Our Knowledge of the Mathematical World
Part I: Epistemological Framework

Peter Koellner

(Harvard University)

Introduction

In this tutorial I want to consider the epistemology of mathematics in light of the independence phenomenon.

The tutorial has two parts:

(1) Epistemological Framework
(2) Mathematical Substance – The Case for New Axioms and the Prospect of Bifurcation

The philosophical interest of mathematics is that from a metaphysical point of view its objects are clear and distinct and from an epistemological view it is the paradigm of *a priori* knowledge; indeed, it appears to be the clearest case of knowledge that we have.

Despite these dual virtues there has always been a tension between these two sides—the metaphysics and the epistemology of mathematics.

For if we take mathematics at face value we seem to be committed to a metaphysics of abstract objects and this seems to face intractable epistemological problems. And if we begin with a tractable epistemology it seems that either we have to give up large tracts of mathematics or succeed in implementing one of the reductionist programs.

This tension increased as mathematics became richer (culminating in set theory and the higher infinite) and the reductionist programs (logicism, formalism, and constructivism) were undermined by the phenomenon of independence.

It would, for example, have been wonderful if logicism had succeeded. For we would have been able to keep everything in mathematics (including the metaphysics of abstract objects) while having a workable epistemology (on which our knowledge of mathematics is a priori and its statements are analytic).

But, alas, it didn't.

To underscore the present state of high tension let us briefly review the independence results:

In the realm of arithmetic the situation is already dramatic:

Theorem (Gödel, 1931)

Assume that PA *is consistent. Then* $\mathrm{PA} \nvdash \mathrm{Con(PA)}$.

The result is much more general and places what appear to be unsurmountable constraints on the *original* reductive programs.

In the realm of analysis and set theory the situation is even more dramatic:

Theorem (Gödel, 1938)

If ZFC *is consistent then* ZFC + CH *is consistent.*

Theorem (Cohen, 1963)

If ZFC *is consistent then* ZFC + ¬CH *is consistent.*

These results are much more general and place what appear to be unsurmountable constraints on the *revised* reductionist programs.

There are two kinds of independence—*vertical* independence and *horizontal* independence.

It is through these results that the one monolithic system once envisaged by logicism was shattered into an array of mathematical systems—the *interpretability order*.

This order is quite complex—it is neither linear nor well-ordered. However, when one restricts to systems that arise naturally in mathematical practice, it turns out that the systems line up along a well-founded stem, which is marked out by *principles of pure strength*, most notably, *large cardinal axioms*.

There are certain phase transitions in this hierarchy which correspond to limitative foundational frameworks. For example, Q corresponds to *strict finitism*, PRA corresponds to *finitism*, and ATR_0 corresponds to *predicativism*.

Two challenges arise:

(1) THE PROBLEM OF EXTENT: How far does this hierarchy extend?
(2) THE PROBLEM OF SELECTION: Is there a correct path through this hierarchy?

Let us return to the tension between metaphysics and epistemology in light of these results.

The more one accepts in terms of extent, the greater the problem of selection.

For example, if one accepts only pure logic then the epistemological problem is relatively easy; if one accepts only strict finitism then the epistemological problem is somewhat harder but still relatively easy; and so on . . . Finally, in the limit, if one accepts *all* of mathematics (including higher set theory) then it would seem one must embrace a very rich metaphysics and thereby face a *very* difficult epistemological problem.

I will argue that for mathematical and philosophical reasons the limitative frameworks are untenable and so we cannot enjoy the way out of this tension that they provide. We are therefore faced with what appears to be an intractable tension between metaphysics and epistemology.

My method of approaching this tension will be to start with default positions—in mathematics, metaphysics, and epistemology—and, working from this starting point, by a process of revision and mutual adjustment, arrive at a full, systematic picture.

