1) and to provide 'a completely new foundation' (Voevodsky 2014a, b, p. 9).

We've seen that set-theoretic foundations arose in response to the serious practical questions in the wake of the profound shift from mathematics as a theory of the world to mathematics as a pure subject in its own right. In contrast, category-theoretic practice was functioning well enough with Grothendieck's understanding; the impetus this time came from the hope for truly unlimited categories (misconstrued at the time as a shortcoming of set-theoretic foundations) and the promise that category theory could do a new and different foundational job (**Essential Guidance**). Univalent foundations takes a page from each book: there was a real practical problem to be addressed, and addressing it introduced a new foundational goal. Let me explain.

Grothendieck's work in category theory was already so complex that 'the intellectual faculties are being strained to their uttermost limit' (Burgess 2015, p. 176), and as younger mathematicians pushed these ideas further, there was some evidence those limits had been breached. Vladimir Voevodsky, one of the leaders in this development and the originator of univalent foundations, describes how the troubles began:

> 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 the claims of the paper were left unsubstantiated.
>
> The 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. (Voevodsky 2014a, p. 8)

Soon, a similar problem hit closer to home. In 1999–2000, Voevodsky lectured at Princeton's Institute for Advanced Study on an approach to motivic cohomology that he, Suslin, and Eric Friedlander had developed, an approach based on earlier work of Voevodsky. That earlier work was written while the jury was still out on Bloch's lemma, so necessarily did without it. As the lectures progressed, the details were carefully scrutinized.

> Only then did I discover that the proof of a key lemma in my [earlier] 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. (ibid.)

Perhaps even worse, in 1998 a counterexample was reported to a 1989 paper of Michael Kaparonov and Voevodsky, but because of the complexities involved, Voevodsky reports that he didn't believe it himself until 2013!

[19] See Awodey (2014, p. 1).

It's easy to sympathize with the cumulative effect of these mishaps on Voevodsky: 'This ... got me scared' (ibid.). It became hard to ignore the fact that proofs in this area were so complex as to be prone to hidden glitches, a worry exacerbated by the further fact that correcting these glitches made the proofs even more complex. To top off the anxiety, at this point Voevodsky was hoping to push even further, into something new he called '2-theories'.

> But to do the work at the level of rigor and precision I felt 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? (Voevodsky 2014a, p. 8)

This, then, is the pressing new problem faced by mathematical practioners in this field: how can we be confident that our proofs are correct? To this point, various sociological checks had been enough – proofs were carefully examined by the community; mathematicians of high reputation were generally reliable; and so on – but those checks had apparently been outstripped.

The need to address this problem gives rise to a new goal – a systematic method for **Proof Checking** – and it seems reasonable to classify this goal, too, as 'foundational'. As we've seen, set-theoretic foundations originated in the embedding of standard mathematics in set theory. For this purpose, as Voevodsky puts it, all we need is to

> ... learn how to translate propositions about a few basic mathematical concepts into formulates of ZFC, and then learn to believe, through examples, that the rest of mathematics can be reduced to these few basic concepts. (Voevodsky 2014a, p. 9)

Here we have the embedding expressed in formal terms. Despite its meta-mathematical virtues, this formal system isn't one in which any mathematician would actually want to prove anything; in fact (as noted earlier), our confidence that there is a formal proof is usually based on our confidence in the informal proof, combined with our informed belief that all informal proofs can be formalized in this way. The demands of **Proof Checking** are quite different: we need a system that can represent actual proofs, 'a tool that can be employed in everyday mathematical work' (Voevodsky 2014a, p. 8).[20]

[20]Awodey traces the roots of univalent foundations in traditional foundational work to Frege rather than Zermelo: 'this new kind of ... formalization could become a practical tool for the working mathematician – just as originally envisaged by Frege, who compared the invention of his Begriffsschrift with that of the microscope, (Awodey 2016a, p. 8, see also Awodey and Coquand (2013, p. 6). While Frege does make this comparison, it involves a contrast between the microscope and the eye: 'because of the range of its possible uses and the versatility with which it can adapt to the most diverse circumstances, the eye is far superior to the microscope' (Frege 1879, p. 6). Frege's formal system 'is a device invented for certain scientific purposes, and one must not condemn it because it is not suited to others' (ibid.). The 'scientific purpose' in question is to determine whether arithmetic can be derived by pure logic; the Begriffsschrift was needed 'to prevent anything intuitive from penetrating here unnoticed ... to keep the chain of inferences free of gaps' (ibid., p. 5). It seems to me likely that Awodey's 'practical tool for the working

