

Truth in Mathematics: The Question of Pluralism*

Peter Koellner

“... before us lies the boundless ocean of unlimited possibilities.”

—Carnap, *The Logical Syntax of Language*

The discovery of non-Euclidean geometries (in the 19th century) undermined the claim that Euclidean geometry is the one true geometry and instead led to a plurality of geometries no one of which could be said (without qualification) to be “truer” than the others. In a similar spirit many have claimed that the discovery of independence results for arithmetic and set theory (in the 20th century) has undermined the claim that there is one true arithmetic or set theory and that instead we are left with a plurality of systems no one of which can be said to be “truer” than the others. In this paper I will investigate such pluralist conceptions of arithmetic and set theory. I will begin with an examination of what is perhaps the most sophisticated and developed version of the pluralist view to date—namely, that of Carnap in *The Logical Syntax of Language*—and I will argue that this approach is problematic and that the pluralism involved is too radical. In the remainder of the paper I will investigate the question of what it would take to establish a more reasonable pluralism. This will involve mapping out some mathematical scenarios (using recent results proved jointly with Hugh Woodin) in which the pluralist could arguably maintain that pluralism has been secured.

*I would like to thank Bill Demopoulos, Iris Einheuser, Matti Eklund, Michael Friedman, Warren Goldfarb, Daniel Isaacson, Oystein Linnebo, John MacFarlane, Alejandro Pérez Carballo, and Thomas Ricketts for helpful discussion.

Here is the plan of the paper. In §1, I give a brief historical overview of the emergence of pluralism and an accompanying minimalist conception of philosophy, starting with Descartes and ending with Carnap. In §2, I investigate Carnap’s pluralism and his minimalist conception of philosophy, arguing that although Carnap’s pluralism is defensible in some domains (for example, with respect to certain metaphysical and purely notational questions), both his pluralism in physics and his pluralism in mathematics are untenable. In the course of the argument we shall see that mathematics is rather resilient to attempts to prove that it is fleeting. I also argue that Carnap’s minimalist conception of philosophy is too extreme—instead of placing philosophy before science it places philosophy after science. In contrast, I suggest that there is room (and historical precedent) for a more meaningful engagement between philosophy and the exact sciences and, moreover, that it is through such an engagement that we can properly approach the question of pluralism in mathematics. The rest of the paper is devoted to exploring whether through such an engagement we could be led to a defensible pluralism. In §3, I lay the groundwork for this new orientation by drawing a structural parallel between physics and mathematics. Einstein’s work on special relativity is given as an exemplar of the kind of meaningful engagement I have in mind. For it is with Einstein that we come to see, for reasons at the intersection of philosophy and science, that statements once thought to have absolute significance (such as “ A is simultaneous with B ”) are ultimately relativized. Our question then is whether something similar could happen in arithmetic or set theory, that is, whether for reasons at the intersection of philosophy and mathematics—reasons sensitive to actual developments in mathematics—we could come to see that statements once thought to have absolute significance (such as the Continuum Hypothesis (CH)) are ultimately relativized. In §4, I investigate the region where we currently have reason to believe that there is no pluralism concerning mathematical statements (this includes first- and second-order arithmetic). Finally, in §5, I begin by considering some standard arguments to the effect that our current mathematical knowledge already secures pluralism in set theory (for example, with regard to CH). After arguing that this is not the case, I map out a scenario (the best to my knowledge) in which one could arguably maintain that pluralism holds (even at the level of third-order arithmetic). This scenario has the virtue that it is sensitive to future developments in mathematics. In this way, by presenting such scenarios, we will, with time, be able to test just how resilient mathematics really is.

1 The Emergence of Pluralism

To introduce some of our main themes let us begin by reviewing some key developments in the history of the relation between philosophy and the exact sciences, starting with Descartes (in the 17th century) and ending with Carnap (in the first half of the 20th century). We shall see that under the influence of two major scientific revolutions—the Newtonian and the Einsteinian—the pendulum swung from a robust conception of mathematics and maximalist conception of philosophy (one in which philosophy is prior to the exact sciences) to a pluralist conception of mathematics and minimalist conception of philosophy (one that places philosophy after the exact sciences).¹

1.1 Descartes

Descartes distinguished between “natural philosophy”—which concerns the corporal world—and “first philosophy” (or “metaphysics”)—which concerns the incorporeal world. From the 17th to the 19th century the term “natural philosophy” was used for what we now call the exact sciences. For Descartes, first philosophy was prior to natural philosophy in that it was required to lay the rational foundation for natural philosophy.

This view of the privileged status of first philosophy with respect to natural philosophy came to be challenged by developments in physics. To begin with, this approach did not fulfill its promise of a comprehensive and unified natural philosophy. Indeed, in many cases, it led to results that were at variance with experience. For example, it led Descartes (in the *The Principles of Philosophy* (1644)) to formulate laws of impact that were at odds with observation. Huygens then found the correct laws through an ingenious blend of experience and reason (here embodied in the principle of relativity). In this way natural philosophy gained ground as an independent discipline. The true vindication of natural philosophy as an independent discipline came with Newton’s *The Mathematical Principles of Natural Philosophy* (1687). Indeed it is here—in natural philosophy and not in first philosophy—that one finds the first true hope of a comprehensive and adequate account of the corporal world. Moreover, Newton’s developments created additional trouble for first philosophy since there were aspects of Newton’s system—such as action at

¹In my overview of this historical development I am very much indebted to the work of Michael Friedman. See especially his *Dynamics of Reason*.

a distance and absolute space—that were incompatible with the apparent results of first philosophy.

1.2 Kant

Kant rejected the pretensions of first philosophy and instead took Newtonian physics to be an exemplar of theoretical reason. On Kant’s view, the task for philosophy proper is not to arrive at natural philosophy from some higher, privileged vantage point,² but rather to take natural philosophy as given and then articulate the necessary conditions for its possibility.³ The Kantian solution involves the notion of the constitutive a priori: The basic laws of logic, arithmetic, Euclidean geometry, and the basic laws of Newtonian physics are not things we “find in the world”, rather they are things we “bring to the world”—they are the necessary conditions of our experience of the world (in particular, they are necessary for the formulation of the Newtonian law of universal gravitation).⁴

This view of the central and secure status of both Euclidean geometry and the basic laws of Newtonian physics came to be challenged (like its predecessor) by developments in mathematics and physics. First, the discovery of non-Euclidean geometries in the 19th century dethroned Euclidean geometry from its privileged position and led to the important distinction between pure (formal) geometry and applied (physical) geometry. Second, many of the basic Newtonian laws were overturned by Einstein’s special theory of relativity. Finally, Einstein’s general theory of relativity provided reason to believe that the geometry of physical space is actually non-Euclidean.

²In a famous passage in the *Critique of Pure Reason* (1781/1787), Kant wrote: “With respect to the question of unanimity among the adherents of metaphysics in their assertions, it is still so far from this that it is rather a battle ground, which seems to be quite peculiarly destined to exert its forces in mock combats, and in which no combatant has ever yet been able to win even the smallest amount of ground, and to base on his victory an enduring possession. There is therefore no doubt that its procedure has, until now, been a merely random groping, and, what is worst of all, among mere concepts.”

³Thus, in the *Prolegomena to Any Future Metaphysics* (1783) the first and second parts of the transcendental problem are “How is pure mathematics possible?” and “How is the science of nature possible?”.

⁴There is a subdivision within these a priori truths: Logic is analytic, while arithmetic, Euclidean geometry and the basic laws of Newtonian physics are synthetic.

1.3 Reichenbach

Reichenbach accommodated these developments by relativizing the notion of the constitutive a priori.⁵ The key observation concerns the fundamental difference between definitions in pure geometry and definitions in physical geometry. In pure geometry there are two kinds of definition: first, there are the familiar *explicit* definitions; second, there are *implicit* definitions, that is the kind of definition whereby such fundamental terms as ‘point’, ‘line’, and ‘surface’ are to derive their meaning from the fundamental axioms governing them.⁶ But in physical geometry a new kind of definition emerges—that of a *physical* (or *coordinative*) definition:

The physical definition takes the meaning of the concept for granted and coordinates to it a physical thing; it is a *coordinative definition*. Physical definitions, therefore, consist in the coordination of a mathematical definition to a “piece of reality”; one might call them *real definitions*. (Reichenbach (1924), p. 8)

As an example, Reichenbach gives the coordination of “beam of light” with “straight line”.⁷

Now there are two important points about physical definitions. First, some such correlation between a piece of mathematics and “a piece of physical reality” is necessary if one is to articulate the laws of physics (e.g. consider “force-free moving bodies travel in straight lines”). Second, given a piece of pure mathematics there is a great deal of freedom in choosing the coordinative definitions linking it to “a piece of physical reality”.⁸ So we have here a conception of the a priori which (by the first point) is constitutive (of the empirical significance of the laws of physics) and (by the second point) is relative. Moreover, on Reichenbach’s view, in choosing between two empirically equivalent theories that involve different coordinative definitions, there

⁵Reichenbach was one of the five students to attend Einstein’s first course on the general theory of relativity in 1919. His central early works on this subject are Reichenbach (1920) and Reichenbach (1924).

⁶Reichenbach (1924, p. 3) refers the reader to Hilbert’s work on the foundations of geometry for more on the notion of implicit definition.

⁷Reichenbach stresses that such definitions involve an element of idealization and that the physical notions concerned (such as “beam of light”) are theory-laden. A coordinative definition is thus to be distinguished from an operational definition.

⁸“... coordinative definitions are arbitrary, and “truth” and “falsehood” are not applicable to them.” (Reichenbach (1924), p. 9)

is no issue of “truth”—there is only the issue of simplicity.⁹

Now, Reichenbach went beyond this and he held a more radical thesis—in addition to advocating pluralism with respect to physical geometry (something made possible by the free element in coordinative definitions), he advocated pluralism with respect to pure mathematics (such as arithmetic and set theory). According to Reichenbach, this view is made possible by the axiomatic conception of Hilbert, wherein axioms are treated as “implicit definitions” of the fundamental terms:

The problem of the axioms of mathematics was solved by the discovery that they are definitions, that is, arbitrary stipulations which are neither true nor false,¹⁰ and that only the logical properties of a system—its consistency, independence, uniqueness, and completeness—can be subjects of critical investigation. (Reichenbach (1924), p. 3)

On this view there is a plurality of consistent formal systems and the notions of “truth” and “falsehood” do not apply to these systems; the only issue in choosing one system over another is one of convenience for the purpose at hand and this is brought out by investigating their metamathematical properties, something that falls within the provenance of “critical investigation”, where there *is* a question of truth and falsehood.

This radical form of pluralism came to be challenged by Gödel’s discovery of the incompleteness theorems. To begin with, through the arithmetization of syntax, the metamathematical notions that Reichenbach takes to fall within the provenance of “critical investigation” were themselves seen to be a part of arithmetic. Thus, one cannot, on pain of inconsistency, say that there is a question of truth and falsehood with regard to the former but not the latter. More importantly, the incompleteness theorems buttressed the view that truth outstrips consistency. This is most clearly seen using Rosser’s

⁹In his discussion of Einstein’s particular definition of simultaneity, after noting its simplicity, Reichenbach writes: “This simplicity has nothing to do with the truth of the theory. The truth of the axioms decides the empirical truth, and every theory compatible with them which does not add new empirical assumptions is equally true.” (Reichenbach (1924), p. 11)

¹⁰Reichenbach is extending the Hilbertian thesis concerning implicit definitions since although Hilbert held this thesis with regard to formal geometry he did not hold it with regard to arithmetic. (I am indebted to Richard Zach for discussion on this point.) Later I shall argue that this extension is illegitimate.

strengthening of the first incompleteness theorem as follows: Let T be an axiom system of arithmetic that (a) falls within the provenance of “critical investigation” and (b) is sufficiently strong to prove the incompleteness theorem.¹¹ Then, assuming that T is consistent (something which falls within the provenance of “critical investigation”), by Rosser’s strengthening of the first incompleteness theorem, there is a Π_1^0 -sentence φ such that (provably within $T + \text{Con}(T)$) both $T + \varphi$ and $T + \neg\varphi$ are consistent. However, not both systems are equally legitimate. For it is easily seen that if a Π_1^0 -sentence φ is independent from such a theory, then it must be true.¹² So, although $T + \neg\varphi$ is consistent, it proves a false arithmetical statement.¹³

1.4 Carnap

Nevertheless, in full awareness of the incompleteness theorems, in *The Logical Syntax of Language* (1934), Carnap held a view that, on the face of it at least, appears to be quite similar to Reichenbach’s view, both with regard to the thesis concerning the conventional element in physical geometry and the thesis concerning the purely conventional nature of pure mathematics. However, Carnap’s position is much more subtle and sophisticated. Indeed I think that it is the most fully developed account of a pluralist conception of mathematics that we have and, for this reason, I will examine it in detail in the next section.