The spirit of this approach is nicely captured in the following quote of Russell:

> *Universal skepticism, though logically irrefutable, is practically barren; it can only, therefore, give a certain hesitancy to our beliefs, and cannot be used to substitute other beliefs for them. ("Our Knowledge of the External World", 1914)*

I agree. There is no Archimedean point. We must rebuild our system from within, by a process of mutual revision and adjustment.

I will argue that once we are clear on philosophical matters, we will break free from the grip of pseudo-problems that appear to stand in the way of this rather straightforward approach, and, that in the end, we will see that there is no conflict between the apparently rich metaphysical picture and a tenable epistemology.

Because of time limitations I will focus on the epistemological side of the story but, along the way, I will say something about the metaphysical side.

The Axiomatic Method

Let us then begin, in medias res, by taking the notion of *reason* at face value.

In reasoning about a domain the question "What is the reason for *that*?" leads us back, in stepwise fashion, until ultimately we hit bedrock at what are traditionally called the *first principles* or *axioms*.

Aristotle

The first systematic account of the structure of reason appears in Aristotle's *Posterior Analytics*. There are several points I want to adopt from this work.

(1) There are two kinds of reasons—*justificatory* reasons and *explanatory*. The *Posterior Analytics* is concerned with explanatory reasons, while our concern will ultimately be with justificatory order. But the structure of the two kinds of reasons is similar—except at one crucial point—and so we shall learn something structural from his account.

(2) Aristotle presupposes that for each special science—biology, physics, mathematics, etc.—we have reached the stage where we have a *pre-theoretic* grasp of the fundamental concepts and we have *pre-theoretic* knowledge of the basic propositions.

(3) The aim is to advance (for each special science (like mathematics)) from the state of *pre-theoretic* knowledge to the state of *scientific* knowledge and this involves knowing the *explanatory* reasons for the various propositions. In the end, for each special science, the propositions are re-ordered into their *true explanatory order*.

(4) The explanatory order has an axiomatic structure: At the base lay the *first principles* (which Aristotle calls "unexplained explanations") and from that basis one proceeds, in stepwise fashion, to give explanations of other propositions, by giving a *demonstration* (which, in Aristotle's system, is composed of explanatory syllogisms).

(5) Aristotle does not feel that he has to *explain* either the notion of *explanatory reason* or the first principles (the "unexplained explanations"). For these lie *outside* of the explanatory order and explanations are given *within* the explanatory order.

(6) The roots of the (original) distinction between *a prior* and *a posterior* knowledge are to be found in this account. For Aristotle calls the items lower in the order the *prior* items, and he calls the items higher in the order the *posterior* items. To know a proposition *a prior* (in the original, *from-grounds* sense) is to demonstrate (that is, explanatorily deduce) it from its ultimate explanatory grounds.

(7) Still, the question arises: "How do we *grasp* first principles?" Some have taken Aristotle to invoke *rational intuition*. But his actual account is more subtle—it involves an account (in terms of the explanatory order) of our ascent to sophisticated conceptual understanding.

The last two points—the roots of the original notion of *a priority* and the question of our *grasp* of first principles—are points that we shall now elaborate.

Ockham

The original distinction between the *a priori* and *a posteriori* is usually attributed to Albert of Saxony (for example in the Encyclopedia Britannica). But this is a mistake: It appears earlier Chapter 17 of Ockham's *Summa Logicae*, Part III, Tractate II, which is largely a commentary on the *Posterior Analytics*, where the roots of the distinction lie.

Port-Royal Logic

The distinction entered the early modern world through the Port-Royal Logic, which again was highly inspired by the *Posterior Analytics*.

> *These are subjects whose truth it is capable of finding and understanding, either by proving effects by their causes, which is called an a priori proof, or, on the contrary, by demonstrating causes by their effects, which is called an a posteriori proof. (Chapter 2, Part IV, Port-Royal Logic)*

Here, as in Aristotle, cause is understood in a general way—in terms of the four explanatory causes—and the idea of apriority articulated is the same as that implicit in Aristotle.