Now there are actually several proof checking technologies on offer these days, some even based on set theory. In his contribution to this volume, Paulson touches on a range of options and remarks that 'every formal calculus ... will do some things well, other things badly and many other things not at all' (Paulson 2019, Chap. 20, pp. 437–453). The proponents of univalent foundations have their own preferred system, combining ideas from Martin-Löf's type theory with insights from the study of computer languages – a system called 'the calculus of inductive constructions' (CIC). The project is to express ordinary mathematical reasoning in these terms – a process that might 'become as natural as typesetting ... papers in TeX' (UFP 2013, p. 10) – and to apply the associated proof assistant (Coq) to mechanically check the validity of those arguments.

Obviously this is a heady undertaking, still in its early stages,[21] but the ambitions of these theorists go beyond the original goal of testing the complex arguments of homotopy theory: Voevodsky holds that univalent foundations, 'like ZFC-based foundations and unlike category theory, is a complete foundational system' (Voevodsky 2014a, p. 9).[22] By this he means that both set-theoretic and univalent foundations aim to provide three things:

(1) a formal language and rules of deduction: first-order logic with the axioms of set theory, on the one hand; the aforementioned deductive system CIC, on the other.
(2) an intuitive interpretation of this deductive system: the iterative hierarchy, on the one hand; homotopy types, on the other.[23]
(3) a method for encoding mathematics: the well-known embedding of mathematics in set theory, on the one hand; an encoding in homotopy types on the other.

The presence of (3) indicates that **Generous Arena** and **Shared Standard** are goals of univalent foundations, though Voevodsky admits that 'this is ... the least understood part of the story' (Voevodsky 2014a, p. 9).

The question that needs answering is whether this encoding in homotopy types is like set theory's proof that there is a set-theoretic surrogate or like category theory's postulation of a category with the desired features – recalling von Staudt vs. Hilbert – as only the former serves to unite the encodings in a single **Generous Arena**. There's probably an easy answer to this question, but if so, it's unknown to me. Voevodsky's strong analogy between set-theoretic and univalent foundations, summarized above, suggests the former; while some of Awodey's remarks appear to lean toward to latter. The move to univalent foundations, Awodey writes,

---

mathematician' would be analogous to the eye, not the microscope, that serving as such a practical tool is one of those purposes for which the microscope and Frege's formal system are 'not suited'.

[21]Cf. UFP 2013, p. 2: 'univalent foundations is very much a work in progress'.

[22]I don't know what Voevodsky finds lacking in category-theoretic foundations – perhaps that it fails to provide a **Generous Arena**?

[23]Interestingly, Voevodsky (2014a) observes that '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' (p. 9).

> ... has the practical effect of simplifying and shortening many proofs by taking advantage of a more axiomatic approach, as opposed to more laborious analytic [e.g., set-theoretic] constructions. (Awodey 2016b, p. 3)

In a footnote, Awodey alludes to Russell's famous remark about 'the advantages of theft over honest toil' (ibid.).

In broad outline, it appears that the foundational theory into which mathematics is to be embedded begins by postulating a countable hierarchy of 'universes' (UFP 2013, p. 549) that obey a series of 'rules' (ibid., pp. 549–552). To this 'type theory', we add three axioms of homotopy theory: function extensionality, univalence, and higher inductive types (ibid., §A.3). Set theory, for example, is encoded as the category of all the 0-types in one or another of these universes, together with the maps between them (ibid., pp. 398, 438). So far, this looks more like honest toil than like theft. But to get even to ETCS, we have to add the axiom of choice, which incidentally brings with it the law of excluded middle (ibid., §10.1.5). If we simply assert that there is such a category, our procedure begins to look more like the axiomatic method of category-theoretic foundations – start with CCAF and add axioms as needed, asserting the existence of individual categories with the desired features for the various areas of mathematics – and we've seen that this sort of approach doesn't even aim for a **Generous Arena**. I'm in no position to assess how far univalent foundations extends in this direction – whether these are minor variations that can be handled with careful import/export restrictions or something more Hilbert-like – so I leave this as a question to its proponents: is your theory intended to provide a **Generous Arena** for all branches of mathematics and a **Shared Standard** of proof – and if so, how?