What I want to point out here is that Carnap also articulated an accompanying minimalist conception of philosophy, one according to which “[p]hilosophy is to be replaced by the logic of science—that is to say, by the logical analysis of the concepts and sentences of the sciences, for *the logic of science is nothing other than the logical syntax of the language of science.*” (Carnap (1934), p. xiii) Thus the pendulum has gone full swing: We started with a robust conception of mathematics and maximalist conception of philosophy (where philosophy comes before the exact sciences) and through the

¹¹A natural choice for such an axiom system is Primitive Recursive Arithmetic (PRA) but much weaker systems suffice, for example, $\text{I}\Delta_0 + \text{exp}$. Either of these systems can be taken as T in the argument that follows.

¹²The point being that T is Σ_1^0 -complete (provably so in T).

¹³For the reader concerned that this argument involves the notion of truth in a problematic way, notice (as we have indicated in the parenthetical remarks) that it can be implemented in $T + \text{Con}(T)$ (which is taken to fall within the provenance of “critical investigation”); that is, $T + \text{Con}(T)$ proves that $T + \varphi$ and $T + \neg\varphi$ are consistent and it *also* proves φ .

influence of two scientific revolutions we were led to a pluralist conception of mathematics and minimalist conception of philosophy (where philosophy comes after the exact sciences).¹⁴

In the next section I will criticize both Carnap's pluralist conception of mathematics and his minimalist conception of philosophy and suggest that it is through a more meaningful engagement between philosophy and mathematics that we can properly address the question of pluralism.

2 Carnap on the Foundations of Logic and Mathematics

In *The Logical Syntax of Language* (1934)¹⁵ Carnap defends the following three distinctive philosophical theses: (1) The thesis that logic and mathematics are analytic and hence without content¹⁶ and purely formal.¹⁷ (2) A radical pluralist conception of pure mathematics (embodied in his Principle of Tolerance) according to which the meaning of the fundamental terms is determined by the postulates governing them and hence any consistent set of postulates is equally legitimate.¹⁸ (3) A minimalist conception of philosophy in which most traditional questions are rejected as pseudo-questions and the task of philosophy is identified with the study of the logical syntax of the

¹⁴The above historical sketch has been necessarily brief and so some qualifications and comments are in order. First, there are other views that reject first philosophy and largely place philosophy after science, for example, various forms of naturalism. Second, the above division—philosophy-before-science versus philosophy-after-science—is not intended as a complete classification of views—there are degrees. (In fact, even in the extreme case where a view attempts to place philosophy entirely after science, if it does so for philosophical reasons, then it would seem to involve first philosophy at the meta-level. To overcome such an impurity, the advocate of such a view might regard the meta-philosophical view not as a philosophical thesis but rather as a proposal. Below we shall see this as an attractive interpretation of Carnap.) Finally, it is worth mentioning that Carnap did not have an uncritical attitude toward science and he did a good deal of work that one might say lies at the intersection of philosophy and science (take, for example, his work on probability and entropy).

¹⁵For definiteness I shall focus almost exclusively on the position held by Carnap in this work. Unless otherwise specified all references in this section are to this work.

¹⁶See pp. xiv, 7, 41.

¹⁷See pp. 1–2, 258.

¹⁸See p. xv.

language of science.¹⁹

I will deal with these three philosophical theses in order in the following three subsections.²⁰

2.1 Logic, Mathematics, and Content

The first philosophical thesis—that logical and mathematical truths are analytic and hence without content—involves Carnap’s distinctive account of the notions of analyticity and content. So it is here that we shall begin. But first we need to make a few terminological remarks since some Carnap’s terminology differs from modern terminology and, where there is overlap, his usage—most notably the term ‘syntax’—is often out of step with modern usage.

2.1.1 Some Key Terminology

A central distinction for Carnap is that between *definite* and *indefinite* notions. A *definite* notion is one that is recursive, such as “is a formula” and “is a proof of φ ”. An *indefinite* notion is one that is non-recursive, such as “is an ω -consequence of PA” and “is true in $V_{\omega+\omega}$ ”. This leads to a distinction between (i) the *method of derivation* (or *d-method*), which investigates the semi-definite (recursively enumerable) metamathematical notions, such as *demonstrable*, *derivable*, *refutable*, *resolvable*, and *irresolvable*, and (ii) the *method of consequence* (or *c-method*), which investigates the (typically) non-recursively enumerable metamathematical notions such as *consequence*, *analytic*, *contradictory*, *determinate*, and *synthetic*.

A *language* for Carnap is what we would today call a formal axiom system.²¹ The rules of the formal system are definite (recursive)²² and Carnap

¹⁹See p. 279.

²⁰This section is a summary of the fuller discussion in Koellner (2009), to which the reader is referred for further details and a discussion of Carnap’s views during later periods. In my thinking about *The Logical Syntax of Language* I have benefited from Friedman (1999c), Gödel (1953/9), and Goldfarb & Ricketts (1992). After writing Koellner (2009), Warren Goldfarb drew my attention to Kleene’s review (Kleene (1939)). I am in complete agreement with what Kleene has to say and there is some overlap between our discussions, though my discussion goes a good deal further.

²¹See p. 167.

²²Strictly speaking Carnap does not prohibit indefinite rules (see p. 172) but in all of the cases he considers (Language I, Language II, etc.) the rules are definite.

is fully aware that a given language cannot include its own c-notions (see Theorem 60c.1).

The *logical syntax* of a language is what we would today call metatheory. It is here that one formalizes the c-notions for the (object) language. From among the various c-notions Carnap singles out one as central, namely, the notion of (direct) consequence; from this c-notion all of the other c-notions can be defined in routine fashion.²³

2.1.2 The Analytic/Synthetic Distinction

We now turn to Carnap's account of his fundamental notions, most notably, the analytic/synthetic distinction and the division of primitive terms into 'logico-mathematical' and 'descriptive'. Carnap actually has two approaches. The first approach occurs in his discussion of specific languages—Languages I and II. Here he starts with a division of primitive terms into 'logico-mathematical' and 'descriptive' and upon this basis defines the c-notions, in particular the notions of being analytic and synthetic. The second approach occurs in the discussion of general syntax. Here Carnap reverses procedure: he starts with a specific c-notion—namely, the notion of direct consequence—and he uses it to define the other c-notions and draw the division of primitive terms into 'logico-mathematical' and 'descriptive'.

A. The First Approach. In the first approach Carnap introduces two languages—Language I and Language II. It is important to note (as we have above) that here by 'language' Carnap means what we would call a 'formal system'.²⁴ The background languages (in the modern sense) of Language I and Language II are quite general—they include expressions that we would call 'descriptive'. Carnap starts with a demarcation of primitive terms into 'logico-mathematical' and 'descriptive'. The expressions he classifies as 'logico-mathematical' are exactly those included in the modern versions of these systems; the remaining expressions are classified as 'descriptive'. Language I is a version of PRA and Language II is a version of finite type theory built over PA. The d-notions for these languages are the standard proof-theoretic ones. So let us concentrate on the c-notions.

²³See p. 168.

²⁴Although Carnap's usage of 'language' is somewhat misleading I will follow him in certain instances—for example, in speaking of 'Language I'—simply because the usage is well-entrenched in his writing and the secondary literature. It will always be clear from context whether I am referring to a system or a language (in the modern sense).

For Language I Carnap starts with a consequence relation based on two rules—(i) the rule that allows one to infer φ if $T \vdash \varphi$ (where T is some fixed Σ_1^0 -complete formal system) and (ii) the ω -rule. It is then easily seen that one has a complete theory for the logico-mathematical fragment, that is, for any logico-mathematical sentence φ , either φ or $\neg\varphi$ is a consequence of the null set. The other c-notions are then defined in the standard fashion. For example, a sentence is *analytic* if it is a consequence of the null set; *contradictory* if its negation is analytic; etc.

For Language II Carnap starts by defining analyticity. His definition is a notational variant of the Tarskian truth definition with *one important difference*—namely, it involves an asymmetric treatment of the logico-mathematical and descriptive expressions. For the logico-mathematical expressions his definition really just is a notational variant of the Tarskian truth definition. But descriptive expressions must pass a more stringent test to count as analytic—they must be such that if one replaces all descriptive expressions in them by variables of the appropriate type, then the resulting logico-mathematical expression is analytic, that is, true.²⁵ In other words, to count as analytic a descriptive expression must be a substitution-instance of a general logico-mathematical truth. With this definition in place the other c-notions are defined in the standard fashion.

The *content* of a sentence is defined to be the set of its non-analytic consequences. It then follows immediately from the definitions that logico-mathematical sentences (of both Language I and Language II) are analytic or contradictory and (assuming consistency) that analytic sentences are without content.

B. The Second Approach. In the second approach, for a given language, Carnap starts with an *arbitrary* notion of direct consequence²⁶ and from this notion he defines the other c-notions in the standard fashion. More importantly, in addition to defining the other c-notion, Carnap also uses the primitive notion of direct consequence (along with the derived c-notions) to effect the classification of terms into ‘logico-mathematical’ and ‘descriptive’. The guiding idea is that “the formally expressible distinguishing peculiarity of logical symbols and expressions [consists] in the fact that each sentence constructed solely from them is determinate” (177).²⁷ He then gives a formal

²⁵The relevant clause is DA I.C.b. on p. 111.

²⁶There are no conditions placed on this notion—for example, it could be a definite notion (such as “provable in T ”).

²⁷A sentence is *determinate* if either it or its negation is valid, that is, a consequence of

definition that aims to capture this idea. His actual definition is problematic for various technical reasons²⁸ and so we shall leave it aside. What is important for our purposes (as shall become apparent in §2.1.3) is the fact that (however the guiding idea is implemented) the actual division between ‘logico-mathematical’ and ‘descriptive’ expressions that one obtains as output is sensitive to the scope of the direct consequence relation with which one starts.

With this basic division in place, Carnap can now draw various derivative divisions, most notably, the division between analytic and synthetic statements: Suppose φ is a consequence of Γ . Then φ is said to be an *L-consequence* of Γ if either (i) φ and the sentences in Γ are logico-mathematical, or (ii) letting φ' and Γ' be the result of unpacking all descriptive symbols, then for every result φ'' and Γ'' of replacing every (primitive) descriptive symbol by an expression of the same genus (a notion that is defined on p. 170), maintaining equal expressions for equal symbols, we have that φ'' is a consequence of Γ'' . Otherwise φ is a *P-consequence* of Γ . This division of the notion of consequence into *L-consequence* and *P-consequence* induces a division of the notion of demonstrable into *L-demonstrable* and *P-demonstrable* and the notion of valid into *L-valid* and *P-valid* and likewise for all of the other d-notions and c-notions. The terms ‘*analytic*’, ‘*contradictory*’, and ‘*synthetic*’ are used for ‘L-valid’, ‘L-contravalid’, and ‘L-indeterminate’.

Again it follows immediately from the definitions that logico-mathematical sentences are analytic or contradictory and that analytic sentences are without content. This is what Carnap says in defense of the first of his three basic theses.