This notion of a priority is quite different from the modern notion. To distinguish it from the modern notion we will call it the '*from-grounds* notion of apriority': A statement about a given domain D can be known *a priori in the from-grounds sense* iff it is provable from the ultimate grounds concerning D.

Leibniz

Leibniz also employed the from-grounds notion of apriority. But he held a substantial equivalence thesis:

($\star$) A statement can be known *a priori in the from-grounds sense* if and only if it is it can be known in a way that is *independent of experience*.

It was through the presumption of this thesis that the original sense of the term ‘a prior’ was quietly replaced by the modern sense. (See Adams and Smit.)

Frege

It is interesting to note, however, that although the modern sense prevailed the original sense was employed by Frege.

Frege's articulation of the notion of apriority is quite similar to that of Aristotle, only he defines the notion in terms of the *justificatory order*—it concerns "the ultimate ground upon which rests the justification": If this ultimate ground consists of "general laws, which themselves neither need nor admit proof" then the statement is said to be *a priori*.

He also articulates the notion of *analyticity* in a similar manner: If the ultimate ground of the ideal justification consists of "general logical laws and definitions" then the statement is said to be *analytic*.

Notice that for Frege (a) the notion of analyticity is defined in *epistemic* terms and (b) the analytic statements form a subclass of the a priori statements.

This completes our account of the development the traditional conception of the axiomatic method and the notion of a priority.

Part of what is useful about this discussion for my purposes is that I wish to employ the original, from-grounds sense of apriority. For while I think that mathematics is apriori in the independent-of-experience sense of the term, one of the key conclusions I will draw is that much of mathematics is actually a posteriori in the *from-grounds* sense of the term.

However, even this statement is only a first-approximation. For the exact notions that I will invoke are more subtle than the above notions: They will not be *absolute* notions; rather they will be *relative* to a parameter. And this will open the way to the *entanglement* of the notions of apriority and aposteriority.

Alteration

Russell

Russell began with the traditional conception of the axiomatic method and apriori justification and he continued to try to push it after his discovery of the paradox in May 1901. He originally saw that a type-theoretic approach would provide a "way out" but he continued to pursue type-free accounts (in his unpublished work from 1901 to 1907), largely because he thought that only such an approach could be grounded in axioms that would be immediate and self-evident.

But eventually he retreated and the type-hierarchy emerged as essential and central. At that point he has to modify the traditional conception of the axiomatic method and the nature of axioms.

This modification involved his *regressive method* for finding the axioms, according to which an axiom need not be immediate and self-evident but rather can be justified in terms of *its consequences*. On this view, the axioms of mathematics need not be more evident than their consequences; rather there is an inversion—an axiom A can be justified if it implies a great deal of consequences φ which are in fact *more evident* than the axiom. In short, the case for axioms can be *a posteriori* in the from-grounds sense.

Gödel

Gödel inherited this view from Russell and he developed it further. He articulated it in terms of his between *intrinsic* and *extrinsic* justifications of new axioms.

Concerning *intrinsic* justifications he writes:

> *[the] axioms of set theory [*ZFC*] by no means form a system closed in itself, but, quite on the contrary, the very concept of set on which they are based suggests their extension by new axioms which assert the existence of still further iterations of the operation "set of"* ...

As examples he mentions small large cardinal axioms such as those asserting the existence of inaccessible and Mahlo cardinals.

> ... *[t]hese axioms show clearly, not only that the axiomatic system of set theory as used today is incomplete, but also that it can be supplemented without arbitrariness by new axioms which only* **unfold the content of the concept of set.** *(Gödel 1964, 260–261).*

Concerning *extrinsic* justifications he wrote:

> *[E]ven disregarding the intrinsic necessity of some new axiom, and even in case it has no intrinsic necessity at all, a probable decision about its truth is possible also in another way, namely, inductively by studying its "success"...*
>
> *... There might exist axioms so abundant in their verifiable consequences, shedding so much light upon a whole field, and yielding such powerful methods for solving problems ... that, no matter whether they are intrinsically necessary, they would have to be accepted at least in the same sense as any well-established physical theory. (Gödel 1964)*

We have here the distinction between *a priori* and *a posteriori* justifications in the *from-grounds* sense.