Whatever the answer to this question may be, further doubts on the viability of univalent foundations for **Generous Arena** and **Shared Standard** arise when we consider **Essential Guidance**, the key new foundational goal of category-theoretic foundations. Following the category theorists, Voevodsky seems to endorse this goal: he holds that 'the main organizational ideas of mathematics of the second half of the 20th century were based on category theory' (Voevodsky 2014a, p. 9); seeks 'a tool that can be employed in everyday mathematical work' (ibid., p. 8); and counts set theory's failure in these areas against its suitability as a foundation.[24] So, for example, it isn't enough that we find a way to embed set theory in the theory of homotopy types; we need to find a way that reveals the true nature of the subject, unlike ZFC:

> The notion of set ... is fundamental for mathematics. ... However, the theory of sets [has] never been successfully formalized. ... The formal theory ZFC ... is not an adequate

[24] Similarly, Awodey bemoans the 'serious mismatch between the everyday practice of mathematics and the official foundations of mathematics in ZFC' (Awodey 2016a, p. 2) and connects univalent foundations with structuralist tendencies in the philosophy of mathematics that frown on the extraneous features of set-theoretic surrogates.

> formalization of the set theory which is used in mathematics. (Voevodsky 2014b, lecture 2, slides 21–22)[25]

Voevodsky takes this to be accomplished in the new foundation:

> As part of Univalent Foundations we now have a formalization of set theory in the form of the theory of types of h-level 2 in MLTT [i.e., Martin-Löf type theory].[26] I believe that this is the first adequate formalization of the set theory that is used in pure mathematics. (Ibid, lecture 3, slide 11)[27]

Set theorists would most likely dispute this claim,[28] but for our purposes, what matters is that the goal of **Essential Guidance** is more or less explicit. And as we've seen, it seems unlikely that any one way of thinking is best for all areas of mathematics, so aiming for **Essential Guidance** tends to undercut **Generous Arena** and **Shared Standard**.

So, given that **Generous Arena** and **Shared Standard** are once again threatened by **Essential Guidance**, likely to return to the province of set theory, what of the other foundational goals? Speaking of the new formal system, Voevodsky remarks

> Currently we are developing new type theories more complicated than the standard Martin-Löf type theory and at the same time more convenient for practical formalization of complex mathematics. Such type theories may easily have over a hundred derivation rules. (Voevodsky 2013, slide 18)

Notice again the contrast with formalized ZFC. The first-order logic used there is designed to be a simple as possible, with as few formation and inference rules as possible, facilitating meta-mathematical study of theories expressed therein. Because the system of univalent foundations is designed to be as natural as possible a format for actual mathematical reasoning, it ends up being considerably more complex, so the goal of **Metamathematical Corral** presumably also remains with set theory. Furthermore, the complexity of univalent foundations leaves the question of consistency unsettled, much as in the early days of pure mathematics, and the solution is the same:

> Thus a careful and formalizable approach is needed to show that the newly constructed type theory is at least as consistent as ZFC with a given structure of universes [that is, with inaccessibles]. (Voevodsky, ibid.)

In other words, the role of 'a foundational system ... as a standard of consistency' (Voevodsky 2014a, p. 8) – **Risk Assessment** – also falls to set theory.[29]

---

[25]Cf. Awodey and Coquand 2013, p. 6: 'the fundamental notion of a set ... in univalent foundations turns out to be definable in more primitive terms'.

[26]Colin Mclarty was kind enough to explain to me that 'types of h-level 2' is just a different terminology for the '0-types in one or another of these universes' in the previous paragraph.

[27]Cf. UFP (2013, p. 9).

[28]I'm not sure what these thinkers take to be wrong with ZFC, but it could be something akin to the category-theorist's conviction that a neutral notion of 'collection' is better understood in top-down function-based terms (as in ETCS) rather than bottom-up element-based terms (as in ZFC).

[29]See also UFP (2013, p. 15).

To sum up, then, **Risk Assessment**, **Metamathematical Corral**, **Generous Arena**, and **Shared Standard** all appear to continue as the province of set-theoretic foundations. We're left with **Proof Checking**, the new goal introduced by univalent foundations. The promise is that ordinary mathematical reasoning will be easily and directly expressed in CIC and the validity of proofs then checked automatically in COQ, and thus that homotopy type theory will provide a framework for reliable **Proof Checking**:

> I now do my mathematics with a proof assistant. I have lots 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. (Voevodsky 2014a, p. 9)

I think we can all agreed that this is a very attractive picture, even if it would only apply to areas of mathematics amenable to this sort of conceptualization.

