THE MATHEMATICIZATION OF NATURE

Jeffrey John Ketland

For Research Degree of PhD in Philosophy

London School of Economics and Political Science

Houghton Street, London

WC2A 2AE

13th August, 1998

ABSTRACT

This thesis defends the Quine-Putnam indispensability argument for mathematical realism and introduces a new indispensability argument for a substantial conception of truth.

Chapters 1 and 2 formulate the main components of the Quine-Putnam argument, namely that virtually all scientific laws quantify over mathematical entities and thus logically presuppose the existence thereof. Chapter 2 contains a detailed discussion of the logical structure of some scientific theories that incorporate or apply mathematics. Chapter 3 then reconstructs the central assumptions of Quine's argument, concluding (provocatively) that "science entails platonism".

Chapter 4 contains a brief discussion of some major theories of truth, including deflationary views (redundancy, disquotation). Chapter 5 introduces a new argument against such deflationary views, based on certain logical properties of truth theories. Chapter 6 contains a further discussion of mathematical truth. In particular, non-standard conceptions of mathematical truth such as "if-thenism" and "hermeneuticism".

Chapter 7 introduces the programmes of reconstrual and reconstruction proposed by recent nominalism. Chapters 8 discusses *modal* nominalism, concluding that modalism is implausible as an interpretation of mathematics (if taken seriously, it suffers from exactly those epistemological problems allegedly suffered by realism). Chapter 9 discusses Field's deflationism, whose central motivating idea is that mathematics is (*pace* Quine and Putnam) dispensable in applications. This turns on a conservativeness claim which, as Shapiro pointed out in 1983, must be incorrect (using Gödel's Theorems).

I conclude in Chapter 10 that nominalistic views of mathematics and deflationist views of truth are both inadequate to the overall explanatory needs of science.

CONTENTS

PREFACE

The problem of how mathematics relates to science (in particular, to physics) has long fascinated me. As a teenager, I was intrigued and puzzled by the fact that physical projectiles move in parabolas, curves that satisfy a simple mathematical equation. Why should physical objects obey mathematical laws? Although I became a theoretical physics graduate, this philosophical puzzle has always interested me.

Several philosophers, mathematicians and physicists have discussed this problem, but illumination is rare. Discussions usually terminate in a "credo", which could be realism (Gardner, Penrose, Stewart), "conceptualism" (Popper, Hersh & Davis) or even formalism (Hilbert, E.T. Bell).

The scales fell from my eyes when I happened quite accidentally upon a second-hand anthology, *From a Logical Point of View*, containing an article 'On What There Is', by a certain Willard Van Orman Quine. At last, an *argument* is presented to the effect that a theory is *about* those entities which its quantifiers must *range over* if it is to be *true*. Furthermore, because natural science incorporates a great deal of mathematics, quantifying in its laws over mathematicalia, any such theory of modern science is committed to mathematicalia. As it transpired, Quinian arguments had been further developed by Putnam in 1971, and had attracted serious attempts at refutation by several recent philosophers of mathematics, notably Hartry Field and Charles Chihara.

Having completed my MA in Philosophy at the University of Warwick in 1992, working in the philosophy of language, logic and science, I decided to work on this philosophical problem of the role of mathematics within science. Joining the London School of Economics in 1994, this thesis now represents four years of research, under the supervision of Professor John Worrall. It explores the application of mathematics in science, and the consequential implications for the interpretation of mathematics. The main conclusion is that if (as scientific realists sensibly claim) we are entitled to assume that a great deal of *natural science* is true, then mathematical realism is unavoidable.

A further, equally important, topic of this thesis concerns the nature of *truth*. I originally felt some sympathy for deflationism, the doctrine that truth is ultimately a rather simple "insubstantial" concept, governed by the totality of so-called "T-sentences", of the form,

'*p*' is true if and only if *p*.

Motivated by an analogy with Hartry Field's deflationism about *mathematics*, I guessed that the totality of such sentences would be conservative: truth-theoretic reasoning should never generate anything "genuinely new". The conservativeness of all the T-sentences would partially explain their obviousness or their status as partial definitions. In fact, the conservativeness of the set of such T-sentences over *any* theory in the object language, is not difficult to prove (indeed Tarski proves it in his 1936). But Tarski's full theory of truth cannot (as John Burgess pointed out to me in a letter) be similarly conservative. For by adding Tarski's theory of truth to Peano Arithmetic, one can prove that Peano Arithmetic is *consistent*. If Tarski's theory were conservative, this would contradict Gödel's Incompleteness Theorem. Alerted to this logical fact, it occurred to me that *any adequate truth theory* must be non-conservative. It follows that deflationism about truth must be wrong. The result is Chapter 5 (which will appear in *Mind*).