The notion of analyticity also plays a role here. For Gödel thinks that the first principles involved in an a priori justification are "true owing to the concepts involved in them". So his view is quite similar to Frege's.

Analyticity and First Principles

This connection between analyticity and apriority appears in contemporary discussions (though those discussions involve the modern, independent-of-experience notion of apriority).

This brings us back to the second point from Aristotle that we wished to elaborate: The question of how we come to *grasp* the first principles?

The modern-day representatives that I have in mind—Bealer, Boghossian, and Peacocke—provide an account of apriority that is grounded in an account of the constitutive conditions on conceptual grasp.

The basic idea is this: For each concept C, let Γ_C be the collection of propositions that an epistemically well-situated and well-functioning person must accept if that person is to be said to grasp the concept C. The case of interest is when a proposition P contains a concept C such that P is in Γ_C. For in such a case, if an epistemically well-situated and well-functioning person is to *grasp* P then that person must *accept* P.

Since in this case “grasp implies acceptance” it is reasonable to call such a proposition *analytic*.

[Indeed this account is actually very much like the account of that Quine gives (in *Word and Object* and *The Roots of Reference*) of a notion of analyticity that is acceptable by his lights (modulo issues connected with indeterminacy of translation).]

But an important issue arises ...

The main examples that proponents of this account focus on are elementary principles of logic, like the law of identity and $\wedge$-introduction and $\wedge$-elimination. Now, it *is* plausible for such principles that to understand the fundamental concepts involved in them one must accept the principles.

But part of what makes this plausible is that in these cases the principles have no *existential import*. In the case, however, of the principles of number theory and the principles of set theory it is quite *implausible* to maintain that simply by understanding the concepts involved one can come to have justification for believing them, for these principles imply the existence of infinitely many objects (and, in the latter case, very large levels of infinity).

This is something that the modern proponents of this approach pass over in silence. For they give the account for logic (which has no existential import) and quickly say that the same account lifts to number theory and set theory (where the existential import is quite substantial).

Gödel was more sensitive to this point. He distinguished between being *tautological* and being *analytic*, where the former is "content-free" and the latter is compatible with having "real content", as in the case of the first principles of set theory.

Still, he accepted that pure understanding of concepts could lead to justified belief in the principles of set theory (and so belief in the higher infinite); he seems to have simply regarded this as something mysterious.

But it is a bit *too* mysterious. One cannot help being reminded of the ontological argument for the existence of God. As in that argument it appears as if we are getting something from nothing.

The key point is that while it is true that *if* we grasp, say, the concepts of number theory, *then* we are justified in believing in infinitely many objects, the account as stated is *conditional*: We only obtain justified belief on the *condition* that we do indeed have the grasp of the concepts in question, and this in turn, requires that *there be* such concepts (hanging together in the appropriate way).

This rather obvious observation opens the way to a more nuanced account.

The New Conception

To fix ideas let us consider number theory and set theory. We say that a statement is analytic *relative* to such a conception if it is a rational condition on grasping the conception that one accepts the statement. But this requires that there be a conception there to be grasped. And, in cases such as this, where we are dealing with concepts that characterize *fundamental* structures, this amounts, I maintain, to the *coherence* of the conception.

So we arrive at *relative* notions of analyticity and apriority (in the from-grounds sense) according to which a statement is analytic or a priori justified *relative* to the coherence of a conception.

This raises the question of how we can come to be justified in the belief that an underlying conception is coherent. This, I maintain, is something that requires *a posteriori* justification.

The obvious objection to this state of affairs is that according to it a prior justification depends on a posteriori justification and a posteriori justification depends on a prior justification. So we appear to have a circle.