13.4 Conclusion

The upshot of all this, I submit, is that there wasn't and still isn't any need to replace set theory with a new 'foundation'. There isn't a unified concept of 'foundation'; there are only mathematical jobs reasonably classified as 'foundational'. Since its early days, set theory has performed a number of these important mathematical roles – **Risk Assessment**, **Generous Arena**, **Shared Standard**, **Meta-mathematical Corral** – and it continues to do so. Demands for replacement of set theory by category theory were driven by the doomed hope of founding unlimited categories and the desire for a foundation that would provide **Essential Guidance**. Unfortunately, **Essential Guidance** is in serious tension with **Generous Arena** and **Shared Standard**; long experience suggests that ways of thinking beneficial in one area of mathematics are unlikely to be beneficial in all areas of mathematics. Still, the isolation of **Essential Guidance** as a desideratum, also reasonably regarded as 'foundational', points the way to the methodological project of characterizing what ways of thinking work best where, and why.

More recent calls for a foundational revolution from the perspective of homotopy type theory are of interest, not because univalent foundations would replace set theory in any of its important foundational roles, but because it promises something new: **Proof Checking**. If it can deliver on that promise – even if only for some, not all, areas of mathematics – that would be an important achievement. Time will tell. But the salient moral is that there's no conflict between set theory continuing to do its traditional foundational jobs while these newer theories explore the possibility of doing others.[30]

[30]Many thanks to Colin McLarty, Lawrence Paulson, and an anonymous referee for very helpful explanations, discussions, and comments.

References

Awodey, S. (2014). Structuralism, invariance, and univalence. *Philosophia Mathematica, 22*, 1–11.

Awodey, S. (2016a). *Univalence as a principle of logic*. Unpublished. Available at https://www.andrew.cmu.edu/user/awodey/

Awodey, S. (2016b). *A proposition is the (homotopy) type of its proofs* (Unpublished). Available at https://www.andrew.cmu.edu/user/awodey/

Awodey, S., & Coquand, T. (2013, Summer). Univalent foundations and the large-scale formalization of mathematics. *Princeton Institute for Advanced Study Letter*. https://www.ias.edu/ideas/2013/awodey-coquand-univalent-foundations

Burgess, J. (2015). *Rigor and structure*. Oxford: Oxford University Press.

Ernst, M. (2015). The prospects of unlimited category theory. *Review of Symbolic Logic, 8*, 306–327.

Frege, G. (1879). *Begriffsschrift* (S. Bauer-Mengelberg, Trans., reprinted in von Heijenoort, Ed.). From Frege to Gödel, (Cambridge, MA: Harvard University Press, 1967), pp. 5–82.

Frege, G. (1884) *Foundations of arithmetic* (J. L. Austin, Trans.). (Oxford: Blackwell, 1980

Mac Lane, S. (1971). Categorical algebra and set-theoretic foundations. In D. Scott & T. Jech (Eds.), *Axiomatic set theory, proceedings of the symposium in pure mathematics of the AMS, UCLA 1967* (pp. 231–240). Providence: AMS.

Mac Lane, S. (1986). *Mathematics: Form and function*. New York: Springer.

Maddy, P. (1997). *Naturalism in mathematics*. Oxford: Oxford University Press.

Maddy, P. (2008). How applied mathematics became pure. *Review of Symbolic Logic, 1*, 16–41.

Maddy, P. (2011). *Defending the axioms*. Oxford: Oxford University Press.

Maddy, P. (2017). 'Set-theoretic foundations', to appear in A. Caicedo et al, *Foundations of mathematics* (Contemporary mathematics, Vol. 609). Providence: AMS.

Paulson, L. (2019). *Formalizing mathematics in simple type theory* (this volume).

Univalent Foundations Program, group author (UFP). (2013). *Homotopy type theory: Univalent foundations of mathematics*. Princeton: Institute for Advanced Study.

Voevodsky, V. (2013, May 8). *Slides for a plenary talk to the Association for Symbolic Logic*. Available at https://www.math.ias.edu/vladimir/lectures

Voevodsky, V. (2014a, Summer). The origins and motivations of univalent foundations. *Princeton Institute for Advanced Study Newsletter*, pp. 8–9. https://www.ias.edu/ideas/2014/voevodsky-origins

Voevodsky, V. (2014b, September) '*Foundations of mathematics: Their past, present and future*', the Bernays lectures I-III, ETH Zurich.