(After I finished this work on truth in December 1997, I received an article from Stewart Shapiro, entitled 'Truth and Proof - Through Thick and Thin', containing a similar argument to the argument sketched above. I have since benefited from a rewarding correspondence with him.)

I thank John Worrall of LSE for his careful supervision of my thesis, for reading countless versions as it gradually developed, and for focusing my style. I also thank David Miller of the University of Warwick for his continuing encouragement of my work. Finally, I acknowledge some technical help from John Burgess at Princeton, who pointed out to me the "well-known fact" that Tarski's theory of truth must be non-conservative.

Most importantly, I thank my wife Blanca Fuentes and my mother Patricia Hall for their kindness, inspiration and love.

CHAPTER 1

The Application of Mathematics

The essential fact is simply that all the pictures which science now draws of nature, and which alone seem capable of according with observational fact, are mathematical pictures.

Sir James Jeans 1930, *The Mysterious Universe*, p. 153.

Philosophically, the applied mathematician is an uncritical Platonist. He takes for granted that *there is a function u(t)*, and that he has a right to use any method he can think of to learn as much as he can about it.

Philip J. Davis & Reuben Hersh 1980 (1990), *The Mathematical Experience*, p. 378.

Most of the literature in the philosophy of mathematics takes the following three questions as central:

(a) How much of mathematics is true? For example, are conclusions arrived at using impredicative set theory true?

(b) What entities do we have to postulate to account for the truth of (this part of) mathematics?

(c) What sort of account can we give of our knowledge of these truths?

A fourth question is sometimes discussed, though usually quite cursorily:

(d) What sort of account is possible of how mathematics is applied to the physical world?

Now, my view is that question (d) is the really fundamental one.

Hartry Field 1980, *Science Without Numbers*, p. vii.

1.1 The Application of Mathematics

Why, in attempting to understand Nature, do we "mathematicize" Her? Why should the attempt to explain and predict the behaviour of concrete physical objects and systems involve such ubiquitous use of mathematical descriptions?

A theory from pure mathematics (such as Peano Arithmetic, Real Analysis or Zermelo-Fraenkel set theory) might be thought of as either,

i. a formal system of uninterpreted axioms, or, alternatively, as,

ii. a fully interpreted description of the properties of certain abstract entities: mathematicalia, like natural numbers, real numbers, functions, sets and structures.

It is perhaps puzzling, on both of these pictures, that the application of mathematics is even possible. But it is not only possible: it seems to be *indispensable*. A physicist (or someone familiar with physical laws) would find it hard even to imagine explanations of physical phenomena without talk of measurable quantities, real- and vector-valued field quantities defined over space-time, governed by differential equations, conservation laws, mathematically formulated symmetry principles, and so on.

Henceforth, we shall refer, somewhat loosely, to the use or application of mathematics within science as "mathematicization" and we shall refer to theories that use mathematical concepts to describe Nature as "mathematicized theories".

Mathematicization goes back a long way. One thinks primarily of *counting* (i.e., arithmetic) and *measuring*, which lead to both geometry (literally "earth measurement") and analysis (the theory of continuous quantities). Mathematicization certainly took off in Antiquity around 500 BC, with the Pythagoreans and the Miletians. Indeed, the Pythagoreans were so captivated by mathematical explanations that one of their slogans, along with not eating beans, was that "all things are numbers". By the time of Plato's Academy around 380BC, mathematics lay at the heart of the curriculum.

However, we begin our study two thousand years later, with one of the founders of the scientific revolution in Europe, with Galileo's famous declaration of 1618:

> Philosophy is written in this grand book—I mean the universe—which stands continually open to our gaze, but it cannot be understood unless one first learns to comprehend the language and interpret the characters in which it is written. It is written in *the language of mathematics*, and its characters are triangles, circles and other geometrical figures, without which it is humanly impossible to understand a single word of it.
>
> (Galileo 1618 (1960), p. 183-184. Emphasis added).

The newly emerging experimental or natural philosophy (i.e., mechanics and astronomy) is to be formulated mathematically. Galileo is perhaps advocating a (generalized) *Pythagoreanism,* the view that the universe is itself fundamentally mathematical[1].

The reasons for Galileo's view are complex, having to do with the debates in astronomy about "saving the appearances"[2], his distinction between the primary qualities of bodies (shape, size, number, position, quantity of motion: qualities independent of human sensation) and secondary qualities (colours, tastes, odours and sounds: relational qualities dependent upon human sensation), and his frequent discussion of the method of idealization and abstraction. Roughly, but not inaccurately, Galileo agreed with Plato that although the appearances (or phenomena) can be complex and messy, they are best viewed as reflecting a hidden reality properly described by mathematics.