But not all circles are the same—some are *viscous*, some are *virtuous*.

I maintain that we have here a case of a virtuous circle, that is, a case of *interdependence* that is quite familiar from our understanding of reason in other domains.

To see this it is best to compare the case of empirical reason: When we set out to justify an empirical proposition—say, a theoretical proposition in a *view*—reason pushes us backward until ultimately we hit the bedrock of experience. But what propositions an experience justifiably supports is dependent on the view we bring to bear on it. So, in this familiar case, we also have interdependence. But it is virtuous, not viscous.

Our present case is simply an instance of this more general phenomenon.

The above relative notions are defined in terms of the *justificatory order*. This order also requires modification.

For on the traditional conception the first principles are regarded as immediate and *self-evident*. But in our treatment of analyticity we saw that in the case of principles with existential import this is not plausible.

In place of the notion of self-evidence we will invoke the relative notion of *being more intrinsically plausible (or evident) than* and the corresponding notion of *degrees of intrinsic plausibility (or evidentness)*.

One virtue of this change is what while there is disagreement on which statements are "self-evident" there is a great deal of agreement, in many cases, on when one statement is more intrinsically plausible (or evident) than another.

Here are some examples:

(1) Over the base theory EFA, the Finite Ramsey Theorem is equivalent to the totality of superexponentiation.

(2) Over the base theory RCA_0, the Hilbert Basis Theorem is equivalent to the statement that ω^ω is well-ordered.

(3) Over the base theory ACA′, Goodstein's Theorem and the Hydra Theorem are each equivalent to the Σ^0_1-soundness of PA.

(4) Over the base theory RCA_0, Kruskal's Tree Theorem is equivalent to the statement that $\theta\Omega_\omega$ is well-ordered.

(5) Over the base theory ACA', the Exotic Case (of Boolean Relation Theory) is equivalent to the Σ^0_1-soundness of SMAH.

In each case the statement on the right is more intrinsically plausible than the statement on the statement on the left. This is why these are the axioms from which we prove the theorems. The 'reverse' in 'reverse mathematics' references this directionality.

The Search for New Axioms—A Puzzle

In seeking axioms (within a given interpretability degree) the ideal would be a statement in that degree which is more intrinsically evident than any other statement in that degree.

But as one moves to stronger and stronger systems it appears that the most intrinsically evident propositions become *less and less* evident.

So, it would, appear that the search for new axioms is hopeless since as we move to stronger systems our best possible epistemic stance becomes weaker and weaker.

To underscore this problem let us consider several attempts that proponents have thought would (for each interpretability degree) provide statements that are intrinsically evident.

(1) Friedman's Second Program

- Two Primitives: "better than", "much better than".
- Innocuous Axioms: Nothing is better than itself; given two things there is something much better than both.
- Tendentious Axioms: Anything much better than a given thing is also much better than something minimally better than that given thing.

Issue: The system is ultimately borrowing from set-theory.

(2) Neo-Logicism

- As the systems increase in strength they become more and more transparently set-theoretic.

(3) Constructive Hilbert Program

- Rathjen's bound on Martin-Löf Type Theory: Π^1_2-CA.

(4) Reflection Principles

- The bound $\kappa(\omega)$.

It appears, therefore, that the prospect of finding first principles with a high degree of intrinsic plausibility (evidence) is rather dim.

Is there any way to pick up the slack?

Yes. We must resort to *a posteriori* justifications.

It will be useful to distinguish between *first principles* and *axioms*. The first principles are those principles which characterize fundamental structures, like the basis principles of number theory, analysis, and the layers of set theory. Axioms divide into two cases. First, there are those axioms that are minimal points in the justificatory order. But not all axioms need occupy this position. Some axioms admit of a posteriori justification and may actually possess little in the way of intrinsic plausibility.

The goal, in the second part of this tutorial, will be to spell out the nature of such a case for a new axiom.