2.1.3 Criticism #1: The Argument from Free Parameters

The trouble with the first approach is that the definitions of analyticity that Carnap gives for Languages I and II are highly sensitive to the original classification of terms into ‘logico-mathematical’ and ‘descriptive’. And the trouble with the second approach is that the division between ‘logico-mathematical’ and ‘descriptive’ expressions (and hence division between ‘analytic’ and ‘synthetic’ truths) is sensitive to the scope of the direct consequence relation with

the null set.

²⁸For some of these see Quine (1963), though note that Quine’s discussion appears at points to mistakenly assume that the notion of direct consequence that Carnap uses is a d-notion.

which one starts. This threatens to undermine Carnap’s thesis that logico-mathematical truths are analytic and hence without content. Let us discuss this in more detail.

In the first approach, the original division of terms into ‘logico-mathematical’ and ‘descriptive’ is made by stipulation and if one alters this division one thereby alters the derivative division between analytic and synthetic sentences. For example, consider the case of Language II. If one calls only the primitive terms of first-order logic ‘logico-mathematical’ and then extends the language by adding the machinery of arithmetic and set theory, then, upon running the definition of ‘analytic’, one will have the result that true statements of first-order logic are without content while (the distinctive) statements of arithmetic and set theory have content.²⁹ For another example, if one takes the language of arithmetic, calls the primitive terms ‘logico-mathematical’ and then extends the language by adding the machinery of finite type theory, calling the basic terms ‘descriptive’, then, upon running the definition of ‘analytic’, the result will be that statements of first-order arithmetic are analytic or contradictory while (the distinctive) statements of second- and higher-order arithmetic are synthetic and hence have content. In general, by altering the input, one alters the output, and Carnap adjusts the input to achieve his desired output.³⁰

In the second approach, there are no constraints on the scope of the direct consequence relation with which one starts and if one alters it one thereby alters the derivative division between ‘logico-mathematical’ and ‘descriptive’ expressions. Recall that the guiding idea is that logical symbols and expressions have the feature that sentences composed solely of them are determinate. The trouble is that (however one implements this idea) the resulting division of terms into ‘logico-mathematical’ and ‘descriptive’ will be highly sensitive to the scope of the direct consequence relation with which one starts.³¹ For example, let S be first-order PA and for the direct consequence relation take “provable in PA”. Under this assignment Fermat’s Last Theorem will be deemed descriptive, synthetic, and to have non-trivial

²⁹This is more in keeping with the standard use of the term ‘content’. For, in a straightforward sense, the truths of first-order logic do not pertain to a special subject matter (they are perfectly general) while those of arithmetic and set theory do.

³⁰For a fuller discussion—one that involves a discussion of two additional parameters—see Koellner (2009).

³¹Carnap was fully aware of this sensitivity. See, for instance, the example involving $g_{\mu\nu}$ that Carnap gives (on p. 178) right after he draws the division.

content.³² For an example at the other extreme, let S be an extension of PA that contains a physical theory and let the notion of direct consequence be given by a Tarskian truth definition for the language. Since in the metalanguage one can prove that every sentence is true or false, every sentence will be either analytic (and so have null content) or contradictory (and so have total content). To overcome such counter-examples and get the classification that Carnap desires one must ensure that the consequence relation is (i) complete for the sublanguage consisting of expressions that one wants to come out as ‘logico-mathematical’ and (ii) not complete for the sublanguage consisting of expressions that one wants to come out as ‘descriptive’. Once again, by altering the input, one alters the output, and Carnap adjusts the input to achieve his desired output.

To summarize: What we have (in either approach) is not a principled distinction. Instead, Carnap has merely provided us with a flexible piece of technical machinery involving free parameters that can be adjusted to yield a variety of outcomes concerning the classifications of analytic/synthetic, contentful/non-contentful, and logico-mathematical/descriptive. In his own case, he has adjusted the parameters in such a way that the output is a formal articulation of his logicist view of mathematics that the truths of mathematics are analytic and without content. And one can adjust them differently to articulate a number of other views, for example, the view that the truths of first-order logic are without content while the truths of arithmetic and set theory have content. The possibilities are endless. The point, however, is that we have been given no reason for fixing the parameters one way rather than another. The distinctions are thus not principled distinctions. It is trivial to prove that mathematics is trivial if one trivializes the claim.

2.1.4 Criticism #2: The Argument from Assessment Sensitivity

Carnap is perfectly aware that to define c-notions like analyticity one must ascend to a stronger metalanguage. However, there is a distinction that he appears to overlook,³³ namely, the distinction between (i) having a stronger

³²Carnap is fully aware of such counter-examples. See p. 231 of §62 where he notes that his definitions have the consequence that the universal numerical quantifier in Whitehead and Russell’s *Principia Mathematica* is really a *descriptive* symbol, the reason being that the system involves only d-rules and hence (by Gödel’s incompleteness theorem) it will leave some Π_1^0 -sentences undecided.

³³See, for example, pp. 1–2.

system S' that can *define* ‘analytic in S ’ and (ii) having a stronger system S' that can, in addition, *evaluate* a given statement of the form ‘ φ is analytic in S ’. It is an elementary fact that two systems S_1 and S_2 can employ the same definition (from an intensional point of view) of ‘analytic in S ’ (using either the definition given for Language I or Language II) but *differ* on their evaluation of ‘ φ is analytic in S ’ (that is, differ on the extension of “analytic in S ”). Thus, to determine whether ‘ φ is analytic in S ’ holds one needs to access much more than the “syntactic design” of φ —in addition to ascending to an essentially richer metalanguage one must move to a sufficiently strong system to evaluate ‘ φ is analytic in S ’. The first step need not be a big one.³⁴ But for certain φ the second step must be huge.³⁵

In fact, it is easy to see that to answer “Is φ analytic in Language I?” is just to answer φ and, in the more general setting, to answer all questions of the form “Is φ analytic in S ?” (for various mathematical φ and S), where here “analytic” is defined as Carnap defines it for Language II, *just is to answer all questions of mathematics*.³⁶ The same, of course, applies to the c-notion of consequence. So, when in first stating the Principle of Tolerance, Carnap tells us that we can choose our system S arbitrarily and that “no question of justification arises at all, but *only* the question of the syntactical consequences to which one or other of the choices leads” (p. xv, my emphasis)—where here, as elsewhere, he means the c-notion of consequence—*he has given us no assurance, no reduction at all*.

2.2 Radical Pluralism

This brings us to the second philosophical thesis—the thesis of pluralism in mathematics. Let us first note that Carnap’s pluralism is quite radical. We are told that “any postulates and any rules of inference [may] be chosen arbi-

³⁴For example, taking S to be PA one can simply extend the language by adding a truth predicate and extend the axioms by adding the Tarskian truth axioms and allow the truth predicate to figure in the induction scheme. The resulting system S' is only minimally stronger than S . It proves $\text{Con}(\text{PA})$ but not much more.

³⁵To continue the example in the previous footnote, suppose one wishes to show that $\text{Con}(\text{ZF} + \text{AD})$ is analytic in S (which, as I shall argue below, it is). To do this one must move to a system that has consistency strength beyond that of “ZFC + there are ω -many Woodin cardinals”.

³⁶For the first note that $T \vdash_{\omega} \varphi$ if and only if φ (where T is the fixed Σ_1^0 -complete theory) and for the second note that the Tarskian truth definition has the feature that $T(\ulcorner \varphi \urcorner)$ if and only if φ (where T is the truth predicate).

trarily” (xv); for example, the question of whether the Principle of Selection (that is, the Axiom of Choice (AC)) should be admitted is “purely one of expedience” (142); more generally,

The [logico-mathematical sentences] are, from the point of view of material interpretation, expedients for the purpose of operating with the [descriptive sentences]. Thus, in laying down [a logico-mathematical sentence] as a primitive sentence, only usefulness for this purpose is to be taken into consideration. (142)

So the pluralism is quite broad—it extends to AC and even to Π_1^0 -sentences.³⁷

Now, as I argued in §1.3 on Reichenbach, there are problems in maintaining Π_1^0 -pluralism. One cannot, on pain of inconsistency, think that statements about consistency are not “mere matters of expedience” without thinking that Π_1^0 -statements generally are not mere “matters of expedience”. But I want to go further than make such a negative claim. I want to uphold the default view that the question of whether a given Π_1^0 -sentence holds is *not* a mere matter of expedience; rather, such questions fall within the provenance of theoretical reason.³⁸ In addition to being the default view, there are solid reasons behind it. One reason is that in adopting a Π_1^0 -sentence one could

³⁷For further evidence that Carnap’s pluralism is this radical, see xv, 124 and Carnap (1939), p. 27.

³⁸Some readers might be tempted to interpret me as saying that there is a “fact of the matter” concerning Π_1^0 -sentences. I want to resist such a formulation since I am not sure that I understand the phrase “fact of the matter” as it is often employed. I certainly understand this phrase when it is used merely as a point of contrast with “matter of mere expedience”. On this reading, it means no more than that the issue is one of theoretical reason, one concerning something more than mere utility, one having something to do with the truth of one theory over another (not in some robust metaphysical sense of the ‘truth’ but in the ordinary sense). This distinction is not as sharp as one would like but one can point to clear cases (as we have seen above and as we shall see below) and the distinction strikes me as significant. In contrast, the phrase “fact of the matter” is often used in a way that strives for something more—“thick truth”, “Truth with a capital ‘T’”, the idea of “carving reality at the joints”, etc. I cannot think of examples where I could go along with such talk with any confidence. Moreover, it seems to me that such talk buys into the myth that there is some Archimedean vantage point from which we can survey the array of theories and compare them with “reality as it is in and of itself”—in short, a “sideways-on view” (in McDowell’s apt phrase). This is something that I think Kant and Carnap were right to reject and, once we follow them in doing so, we are left with the former, thinner distinction—the one I employ in the text. (See Tait (1986) for a critical discussion of the “myth of the model in the sky”. The reader might press me with the concern that the thinner distinction that I invoke (following Carnap) rests on a similar

always be struck by a counter-example.³⁹ Other reasons have to do with the clarity of our conception of the natural numbers and with our experience to date with that structure. On this basis, I would go further and maintain that for *no* sentence of first-order arithmetic is the question of whether it holds a mere matter of experience. Certainly this is the default view from which one must be moved.

What does Carnap have to say that will sway us from the default view, and lead us to embrace his radical form of pluralism? In approaching this question it is important to bear in mind that there are two general interpretations of Carnap. According to the first interpretation—the *substantive*—Carnap is really trying to *argue* for the pluralist conception. According to the second interpretation—the *non-substantive*—he is merely trying to *persuade* us of it, that is, to show that of all the options it is most “expedient”.⁴⁰

The most obvious approach to securing pluralism is to appeal to the work on analyticity and content. For if mathematical truths are without content and, moreover, this claim can be maintained with respect to an arbitrary mathematical system, then one could argue that even apparently incompatible systems have null content and hence are really compatible (since there is no contentual-conflict).

Now, in order for this to secure radical pluralism, Carnap would have to first secure his claim that mathematical truths are without content. But, as we have argued above, he has not done so. Instead, he has merely provided us with a piece of technical machinery that can be used to articulate any one of a number of views concerning mathematical content and he has adjusted the parameters so as to articulate his particular view. So he has not *secured* the thesis of radical pluralism.⁴¹ Thus, on the substantive interpretation, Carnap has failed to achieve his end.

This leaves us with the non-substantive interpretation. There are a number of problems that arise for this version of Carnap. To begin with, Carnap’s technical machinery is not even suitable for *articulating* his thesis of radical

myth. I do not think that it does. But it would take us too far afield to explore the issue here.)

³⁹This is something that Carnap recognizes.

⁴⁰See Friedman (1999c) for the substantive interpretation and see Goldfarb & Ricketts (1992) for the non-substantive interpretation. In what follows I do not take a stance on the interpretative issue. Instead I criticize both versions of Carnap.

⁴¹There are two other approaches that one might consider. These approaches fail as well. See Koellner (2009).

pluralism since (using either the definition of analyticity for Language I or Language II) there is no metalanguage in which one can say that two apparently incompatible systems S_1 and S_2 both have null content and hence are really contentually compatible. To fix ideas, consider a paradigm case of an apparent conflict that we should like to dissolve by saying that there is no contentual-conflict: Let $S_1 = \text{PA} + \varphi$ and $S_2 = \text{PA} + \neg\varphi$, where φ is any arithmetical sentence, and let the metatheory be $\text{MA} = \text{ZFC}$. The trouble is that on the approach to Language I, although in MT we can prove that each system is ω -complete (which is a start since we wish to say that each system has null content), we can also prove that one has null content while the other has *total* content (that is, in ω -logic, *every* sentence of arithmetic is a consequence). So, we cannot, within MT articulate the idea that there is no contentual-conflict.⁴² The approach to Language II involves a complementary problem. To see this note that while a strong logic like ω -logic is something that one can apply to a *formal system*, a truth definition is something that applies to a *language* (in our modern sense). Thus, on this approach, in MT the definition of analyticity given for S_1 and S_2 is the same (since the two systems are couched in the same language). So, although in MT we can say that S_1 and S_2 do not have a contentual-conflict this is only because we have given a deviant definition of analyticity, one that is blind to the fact that in a very straightforward sense φ is analytic in S_1 while $\neg\varphi$ is analytic in S_2 .

Now, although Carnap's machinery is not adequate to articulate the thesis of radical pluralism (for a given collection of systems) in a given metatheory, under certain circumstances he can attempt to articulate the thesis by changing the metatheory. For example, let $S_1 = \text{PA} + \text{Con}(\text{ZF} + \text{AD})$ and $S_2 = \text{PA} + \neg\text{Con}(\text{ZF} + \text{AD})$ and suppose we wish to articulate both the idea that the two systems have null content and the idea that $\text{Con}(\text{ZF} + \text{AD})$ is analytic in S_1 while $\neg\text{Con}(\text{ZF} + \text{AD})$ is analytic in S_2 . As we have seen no single metatheory (on either of Carnap's approaches) can do this. But it turns out that because of the kind of assessment sensitivity that we discussed in §2.1.4, there are two metatheories MT_1 and MT_2 such that in MT_1 we can say both that S_1 has null content and that $\text{Con}(\text{ZF} + \text{AD})$ is analytic in S_1 , while in MT_2 we can say both that S_2 has null content and that $\neg\text{Con}(\text{ZF} + \text{AD})$ is analytic in S_2 . But, of course, this is simply because (any such metatheory) MT_1 proves $\text{Con}(\text{ZF} + \text{AD})$ and (any such metathe-

⁴²This is related to a point made by Michael Friedman. See p. 226 of Friedman (1999c).

ory) MT_2 proves $\neg\text{Con}(\text{ZF} + \text{AD})$. So we have done no more than (as we must) reflect the difference between the systems in the metatheories. Thus, although Carnap does not have a way of articulating his radical pluralism (in a given metalanguage), he certainly has a way of *manifesting* it (by making corresponding changes in his metatheories).

As a final retreat Carnap might say that he is not trying to persuade us of a *thesis* that (concerning a collection of systems) can be articulated in a given framework but rather is trying to persuade us to adopt a thorough radical pluralism as a “*way of life*”. He has certainly shown us how we can make the requisite adjustments in our metatheory so as to consistently manifest radical pluralism. But does this amount to more than an algorithm for begging the question?⁴³ Has Carnap shown us that there is no question to beg? I do not think that he has said anything persuasive in favour of embracing a thorough radical pluralism as the “most expedient” of the options. The trouble with Carnap’s entire approach (as I see it) is that the question of pluralism has been detached from actual developments in mathematics. To be swayed from the default position something of greater substance is required.

2.3 Philosophy as Logical Syntax

This brings us finally to Carnap’s third philosophical thesis—the thesis that philosophy is the logical syntax of the language of science. This thesis embodies a pluralism more wide-ranging than mathematical pluralism. For just as foundational disputes in mathematics are to be dissolved by replacing the theoretical question of the justification of a system with the practical question of the expedience of adopting the system, so too many philosophical and physical disputes are to be dissolved in a similar fashion:

It is especially to be noted that the statement of a philosophical thesis sometimes . . . represents not an *assertion* but a *suggestion*. Any dispute about the truth or falsehood of such a thesis is quite mistaken, a mere battle of words; we can at most discuss the utility of the proposal, or investigate its consequences. (299–300)

Once this shift is made one sees that

⁴³There are many places where Carnap quite obviously begs the question in the metatheory. See, for example, §§43 and 44 where Carnap discusses intuitionism and predicativism; to people like Brouwer and Poincaré these sections would be maddening.

the question of truth or falsehood cannot be discussed, but only the question whether this or that form of language is the more appropriate for certain purposes. (300)

Philosophy has a role to play in this. For, on Carnap's conception, "as soon as claims to scientific qualifications are made" (280), philosophy just *is* the study of the syntactical consequences of various scientific systems.

I want to focus on this distinction between "the question of truth or falsehood" and "the question of whether this or that form of language is the more appropriate for certain purposes" as Carnap employs it in his discussion of metaphysics, physics, and mathematics.

Let us start with metaphysics. The first example that Carnap gives (in a long list of examples) concerns the apparent conflict between the statements "*Numbers* are classes of classes of things" and "*Numbers* belong to a special primitive kind of objects". On Carnap's view, these material formulations are really disguised versions of the proper formal or syntactic formulations "Numerical expressions are class-expressions of the second-level" and "Numerical expressions are expressions of the zero-level" (300). And when one makes this shift (from the "material mode" to the "formal mode") one sees that "the question of truth or falsehood cannot be discussed, but only the question whether this or that form of language is the more appropriate for certain purposes" (300). There is a sense in which this is hard to disagree with. Take, for example, the systems PA and ZFC – Infinity. These systems are mutually interpretable (in the logician's sense). It seems that there is no theoretical question of choosing one over the other—as though one but not the other were "getting things right", "carving mathematical reality at the joints"—but only a practical question of expedience relative to a given task. Likewise with apparent conflicts between systems that construct lines from points versus points from lines or use sets versus well-founded trees, etc. So, I am inclined to agree with Carnap on such *metaphysical* disputes. I would also agree concerning theories that are mere *notational* variants of one another (such as, for example, the conflict between a system in a typed language and the mutually interpretable system obtained by "flattening" (or "amalgamating domains")). But I think that Carnap goes too far in his discussion of physics and mathematics. I have already discussed the case of mathematics (and I will have much more to say about it below). Let us turn to the case of physics.

Carnap's conventionalism extends quite far into physics. Concerning physical hypotheses he writes:

The construction of the physical system is not affected in accordance with fixed rules, but by means of conventions. These conventions, namely, the rules of formation, the L-rules and the P-rules (hypothesis), are, however, not arbitrary. The choice of them is influenced, in the first place, by certain practical methodological considerations (for instance, whether they make for simplicity, expedience, and fruitfulness in certain tasks). . . . But in addition the hypotheses can and must be tested by experience, that is to say, by the protocol-sentences—both those that are already stated and the new ones that are constantly being added. Every hypothesis must be compatible with the total system of hypotheses to which the already recognized protocol-sentences also belong. That hypotheses, in spite of their subordination to empirical control by means of the protocol-sentences, nevertheless contain a conventional element is due to the fact that the system of hypotheses is never univocally determined by empirical material, however rich it may be. (320)

Thus, while in pure mathematics it is convention and question of expedience all the way (modulo consistency), in physics it is convention and question of expedience modulo empirical data.

I think that this conception of theoretical reason in physics is too narrow. Consider the following two examples: first, the historical situation of the conflict between the Ptolemaic and the Copernican accounts of the motion of the planets and, second, the conflict between the Lorentz's mature theory of 1905 (with Newtonian spacetime) and Einstein's special theory of relativity (with Minkowski spacetime). In the first case, the theories are empirically equivalent to within 3' of arc and this discrepancy was beyond the powers of observation at the time. In the second case, there is outright empirical equivalence.⁴⁴ Yet I think that the reasons given at the time in favour of the Copernican theory were not of the practical variety; they were not considerations of expedience, they were solid theoretical reasons pertaining to truth.⁴⁵

⁴⁴See Janssen (2002) for a discussion of the empirical equivalence of Lorentz's mature theory of 1905 and special relativity.

⁴⁵See Evans (1998), p. 412 and Wilson (1970), p. 109 for six solid reasons.

Likewise with the reasons for Einstein's theory over Lorentz's (empirically equivalent) mature theory of 1905.⁴⁶

2.3.1 Conclusion

Let us summarize the above conclusions. Carnap has not given a substantive account of analyticity and so he has not given a *defense* of his first thesis—the thesis that mathematical truths are analytic and hence formal and without content. Instead he has presented some technical machinery that can be used to formally articulate his view. In the case of his second thesis—that radical pluralism holds in mathematics—the situation is even worse. Not only has he not provided a defense of this thesis, he has not even provided technical machinery that is suitable to *articulate* the thesis (for a given collection of systems) in a given metatheory. All he can do is *manifest* his radical pluralism by mirroring the differences among the candidate systems as differences among their metatheories.

These limitations lead one to think that perhaps Carnap's first two theses were not intended as assertions—theses to be argued for—but rather as suggestions—theses to be adopted for practical reasons as the most expedient of the available options. The trouble is that he has not provided a persuasive case that these theses are indeed the most expedient. In any case, one cannot refute a proposal. One can only explain one's reasons for not following it—for thinking that it is not most expedient.

There is something right in Carnap's motivation. His motivation comes largely from his rejection of the myth of the model in the sky. One problem with this myth is that it involves an alienation of truth (to borrow an apt phrase of Tait).⁴⁷ However, the non-pluralist can agree in rejecting such a myth. Once we reject the myth and the pretensions of first philosophy, we are left with the distinction between substantive theoretical questions and matters of mere expedience. I agree with Carnap in thinking that the choice between certain metaphysical frameworks (e.g. whether we construct

⁴⁶See DiSalle (2006), Friedman (1999a) and Janssen (2002). To be sure, the full vindication of the Copernican theory over the Ptolemaic theory (and Tychonic theory) came with Galileo (and Newton) and the full vindication of special relativity over Lorentz's mature theory came with general relativity. For my purposes it is sufficient that the reasons given before the discovery of the telescope (and Newtonian gravitation theory), in the first case, and the discovery of general relativity, in the second case, have *some* force.

⁴⁷See §72.

lines from points or points from lines) and between certain notational variants (e.g. whether we use a typed language or amalgamate types) are not substantive theoretical choices but rather matters of mere expedience. But I think that he goes too far in saying that the choice between empirically equivalent theories in physics and the choice between arbitrary consistent mathematical systems in mathematics are likewise matters of mere expedience. In many such cases one can provide convincing theoretical reasons for the adoption of one system over another. To deny the significance of such reasons appears to me to reveal a remnant of the kind of first philosophy that Carnap rightfully rejected.

Just as Kant was right to take Newtonian physics as exemplary of theoretical reason and retreat from a maximalist conception of philosophy—one that placed philosophy before science—I think that we should take Einsteinian physics as exemplary of theoretical reason and retreat from a minimalist conception of philosophy—one that places philosophy after science. Similarly, in the case of mathematics, I think we are right to take the developments in the search for new axioms seriously. But we must ensure that the pendulum does not return to its starting point. The proper balance, I think, lies in between, with a more meaningful engagement between philosophy and science. In the remainder of this paper I shall outline such an approach. I believe that what I shall have to say is Carnapian in spirit, if not in letter.

3 A New Orientation

I have embraced Carnap’s pluralism with respect to certain purely metaphysical and notational questions but rejected it with regard to certain statements in physics and mathematics. For example, I agree with Carnap that there is no substantive issue involved in the choice between ZFC and a variant of this theory that uses the ontology of well-founded trees instead of sets. But I disagree in thinking the same with regard to the choice between $ZFC + \text{Con}(ZFC)$ and $ZFC + \neg\text{Con}(ZFC)$. Here there is a substantive difference. In slogan form my view might be put like this: Existence in mathematics is (relatively) cheap; truth is not.⁴⁸

⁴⁸The subject requires further discussion than I can give it here. For example, *how* cheap is existence in mathematics? Does consistency suffice? Consider $ZFC + \text{CH}$ versus $ZFC + \neg\text{CH}$. There is reason to believe that both are consistent. The trouble is that CH and $\neg\text{CH}$ are existential claims and, on a straightforward reading, the objects that they

In the physical case I gave two examples of where I think we must part ways with Carnap. First, the historical situation of the choice between the Ptolemaic and Copernican account of the planets; second, the choice between Lorentz’s mature theory (with Newtonian spacetime) and Einstein’s theory of special relativity (with Minkowski spacetime). In opposition to Carnap, I hold that in each case the choice is not one of “mere expedience”, but rather falls within the provenance of theoretical reason—indeed I take these to be some of the highest achievements of theoretical reason.

In addition to breaking with Carnap on the above (and other) particular claims, I think we should break with his minimalist conception of philosophy—a conception that places philosophy *after* science—and—without reverting to a conception that places it *before*—embrace a conception involving a more meaningful engagement between philosophy and science. Moreover, in doing so, I think that we can properly address the question of pluralism and thereby gain some insight into what it would take to secure a more reasonable pluralism.

In this connection, the theory of special relativity actually serves as a guide in *two* respects. First, as already mentioned, like the Copernican case, it illustrates the point that there are cases where, despite being in a situation where one has empirical equivalence, one can give theoretical reasons for one theory over another. Second, it does this in such a way that once one makes the step to special relativity one sees a kind of pluralism, namely, with respect to the class of all permissible foliations of Minkowski spacetime—the point being that it is in principle impossible to use the laws of physics to single out (in a principled way) any one of these foliations over another⁴⁹—and so we change perspective, see each as standing on a par and bundle them all up into one spacetime: Minkowski spacetime.⁵⁰ The source of this form of pluralism is different than Carnap’s—it is driven by developments at the intersection

assert to exist cannot coexist. I am inclined to think that existence in mathematics is “Consistency + X ” but I do not know how to solve for X .

⁴⁹In the sense that the laws of physics are Lorentz invariant.

⁵⁰Some metaphysicians accept these physical limitations but are not moved by them. For example, in his entry on presentism in the *The Oxford Handbook of Metaphysics* (2003), Thomas Crisp says that we can just stipulate a preferred foliation. Well, we can do *that* but what reason do we have to think that our stipulation captures any structure (either physical or metaphysical) of our universe? We can also stipulate a preferred center or preferred direction. The possibilities are endless. But most would maintain that such stipulations are idle and do not reflect the structure of spacetime. How is the situation with simultaneity any different?

of philosophy and physics and is sensitive to developments that fall squarely within physics.

I want now to examine analogues of these two features in the mathematical case. In Section 4, I will argue that there are cases in mathematics where theoretical reason presents us with a convincing case for one theory over another. Indeed I will argue that theoretical reason goes quite a long way. But there are limitations to our current understanding of the universe of sets and there is a possibility that we are close to an impasse. In Section 5, I will investigate the possibility of being in a position where one has reason to believe that we are faced with a plurality of alternatives that cannot in principle be adjudicated on the basis of theoretical reason and, moreover, where this leads us to reconceive the nature of some fundamental notions in mathematics.

On the way toward this it will be helpful to introduce some machinery and use it to present a very rough analogy between the structure of physical theory and the structure of mathematical theory.

To a first approximation, let us take a physical theory to be a formal system with a set of coordinative definitions. We are interested in the relation of empirical equivalence between such theories. Let us first distinguish two levels of data: The *primary* data are the observational sentences (such as “At time t , wandering star w has longitude and latitude (φ, ϑ) ”) that have actually been verified; the *secondary* data are the observational generalizations (such as “For each time t , wandering star w has longitude and latitude $(\varphi(t), \vartheta(t))$ ”). The primary data (through accumulation) can provide us with inductive evidence of the secondary data. We shall take our notion of empirical equivalence to be based on the secondary data. Thus, two physical theories are *empirically equivalent* if and only if they agree on the secondary data.⁵¹ The *problem of selection* in physics is to select from the equivalence classes of empirically equivalent theories. Some choices—for example, certain purely metaphysical and notational choices—are merely matters of expedience; others—for example, that between Lorentz’s mature theory and special relativity—are substantive and driven by theoretical reason.

In the mathematical case we can be more exact. A theory is just a (recursively enumerable) formal system and as our notion of equivalence we

⁵¹This is not intended as a precise, formal definition—for example, I have said nothing about what it takes to count as data. So the definition is quite flexible. Nevertheless, it is sharp enough to serve for our purposes.

shall take the notion of mutual interpretability in the logician’s sense. The details of this definition would be too distracting to give here;⁵² suffice it to say that the technical definition aims to capture the notion of interpretation that is ubiquitous in mathematics, the one according to which Poincaré provided an interpretation of two-dimensional hyperbolic geometry in the Euclidean geometry of the unit circle, Dedekind provided an interpretation of analysis in set theory, and Gödel provided an interpretation of the theory of formal syntax in arithmetic. Every theory of pure mathematics is mutually interpretable with a theory in the language of set theory. So, for ease of exposition, we will concentrate on theories in the language of set theory.⁵³ We shall assume that all theories contain ZFC – Infinity. As in the case of physics we shall distinguish two levels of data: The *primary* data are the Δ_1^0 -sentences that have actually been verified (such as, to choose a metamathematical example, “ p is not a proof of $\neg\text{Con}(\text{ZFC})$ ”) and the *secondary* data are the corresponding generalizations, which are Π_1^0 -sentences (such as, “For all proofs p , p is not a proof of $\neg\text{Con}(\text{ZFC})$ ”). As in the case of physics, in mathematics the secondary data can be definitely refuted but never definitely verified. Nevertheless, again as in the case of physics, the primary data can provide us with evidence for the secondary data. Finally, given our assumption that all theories contain ZFC – Infinity, we have the following fundamental result: Two theories are mutually interpretable if and only if they prove the same Π_1^0 -sentences, that is, if and only if they agree on the secondary data. Thus, we have a nice parallel with the physical case, with Π_1^0 -sentences being the analogues of observational generalizations. The *problem of selection* in mathematics is to select from the equivalence classes of mutual interpretability (which are called the interpretability *degrees*). Some choices—for example, certain purely metaphysical and notational choices—are mere matters of expedience. But in restricting our attention to a fixed language we have set most of these aside. Other choices are substantive and fall within the provenance of theoretical reason. For example, let T be ZFC – Infinity. One can construct a Δ_2^0 -sentence such that T , $T + \varphi$ and $T + \neg\varphi$ are mutually interpretable.⁵⁴ The choice between these two theories

⁵²See chapter 6 of Lindström (2003).

⁵³For our present purposes there is little (if any) loss of generality in this restriction since our concern now is with theoretical reason and so we do not wish to be distracted by choices between, say, ZFC – Infinity and PA.

⁵⁴Such a sentence is called an *Orey sentence* for T . See Lindström (2003) for the construction of such a sentence.

is *not* a matter of mere expedience.

The question of pluralism has two aspects which can now be formulated as follows: First, setting aside purely metaphysical and notational choices, can the problem of selection be solved in a convincing way by theoretical reason? Second, how far does this proceed; does it go all the way or does one eventually reach a “bifurcation point”; and, if so, could one be in a position to recognize it? Our approach is to address these questions in a way that is sensitive to the actual results of mathematics, where here by ‘results’ we mean results that everyone can agree on, that is, the theorems, the primary data. But while the *data* will lie in mathematics, the *case* will go beyond mathematics⁵⁵ and this is what we mean when we say that the case will lie at the intersection of mathematics and philosophy.

4 The Initial Stretch: First- and Second-Order Arithmetic

There is currently no convincing case for pluralism with regard to first-order arithmetic and most would agree that given the clarity of our conception of the structure of the natural numbers and given our experience to date with that structure such a pluralism is simply untenable. So, most would agree that not just for any Π_1^0 -sentence, but for *any* arithmetical sentence φ , the choice between $PA + \varphi$ and $PA + \neg\varphi$ is not one of mere expedience. I have discussed this above and will not further defend the claim here. Instead I will take it for granted in what follows.

The real concerns arise when one turns to *second-order* arithmetic, *third-order* arithmetic, and more generally, the transfinite layers of the set-theoretic hierarchy.⁵⁶

4.1 Independence in Set Theory

One source of the concern is the proliferation of independence results in set theory and the nature of the forms of independence that arise. Consider

⁵⁵And necessarily so since we are not stating a theorem.

⁵⁶Recall that in our setting the various systems of arithmetic (e.g. PA and PA₂ (the second-order axioms of Peano Arithmetic), etc.) are cast in the language of set theory and that from the point of view of independence there is no loss of generality in this assumption.

the statements PU (Projective Uniformization)⁵⁷ of (schematic) second-order arithmetic and CH (Cantor’s continuum hypothesis), a statement of third-order arithmetic. Combined results of Gödel and Cohen show that if ZFC is consistent then these statements are independent of ZFC: Gödel invented (in 1938) the method of *inner models*. He defined a canonical and minimal inner model L of the universe of sets, V , and he showed that CH holds in L . This had the consequence that ZFC could not refute CH. Cohen invented in (in 1963) the method of *forcing* (or *outer models*). Given a complete Boolean algebra he defined a model $V^{\mathbb{B}}$ and showed that $\neg\text{CH}$ holds in $V^{\mathbb{B}}$. This had the consequence that ZFC could not prove CH. Thus, these results together showed that CH is independent of ZFC. Similar results hold for PU and a host of other questions in set theory. In contrast to the incompleteness theorems, where knowledge of the independence of the sentences produced— Π_1^0 -sentences of first-order arithmetic—actually settles the statement, in the case of PU and CH, knowledge of independence provides *no clue whatsoever* as to how to settle the statement.

4.2 The Problem of Selection for Second-Order Arithmetic

How then are we to solve the problem of selection with respect to second-order arithmetic?

There are two steps in solving the problem of selection for a given degree of interpretability. The first step is to *secure the secondary data* by showing that the Π_1^0 -consequences of the theories in the degree are true. One way to do this is to show $(\text{ZFC} - \text{Infinity}) + \bigcup_{n < \omega} \text{Con}(T \upharpoonright n)$, where T is some theory in the degree and $T \upharpoonright n$ is the first n sentences of T . Having thus secured the secondary data, the second step is to *select from the degree*. For example, the degree of ZFC contains $\text{ZFC} + \neg\text{Con}(\text{ZFC})$, $\text{ZFC} + \text{PU}$, $\text{ZFC} + \neg\text{PU}$, $\text{ZFC} + \text{CH}$ and $\text{ZFC} + \neg\text{CH}$, and many other theories. The full problem of selection for this degree would involve settling all such questions, which is a massive task. For the moment we shall concentrate on deciding between $\text{ZFC} + \text{PU}$ and $\text{ZFC} + \neg\text{PU}$. In what follows I shall presuppose ZFC. I will argue that the choice between $\text{ZFC} + \text{PU}$ and $\text{ZFC} + \neg\text{PU}$ is not one of mere expedience; in fact, one can give strong theoretical reasons for $\text{ZFC} + \text{PU}$.

⁵⁷This is the statement that every projective subset of the plane admits a projective choice function.

The case will involve the problem of selection for a much higher degree. But first a word on the structure of the hierarchy of interpretability and large cardinal axioms.

The structure of the hierarchy of interpretability is more disorderly than one might expect—it forms a distributive lattice that is neither linearly ordered nor well-founded. This is shown via the construction of non-standard theories via coding techniques. Remarkably, however, when one restricts to the natural theories that occur in mathematical practice, one finds that the theories are well-behaved—they are well-ordered under interpretability.⁵⁸

Extensions of ZFC – Infinity via *large cardinal axioms* provide us with a canonical class of representatives within this well-ordered hierarchy in that given a theory T in the hierarchy one can generally find—via the dual techniques of inner model theory and forcing—a theory of the form (ZFC – Infinity) + LCA, where LCA is a large cardinal axiom, such that T and (ZFC – Infinity) + LCA are mutually interpretable.⁵⁹ In this way large cardinal axioms (which are (for the most part) naturally well-ordered) provide a gauge of the strength of the theories in the hierarchy of interpretability. The simplest large cardinal axioms are *reflection principles*.⁶⁰ Some notable stepping in the large cardinal hierarchy are strongly inaccessible cardinals, Mahlo cardinals, measurable cardinals, Woodin cardinals, and supercompact cardinals.

The higher degree that I shall be working with is that of the theory

$$T_1 = \text{ZFC} + \text{there are } \omega\text{-many Woodin cardinals.}$$

There are many theories of interest in this degree. For example, the theories

$$T_2 = \text{ZFC} + \text{there is an } \omega_1\text{-dense ideal on } \omega_1,$$

$$T_3 = \text{ZFC} + \text{AD}^{L(\mathbb{R})}, \text{ and}$$

$$T_4 = \text{ZFC} + \text{AD}^{L(\mathbb{R})} + \text{MA} + \neg\text{CH.}$$

are all in the same degree as T_1 , that is, T_1 , T_2 , T_3 , and T_4 all yield the same secondary data.⁶¹

⁵⁸This is a mystery that calls for clarification.

⁵⁹There is an element of imprecision in this claim due to the lack of precision involved in both the notion of a natural theory and the notion of a large cardinal axiom.

⁶⁰Indeed ZFC is the result of supplementing our base theory ZFC – Infinity with a scheme of first-order reflection principles.

⁶¹The main results here are due to Woodin. See Koellner (2006) for further discussion and references.

The first step is to secure the secondary data. As noted above it suffices to establish $(\text{ZFC} - \text{Infinity}) + \bigcup_{n < \omega} \text{Con}(T \upharpoonright n)$, where T is any one of the above theories. There is a strong case for this but it would take us too far afield to present it here. Let me just say that case rests on the intimate connection between determinacy and inner models of large cardinal axioms and that the case is so strong that set theorists who have investigated the network of theorems in this area (the primary data) are quite confident that T_1 (and hence the other theories in its degree) is consistent.⁶²

The second step is to provide theoretical reasons for some of the theories in the degree. One cannot accept all of them since they are not mutually consistent (for example, T_2 and T_4 contradict one another). Nevertheless, I think that a very strong case can be made for T_3 . Moreover, T_3 implies PU and hence resolves the problem of selection for $\text{ZFC} + \text{PU}$ versus $\text{ZFC} + \neg\text{PU}$. Again the case for this is based upon (but goes beyond) the primary data, namely, a large network of mathematical theorems. The case is quite involved and so we shall only give an overview. For details see Koellner (2006) and the references therein.

(1) $\text{AD}^{L(\mathbb{R})}$ is an axiom that has some degree of intrinsic plausibility.⁶³ But, of course, this is just a starting point. One must look to its consequences and connections with other statements to determine, for example, whether its consequences are intrinsically plausible and whether it is implied by other intrinsically plausible axioms.

(2) $\text{AD}^{L(\mathbb{R})}$ has a number of intrinsically plausible consequences—for example, that there is no paradoxical decomposition of the unit sphere using pieces that are definable (in the precise sense of being in $L(\mathbb{R})$). In fact, in addition to implying that all subsets of reals in $L(\mathbb{R})$ are Lebesgue measurable, $\text{AD}^{L(\mathbb{R})}$ implies the other regularity properties for such sets of reals, such as that they have the property of Baire and the perfect set property.

⁶²For a large piece of the primary data see the articles on determinacy and inner model theory in the forthcoming *Handbook of Set Theory*, in particular, Koellner & Woodin (2009b). It is noteworthy that the reasons are not merely inductively based on the fact that a contradiction has not yet been found. There is similar inductive support for the consistency of Quine's system NF but few are confident that it is consistent. The reasons for consistency are more involved and make for a very strong case. To underscore the strength of the case, at the Gödel centenary in Vienna in 2006, Woodin announced that should anyone prove one of these theories inconsistent he would resign his post and demand that his position be given to the person who established the inconsistency. (This is not an advisable strategy for securing tenure.)

⁶³For more on this notion, see Chapter 9 of Parsons (2008).

In addition, $\text{AD}^{L(\mathbb{R})}$ implies that Σ_1^2 -uniformization holds in $L(\mathbb{R})$. These consequences are all intrinsically plausible and these results generalize the features of Borel sets that can be established in ZFC to the level of $L(\mathbb{R})$. In short, $\text{AD}^{L(\mathbb{R})}$ has what appear to be the correct consequences for the structure theory of the sets of reals in $L(\mathbb{R})$ and this is evidence for $\text{AD}^{L(\mathbb{R})}$.

(3) In the above, there is an obvious analogy with the hypothetico-deductive method in physics, the analogue of a physical theory being $\text{AD}^{L(\mathbb{R})}$ and the analogue of observational data being the intrinsically plausible statements. One difference of course is that the notion of an intrinsically plausible statement is not very sharp; another is that judgments of intrinsic plausibility are less secure than observational statements. Remarkably, as if to compensate for this shortcoming, the mathematical case provides us with something more. To bring this out consider the concern—which arises also in the physical case—that there might be other theories with the same intrinsically plausible (cf. observational) consequences. In the physical case one can never allay this concern. Remarkably, in the mathematical case one can: $\text{AD}^{L(\mathbb{R})}$ is the *only* theory that has the above intrinsically plausible consequences, that is, the intrinsically plausible consequences themselves *imply* $\text{AD}^{L(\mathbb{R})}$ (a result of Woodin).

(4) The pattern in (2) and (3)—where one draws intrinsically plausible consequences from $\text{AD}^{L(\mathbb{R})}$ and then *recovers* $\text{AD}^{L(\mathbb{R})}$ —repeats itself with respect to *other* classes of intrinsically plausible consequences. See Koellner (2006) for some examples.

(5) Let us now turn from consequences to other connections. To begin with there are other intrinsically plausible axioms—most notably large cardinal axioms (by groundbreaking work of Martin, Steel, and Woodin)—that imply $\text{AD}^{L(\mathbb{R})}$.

(6) Large cardinal axioms and axioms of definable determinacy (such as $\text{AD}^{L(\mathbb{R})}$) spring from entirely different sources. Yet there is an intimate connection between them. It turns out that $\text{AD}^{L(\mathbb{R})}$ is *equivalent* to a statement asserting the existence of inner models of certain large cardinals.⁶⁴ We have here a case where intrinsically plausible principles from completely different domains reinforce one another.

(7) Not only do large cardinal axioms imply $\text{AD}^{L(\mathbb{R})}$, many other theories imply $\text{AD}^{L(\mathbb{R})}$. For example, *both* T_2 and T_4 imply $\text{AD}^{L(\mathbb{R})}$ *despite the fact that* T_2 *and* T_4 *are incompatible*.

⁶⁴See §8.1 of Koellner & Woodin (2009b).

(8) The phenomenon in (7) is quite general. Time after time it is shown that a strong theory, which on the face of it has nothing to do with $\text{AD}^{L(\mathbb{R})}$, actually implies (through a deep result) $\text{AD}^{L(\mathbb{R})}$. The technique for establishing this—the core model induction—provides evidence for the claim that *all* sufficiently strong natural theories imply $\text{AD}^{L(\mathbb{R})}$. In this sense $\text{AD}^{L(\mathbb{R})}$ lies in the “overlapping consensus” of all sufficiently strong natural theories. As one climbs the hierarchy of interpretability along any natural path—however, remote the apparent subject matter is from determinacy—it appears that one cannot avoid laying down something that outright implies $\text{AD}^{L(\mathbb{R})}$.

This is only a sample of the network of theorems (primary data) upon which the case for $\text{AD}^{L(\mathbb{R})}$ is based. To be sure, the case itself goes beyond the data, as must *any* case for a new axiom, in light of the incompleteness theorems. But it is remarkable that such a strong case can exist. These arguments do not merely provide practical reasons for adopting $\text{AD}^{L(\mathbb{R})}$ as a matter of expedience; they provide theoretical reasons for accepting $\text{AD}^{L(\mathbb{R})}$.⁶⁵

Thus, we have solved the problem of selection with respect to $\text{AD}^{L(\mathbb{R})}$, an axiom that concerns the structure $L(\mathbb{R})$. Since $\text{AD}^{L(\mathbb{R})}$ implies PU we have also solved the problem of selection for PU. It turns out that in solving the problem of selection for $\text{AD}^{L(\mathbb{R})}$, we have solved the problem of selection for a host of other statements concerning $L(\mathbb{R})$. In fact, there is a sense in which $\text{AD}^{L(\mathbb{R})}$ is so central that in solving the problem of selection for it we have given an “effectively complete” solution of the problem of selection for the entire theory of $L(\mathbb{R})$. We shall spell this out in some detail since once we are armed with such a “complete solution” at the level of $L(\mathbb{R})$ we will next ask how far such a solution extends. This will form the basis of our search for a more reasonable form of pluralism.

The axiom $\text{AD}^{L(\mathbb{R})}$ appears to be “effectively complete” for the theory of $L(\mathbb{R})$. Let us first elaborate, then quantify this. A comparison with PA is useful here. Of course, neither PA nor $\text{ZFC} + \text{AD}^{L(\mathbb{R})}$ is complete because of the incompleteness theorems. However, there are very few statements of prior mathematical interest that are known to be independent of PA. The classic example of such a statement is the Paris-Harrington. Still, there are very few such statements and for this reason people are inclined to regard PA

⁶⁵A staunch skeptic could refrain from going beyond the primary data, committing to a statement only when it has been secured in the form of a theorem (“statement φ is provable in system S ”). Likewise, in the physical case, a staunch skeptic could refrain from going beyond the primary data, committing to a statement only when it has been verified observationally. Each is consistent; neither is reasonable.

as “effectively complete”. The case with axioms of determinacy is even more dramatic. Let PD be the restriction of the axiom $\text{AD}^{L(\mathbb{R})}$ to the domain of second-order number theory. In contrast to PA, there are many statements of prior mathematical interest that are independent of PA_2 , for example, PU. But when one adds PD to PA_2 this ceases to be the case. In fact, $\text{PA}_2 + \text{PD}$ appears to be more complete than PA in that, for example, there is no analogue of the Paris-Harrington theorem. Similar considerations apply to $\text{AD}^{L(\mathbb{R})}$.

Let us try to quantify this and erase the scare quotes around “effectively complete”. To do this we shall henceforth assume large cardinal axioms.⁶⁶ Our goal is to introduce a strong logic that sharpens the notion of “effective completeness” not directly of PD and $\text{AD}^{L(\mathbb{R})}$ but of the large cardinal axioms that imply these axioms. We shall use the hypothesis that there is a proper class of Woodin cardinals, which we shall abbreviate ‘PCWC’.

The motivating result on “effective completeness” is the following:

Theorem 4.1 (Woodin). *Assume ZFC + PCWC. Suppose φ is a sentence and that \mathbb{B} is a complete Boolean algebra. Then*

$$L(\mathbb{R}) \models \varphi \text{ iff } L(\mathbb{R})^{V^{\mathbb{B}}} \models \varphi.$$

Our main (and very powerful) technique for establishing independence in set theory is set forcing—the construction of such models $V^{\mathbb{B}}$. The above theorem shows that in the presence of a proper class of Woodin cardinals (which implies $\text{AD}^{L(\mathbb{R})}$) this technique *cannot* be used to establish independence with respect to statements about $L(\mathbb{R})$.⁶⁷

The aim of the strong logic is to capture this “freezing” or “sealing” by “factoring out” the effects of forcing.

Definition 4.2 (Woodin). Suppose that T is a countable theory in the language of set theory and φ is a sentence. Then

$$T \models_{\Omega} \varphi$$

if for all complete Boolean algebras \mathbb{B} and for all ordinals α ,

$$\text{if } V_{\alpha}^{\mathbb{B}} \models T \text{ then } V_{\alpha}^{\mathbb{B}} \models \varphi.$$

⁶⁶As noted above, many of these axioms are intrinsically plausible and some (though not all) of the above considerations to $\text{AD}^{L(\mathbb{R})}$ apply to them.

⁶⁷And this is not because the large cardinals are somehow throwing a wrench into the machinery of forcing. In fact, they fuel that machinery by generating more forcing extensions.

A theory T is Ω -satisfiable if there exists an ordinal α and a complete Boolean algebra \mathbb{B} such that $V_\alpha^{\mathbb{B}} \models T$.

This notion of semantic implication is robust in that large cardinal axioms imply that the question of what implies what cannot be altered by forcing:

Theorem 4.3 (Woodin). *Assume ZFC+PCWC. Suppose that T is a countable theory in the language of set theory and φ is a sentence. Then for all complete Boolean algebras \mathbb{B} ,*

$$T \models_{\Omega} \varphi \text{ iff } V^{\mathbb{B}} \models \text{“}T \models_{\Omega} \varphi\text{.”}$$

We are now in a position to reformulate the theorem on “freezing the theory of $L(\mathbb{R})$ ” in terms of Ω -logic (Theorem 4.1).

Definition 4.4. A theory T is Ω -complete for a collection of sentences Γ if for each $\varphi \in \Gamma$, $T \models_{\Omega} \varphi$ or $T \models_{\Omega} \neg\varphi$.

Theorem 4.5. *Assume ZFC + PCWC. Then ZFC is Ω -complete for the collection of sentences of the form “ $L(\mathbb{R}) \models \varphi$ ”.*⁶⁸

In this sense, large cardinal axioms give an Ω -complete picture of the theory of $L(\mathbb{R})$. Furthermore, $\text{AD}^{L(\mathbb{R})}$ lies at the heart of this picture. This is how we shall quantify our success in solving the problem of selection with respect to statements of $L(\mathbb{R})$: Assuming that there is a proper class of Woodin cardinals, we have a theory that settles (in Ω -logic) *every* statement about $L(\mathbb{R})$. Our goal now is to see how far such Ω -complete pictures extend and whether we could eventually reach a “bifurcation point”.

5 Bifurcation Scenarios

We now turn to “bifurcation scenarios”, that is, scenarios where not only have our (current) theoretical reasons have failed to settle a given question, say CH, but where we have reason to believe that no further theoretical reasons

⁶⁸Although we have stated the Ω -completeness with respect to ZFC, the large cardinals are really doing the work. For this reason it is perhaps more transparent to formulate the result by saying that “ZFC + there is a proper class of Woodin cardinals” is Ω -complete for the collection of sentences of the form “ $L(\mathbb{R}) \models \varphi$ ”, noting that under this formulation the stated Ω -completeness is trivial unless our background assumptions guarantee that “ZFC + there is a proper class of Woodin cardinals” is Ω -satisfiable.

can settle the question.⁶⁹ One difficulty in articulating such scenarios is that the space of theoretical reasons one might give is not something that one can survey in advance.⁷⁰ We have to do our best and work with what we have.

5.1 An Initial Pass

Some have claimed that the early independence results in set theory already suffice to secure such a position. For example, it is claimed that the independence of CH with respect to ZFC shows that the choice between ZFC + CH and ZFC + \neg CH is one of mere expedience. It is maintained that although there may be practical reasons in favour of adopting one axiom over the other (say for a given purpose at hand) there are no theoretical reasons that one can give for one over the other.

Now, in response, one might argue that independence from ZFC alone cannot suffice since (under reasonable assumptions) $\text{Con}(\text{ZFC})$ is independent of ZFC and yet the choice between ZFC + $\text{Con}(\text{ZFC})$ and ZFC + $\neg\text{Con}(\text{ZFC})$ is hardly one of mere expedience. But to this the critic can respond the background assumption that ZFC is consistent settles the question under consideration. The critic can point out that in contrast to $\text{Con}(\text{ZFC})$ (and any Π_1^0 -sentence for that matter), knowledge of the independence of CH does not settle CH.

However, there are Π_2^0 Orey sentences⁷¹ in the language of arithmetic that also have this feature and yet (as we have argued above) the position that questions concerning them are questions of mere expedience is simply untenable. The critic must therefore cite some distinctive feature of the nature of the independence of CH with respect to ZFC, one that differentiates it from the case of such Π_2^0 -sentences.

Perhaps the key difference is that CH is a statement of prior mathematical interest shown to be independent via set forcing. To this there are two responses. First, why should this matter? We are interested in truth not

⁶⁹Compare the difference between (a) knowing that our current understanding of the physical world does not enable us to detect the luminiferous ether and (b) having reason to believe that no physical understanding will enable us to detect a luminiferous ether.

⁷⁰However, the situation in physics is similar—for example, one cannot rule out definitively the possibility that we might one day find a foliation that has fundamental physical significance and hence that the ultimate laws of physics are not Lorentz invariant.

⁷¹Recall that φ is an *Orey sentence* for T if T , $T + \varphi$ and $T + \neg\varphi$ are mutually interpretable.

human interest. Second, PU is also such a statement and yet we have argued that the choice between $ZFC + PU$ and $ZFC + \neg PU$ is not one of mere expedience.⁷²

So I do not think that the critic has a case here. Nevertheless, I want to see how one might respond on the critic's behalf. The goal will be to gain insight into what it would take to have theoretical reasons (driven by the primary data of mathematics) to believe that a given question of pure mathematics, say CH, is one of mere expedience.

5.2 A More Promising Approach

Let us begin by considering how far the case for $AD^{L(\mathbb{R})}$ and large cardinal axioms extends. Gödel had high expectations for large cardinal axioms. Indeed he went so far as to entertain a generalized completeness theorem for them:

It is not impossible that for such a concept of demonstrability [namely, provability from true large cardinal axioms] some completeness theorem would hold which would say that every proposition expressible in set theory is decidable from the present axioms plus some true assertion about the largeness of the universe of all sets. (Gödel (1946), p. 151)

As a test case he chose CH, a statement of third-order arithmetic.

As we have seen, there has been a partial realization of this program in that large cardinal axioms provide an Ω -complete picture of second-order arithmetic and, in fact, all of $L(\mathbb{R})$. How far does this proceed? In a sense that can be made precise it holds “below CH”.⁷³ Unfortunately, it fails at the “level of CH”, namely, Σ_1^2 , as follows from a series of results originating with Levy and Solovay:

Theorem 5.1. *Assume L is a standard large cardinal axiom. Then $ZFC + L$ is not Ω -complete for Σ_1^2 .*⁷⁴

⁷²It would be of interest to further investigate the analogies and disanalogies between independence in arithmetic and set theory and the bearing of such results on philosophical positions in each domain.

⁷³See §3.3 of Koellner (2006).

⁷⁴This theorem is stated informally since the notion of a “standard large cardinal axiom” is not precise. However, one can cite examples from across the large cardinal hierarchy.

Although large cardinal axioms do not provide an Ω -complete picture of Σ_1^2 , it turns out that one can obtain such a picture provided one supplements large cardinal axioms. Remarkably, CH itself is such a statement.

Theorem 5.2 (Woodin, 1985). *Assume ZFC and that there is a proper class of measurable Woodin cardinals. Then ZFC + CH is Ω -complete for Σ_1^2 .*

Moreover, up to Ω -equivalence, CH is the unique Σ_1^2 -statement that is Ω -complete for Σ_1^2 . Thus, up to Ω -equivalence, there is a *unique* Σ_1^2 -sentence which (along with large cardinal axioms) provides an Ω -complete picture of Σ_1^2 , namely, CH.

If one shifts perspective from Σ_1^2 to $H(\omega_2)$, there is a companion result for \neg CH, *assuming the Strong Ω Conjecture*.⁷⁵

Theorem 5.3 (Woodin). *Assume ZFC + PCWC. Assume the Strong Ω -Conjecture.*

- (1) *There is an axiom A such that*
 - (i) *ZFC + A is Ω -satisfiable and*
 - (ii) *ZFC + A is Ω -complete for the structure $H(\omega_2)$.*
- (2) *Any such axiom A has the feature that*

$$\text{ZFC} + A \models_{\Omega} "H(\omega_2) \models \neg\text{CH}."$$

Thus, assuming that there is a proper class of Woodin cardinals and that the Strong Ω Conjecture holds, there is an Ω -complete picture of $H(\omega_2)$ and any such picture involves a *failure* of CH.

For example, for L one can take “there is a measurable cardinal”, “there is a proper class of Woodin cardinals”, or “there is a non-trivial embedding $j : L(V_{\lambda+1}) \rightarrow L(V_{\lambda+1})$ with critical point below λ ”.

⁷⁵Here $H(\omega_2)$ is the set of all sets that have hereditary cardinality less than ω_2 . The Strong Ω Conjecture is an outstanding conjecture in set theory. It is the conjunction of the Ω Conjecture and the statement “the AD^+ Conjecture is Ω -valid”, where the Ω Conjecture is a conjectured completeness theorem for Ω -logic and the AD^+ Conjecture is the following conjecture: Suppose that A and B are sets of reals such that $L(A, \mathbb{R}) \models \text{AD}^+$, $L(B, \mathbb{R}) \models \text{AD}^+$, and the sets in $\mathcal{P}(\mathbb{R}) \cap (L(A, \mathbb{R}) \cup L(B, \mathbb{R}))$ are ω_1 -universally Baire. Then either $(\Delta_1^2)^{L(A, \mathbb{R})} \subseteq (\Delta_1^2)^{L(B, \mathbb{R})}$ or $(\Delta_1^2)^{L(B, \mathbb{R})} \subseteq (\Delta_1^2)^{L(A, \mathbb{R})}$. See Woodin (1999) for definitions of the remaining terms. We caution the reader that in the existing literature one finds Theorems 5.3 and 5.4 stated with the Ω Conjecture in place of the Strong Ω Conjecture. However, Woodin recently discovered that the proofs require the latter.

These two results raise the spectre of bifurcation at the level of CH. There are two key questions. First, are there recursive theories with higher degrees of Ω -completeness? Second, is there a unique such theory (with respect to a given level of complexity)? The answers to these questions turn on the Strong Ω Conjecture. For ease of exposition, for the remainder of this paper we shall assume that there is a proper class of Woodin cardinals.

If the Strong Ω Conjecture holds, then one cannot have an Ω -complete picture of third-order arithmetic.

Theorem 5.4 (Woodin). *Assume ZFC+PCWC. Assume the Strong Ω Conjecture. Then there is no recursively enumerable theory A such that $\text{ZFC} + A$ is Ω -complete for Σ_3^2 .*

However, if the Strong Ω Conjecture fails, then such higher levels of Ω -completeness may be possible. In fact, there may be a (recursively enumerable) sequence of axioms \vec{A} such that for some large cardinal axiom L the theory $\text{ZFC} + L + \vec{A}$ is Ω -complete for *all* of third-order arithmetic. Going further it could be the case that for each specifiable fragment V_λ of the universe of sets there is a large cardinal axiom L and a (recursively enumerable) sequence of axioms \vec{A} such that $\text{ZFC} + L + \vec{A}$ is Ω -complete for the theory of V_λ . Moreover, it could be the case that any other theory with this feature, say $\text{ZFC} + L + \vec{B}$, agrees with $\text{ZFC} + L + \vec{A}$ on the computation of the theory of V_λ in Ω -logic. This would mean that there is a *unique* Ω -complete picture of the universe of sets up to V_λ . Furthermore, it could be the case that all of these Ω -complete pictures cohere. This would give us a unique Ω -complete picture of (the successive layers of) the entire universe of sets.

One could argue that such an Ω -complete picture of the entire universe of sets is the most that one could hope for. Should uniqueness hold that would be the end of the story from the perspective of Ω -logic for there would be nothing about the (specifiable fragments of) the universe of sets that could not be settled on the basis of Ω -logic in this unique Ω -complete picture.

Unfortunately, uniqueness *must* fail.

Theorem 5.5 (K. and Woodin). *Assume ZFC + PCWC. Suppose L is a large cardinal axiom and \vec{A} is a (recursively enumerable) sequence of axioms such that*

$$\text{ZFC} + L + \vec{A} \text{ is } \Omega\text{-complete for the theory of } V_\lambda,$$

where V_λ is some specifiable fragment of the universe at least as large as $V_{\omega+2}$. Then there exists a (recursively enumerable) sequence of axioms \vec{B} such that

$$\text{ZFC} + L + \vec{B} \text{ is } \Omega\text{-complete for the theory of } V_\lambda$$

but which differs from $\text{ZFC} + L + \vec{A}$ on CH.⁷⁶

Thus, if there is one Ω -complete picture of such a level of the universe (and hence of arbitrarily large such levels), then there is necessarily an *incompatible* Ω -complete picture.⁷⁷

Suppose then that there is one such theory (or a sequence of such theories for higher and higher levels, all of which extend one another). The advocate of pluralism might argue as follows: Such an Ω -complete picture is the most that theoretical reason could hope to achieve. Given that from one such picture (of the form $\text{ZFC} + L + \vec{A}$) we can generate others (say $\text{ZFC} + L + \vec{B}$) that are incompatible, and given that there is great flexibility (in the choice of \vec{B}) and that we can pass from one to another (by altering \vec{B}), this shows that the choice is merely one of expedience. The choice of \vec{B} is analogous to the choice of a timelike vector in Minkowski spacetime. The choice of such a vector in Minkowski spacetime induces a foliation of the space relative to which one can ask whether “ A is simultaneous with B ” but there is no absolute significance to such questions independent of the choice of a timelike vector. Likewise, the choice of the sequence \vec{B} (in conjunction with $\text{ZFC} + L$) provides an Ω -complete picture relative to which one can ask whether “CH holds” and many other such questions but there is no absolute significance to such questions independent of the choice of \vec{B} . As in the case of special relativity we need to change perspective. There is no sense in searching for “the correct” Ω -complete picture, just as there is no sense in searching for “the correct” foliation. Instead of the naive picture of the universe of sets with which we started we are ultimately driven to a new picture, one that deems questions we originally thought to be absolute to be ultimately relativized.

I do not want to endorse this position. I am merely presenting it on behalf of the advocate of pluralism as the best mathematically driven scenario

⁷⁶For a more precise statement (one that spells out the notion of “specifiable fragment” (there called “robustly specifiable fragment”)) see Koellner & Woodin (2009a).

⁷⁷It should be stressed that the choice of CH is just for illustration. Given one such theory one has a great deal of control in generating others.

that I can think of where one could arguably maintain that we had been driven by the primary data of mathematics to shift perspective and regard certain questions of mathematics (such as CH) as based on choices of mere expedience.

A key virtue of this scenario is that it is sensitive to future developments in mathematics—to rule it out it suffices to prove the Strong Ω Conjecture and to establish it it suffices to find one such Ω -complete theory. In this way, by presenting mathematically precise scenarios that are sensitive to mathematical developments, the pluralist and non-pluralist can give the question of pluralism “mathematical traction” and, through time, test the robustness of mathematics.

What can we say at present concerning the above pluralist scenario? Although the scenario is an open mathematical possibility there are reasons to think that such a scenario cannot happen. For there is growing evidence for the Strong Ω Conjecture⁷⁸ and, as noted above, this conjecture rules out the existence of one (and hence many) such Ω -complete theories. Thus, one way to definitively rule out the above pluralist scenario is to prove the Strong Ω Conjecture.

Should it turn out that the Strong Ω Conjecture is true then the pluralist would have to retreat and present another scenario. Let us consider two such scenarios.

The first scenario builds on Theorem 5.2 which shows that CH is a Σ_1^2 -sentence such that $ZFC + L + CH$ is Ω -complete for Σ_1^2 (where L is a large cardinal axiom) and, moreover, that CH is the unique such sentence (up to Ω equivalence). If the Strong Ω Conjecture holds then (by Theorem 5.4) this result is close to optimal in that there is no recursively enumerable theory A such that $ZFC + A$ is Ω -complete for Σ_3^2 . However, it is an open possibility that there an axiom A such that for some large cardinal axiom L , $ZFC + L + A$ is Ω -complete for Σ_2^2 . Let us assume that this possibility is realized and consider the question of uniqueness. For each A such $ZFC + L + A$ is Ω -complete for Σ_2^2 (where L is a large cardinal axiom) let T_A be the Σ_2^2 theory computed by $ZFC + L + A$ in Ω -logic. The question of uniqueness simply asks whether T_A is unique. A refinement of the techniques used to prove Theorem 5.5 can be used to show that uniqueness must fail. The first pluralist scenario is this: There are incompatible Ω -complete pictures of Σ_2^2 (granting large cardinal axioms) and the choice between them is one of mere

⁷⁸See Woodin (2009).

expedience.

The tenability of this scenario rests on the impossibility of giving theoretical reasons for one such T_A over another. Remarkably, from among all of the theories T_A there is *a single one* that stands out. For it is known (by a result of Woodin in 1985) that if there is a proper class of measurable Woodin cardinals then there is a forcing extension satisfying all Σ_2^2 sentences φ such that $\text{ZFC} + \text{CH} + \varphi$ is Ω -satisfiable. (See Larson, Ketchersid & Zapletal (2008).) It follows that if the question of existence is answered positively with an A that is Σ_2^2 then T_A must be this maximum Σ_2^2 theory and, consequently, all T_A agree when A is Σ_2^2 .⁷⁹ So, assuming that all such T_A contain CH and that there is a T_A where A is Σ_2^2 , then, although not all T_A agree (when A is arbitrary) there is one that stands out, namely, the one that is maximum for Σ_2^2 sentences.

The second scenario is based on Theorem 5.3 which shows that (granting the Strong Ω Conjecture) there is an axiom A such that $\text{ZFC} + A$ is Ω -complete for $H(\omega_2)$ and, moreover, any such axiom has the feature that $\text{ZFC} + A \models_{\Omega} "H(\omega_2) \models \neg\text{CH}"$. For each such axiom A let T_A be the theory of $H(\omega_2)$ as computed by $\text{ZFC} + A$ in Ω -logic. Thus, the theorem shows that all such T_A agree in containing $\neg\text{CH}$. The question then naturally arises whether T_A is unique. A refinement of the techniques used to prove Theorem 5.5 can be used to answer this question negatively. And again the pluralist might use this “local bifurcation” result to ground the case for pluralism. But again, there is a T_A that stands out, namely, the maximum theory given by the axiom (*). (See Woodin (1999).)

In the course of this paper we have seen a number of increasingly sophisticated cases for pluralism. In each instance the case faltered for reasons that are sensitive to actual developments in mathematics. Perhaps there are deeper theorems in this vein that would lead us to embrace pluralism. The point I wish to make is that the real question of pluralism is a deep one, one that requires the combined efforts of philosophy and mathematics. It is through exploring the boundless ocean of unlimited possibilities that we can gain a sure footing in what is actual.

⁷⁹A natural conjecture is that \diamond is such an A . But even if \diamond is not such an axiom A it will be in T_A .

References

- Carnap, R. (1934). *Logische Syntax der Sprache*, Springer, Wien. Translated by A. Smeaton as *The Logical Syntax of Language*, London: Kegan Paul, 1937.
- Carnap, R. (1939). Foundations of Logic and Mathematics, *International Encyclopedia*, vol. I, no. 3, University of Chicago Press, Chicago.
- DiSalle, R. (2006). *Understanding Spacetime: The Philosophical Development of Physics from Newton to Einstein*, Cambridge University Press, Cambridge.
- Evans, J. (1998). *The History and Practice of Ancient Astronomy*, Oxford University Press.
- Friedman, M. (1992). *Kant and the Exact Sciences*, Harvard University Press, Cambridge, Massachusetts.
- Friedman, M. (1999a). *Dynamics of Reason: The 1999 Kant Lectures at Stanford University*, CSLI Publications.
- Friedman, M. (1999b). *Reconsidering Logical Positivism*, Cambridge University Press.
- Friedman, M. (1999c). Tolerance and analyticity in Carnap's philosophy of mathematics, in (*Friedman 1999b*), Cambridge University Press.
- Gödel, K. (1946). Remarks before the Princeton bicentennial conference on problems in mathematics, in (*Gödel 1990*), Oxford University Press, pp. 150–153.
- Gödel, K. (1953/9). Is mathematics syntax of language?, in (*Gödel 1995*), Oxford University Press, pp. 334–362.
- Gödel, K. (1990). *Collected Works, Volume II: Publications 1938–1974*, Oxford University Press, New York and Oxford.
- Gödel, K. (1995). *Collected Works, Volume III: Unpublished Essays and Lectures*, Oxford University Press, New York and Oxford.

- Goldfarb, W. & Ricketts, T. (1992). Carnap and the philosophy of mathematics, in D. Bell & W. Vossenkuhl (eds), *Science and Subjectivity*, Akademie, Berlin, pp. 61–78.
- Janssen, M. (2002). Reconsidering a scientific revolution: The case of Einstein versus Lorentz, *Physics in Perspective* 4(4): 421–446.
- Kleene, S. C. (1939). Review of *The Logical Syntax of Language*, *Journal of Symbolic Logic* 4(2): 82–87.
- Koellner, P. (2006). On the question of absolute undecidability, *Philosophia Mathematica* 14(2): 153–188. Revised and reprinted in *Kurt Gödel: Essays for his Centennial*, edited by Solomon Feferman, Charles Parsons, and Stephen G. Simpson. Lecture Notes in Logic, 33. Association of Symbolic Logic, 2009.
- Koellner, P. (2009). Carnap on the foundations of logic and mathematics. Unpublished.
- Koellner, P. & Woodin, W. H. (2009a). Incompatible Ω -complete theories, *The Journal of Symbolic Logic*. Forthcoming.
- Koellner, P. & Woodin, W. H. (2009b). Large cardinals from determinacy, in M. Foreman & A. Kanamori (eds), *Handbook of Set Theory*, Springer. Forthcoming.
- Larson, P., Ketchersid, R. & Zapletal, J. (2008). Regular embeddings of the stationary tower and Woodin’s Σ_2^2 maximality theorem. Preprint.
- Lindström, P. (2003). *Aspects of Incompleteness*, Vol. 10 of *Lecture Notes in Logic*, second edn, Association of Symbolic Logic.
- Parsons, C. (2008). *Mathematical Thought and its Objects*, Cambridge University Press.
- Quine, W. V. (1963). Carnap and logical truth, in P. A. Schilpp (ed.), *The Philosophy of Rudolph Carnap*, Vol. 11 of *The Library of Living Philosophers*, Open Court, pp. 385–406.
- Reichenbach, H. (1920). *Relativitätstheorie und Erkenntnis Apriori*, Springer, Berlin. Translated by M. Reichenbach as *The Theory of Relativity and A Priori Knowledge*. Los Angeles: University of California Press, 1965.

- Reichenbach, H. (1924). *Axiomatik der relativistischen Raum-Zeit-Lehre*, Vieweg, Braunschweig. Translated by M. Reichenbach as *Axiomatization of the Theory of Relativity*. Los Angeles: University of California Press, 1969.
- Tait, W. W. (1986). Truth and proof: The Platonism of mathematics, *Synthese* (69): 341–370. Reprinted in (Tait 2005).
- Tait, W. W. (2005). *The Provenance of Pure Reason: Essays in the Philosophy of Mathematics and Its History*, Logic and Computation in Philosophy, Oxford University Press.
- Wilson, C. (1970). From Kepler’s laws, so-called, to universal gravitation: Empirical factors, *Archive for History of Exact Sciences* **6**(2): 89–170.
- Woodin, W. H. (1999). *The Axiom of Determinacy, Forcing Axioms, and the Nonstationary Ideal*, Vol. 1 of *de Gruyter Series in Logic and its Applications*, de Gruyter, Berlin.
- Woodin, W. H. (2009). *Suitable Extender Sequences*. To appear.