Three hundred years later, in the 1920s, two highly successful British scientists, Jeans and Eddington[3], took up the Pythagorean-Platonic view. Echoing Plato, Jeans wrote:

> From the intrinsic evidence of his creation, the Great Architect of the Universe now begins to appear as a pure mathematician.
>
> (Jeans 1930, p. 167).
>
> ... the final truth about a phenomenon resides in the mathematical description of it; so long as there is no imperfection in this our knowledge of the

---

[1] It should be noted, however, that 'geometry' is ambiguous. It can be construed, along with Euclid (and Einstein), as a part of *physics*: the study of physical space, with a subject matter of *physical* "geometricalia": points and regions. Geometry can, in contrast, be construed as the study of purely mathematical spaces (e.g., topological manifolds, orbifolds, etc.) and then becomes a part of *pure mathematics*. It is not clear how Galileo would have sided on these deeply distinct construals of geometry.

Within the purely physical interpretation of geometry, "triangles, circles and other geometrical figures" may be treated as allegedly nominalistically acceptable *concrete physical entities*. This is Field's approach, sometimes referred to as "geometrical nominalism", discussed in Chapter 9 below.

[2] C.f., Osiander's notorious preface to the first edition of Copernicus 1543, *De Revolutionibus Orbium Caelestium*. Osiander, a cleric, argued that the heliocentric hypothesis is nothing more that a "mathematical model" for "saving the appearances" of astronomical observation. It is more *convenient* that the Ptolemaic theory, but not any more *true* as a description of the real motions of the Earth and other bodies.

[3] Jeans was Astronomer Royal and worked on stellar combustion. Eddington's 1919 solar eclipse expedition led to the discovery that light was deflected by the Sun *twice as much* as Newton's theory predicted and *precisely as much* as Einstein's theory predicted: Newtonian Mechanics was finally dethroned (this event was to inspire the teenage Popper in his search for a Demarcation Principle). Quizzed by a journalist, "Apparently, only three people in the world understand Einstein's relativity theory", Eddington famously quipped, "Who's the third?".

> phenomenon is complete ... the making of models or pictures to explain mathematical formulae and the phenomena they describe, is not a step towards, but a step away from, reality.
>
> (Jeans 1930, p. 177).

Eddington added:

> Nowadays we do not encourage the engineer to build the world for us out of his material, but we turn to the mathematician to build it out of his.
>
> (Eddington 1928, p. 209).

These forays by scientists into philosophical matters did not go unnoticed by professional philosophers. For example, when Stebbing[4] came to discuss these doctrines of Jeans and Eddington in her 1937, she dismissed them as "at best misleading and, all too often, extravagant nonsense"[5]. Slightly later, but in a similar spirit, Russell added:

> Pythagoras, as everyone knows, said that "all things are numbers". This statement, interpreted in a modern way, is logically nonsense, but what he meant was not exactly nonsense. ... It is only in recent times that it has been possible to say clearly where Pythagoras was wrong.
>
> (Russell 1945, p. 54 & p. 56).

In a witty and memorable passage, Popper argued against the idea that Nature herself is mathematical:

> Similarly, the success, or even the truth, of simple statements, or of mathematical statements, or of English statements, ought not to tempt us to draw the inference that the world is intrinsically simple, or mathematical, or British. All these inferences have been drawn by some philosopher or other; but on reflection, there is little to recommend them.
>
> (Popper 1982, p. 43).

Nevertheless, despite this philosophical scepticism, many scientists who reflect on the use (even the success) of mathematics in describing the physical world find themselves led rightly or wrongly to the Pythagorean-Platonic conclusion. A recent example is Penrose:

> To speak of Plato's world at all, one is assigning some kind of reality to it which is in some way comparable to the reality of the physical world. On the other hand, the reality of the physical world itself seems more nebulous than it had seemed before the advent of the SUPERB theories of relativity and quantum mechanics. The very precision of these theories has provided an almost abstract

[4] Stebbing 1937.

[5] Quoted in Gjertsen 1989, p. 146.

> mathematical reality for actual physical reality. Is this in any way a paradox? How can concrete reality become abstract and mathematical? This is perhaps the other side of the coin to the question of how abstract mathematical concepts can achieve an almost concrete reality in Plato's world. Perhaps in some sense *the worlds are actually the same.*
>
> (Penrose 1989, p. 556).

In a similar vein, Gardner has written:

> To a realist, mathematical structure is mind-independent in two senses. The universe is not shapeless, but patterned in ways that are described by mathematics. In addition, mathematicians investigate purely abstract structures, defined by formal systems, which may or may not have applications to the physical world. The proper attitude to take toward the ontological status of these abstract systems is, of course, one of the great unending controversies in metaphysics. I will here say only that almost all mathematicians agree with Hardy that a mathematician discovers truths that are independent of his culture and that those truths are qualitatively different from the conventions of traffic regulations or codes of etiquette.
>
> (Gardner 1996, p. 268).

1.2 Explaining the Role of Mathematics Within Science

To evaluate the debate concerning the implications of the application of mathematics, we need to go into more detail about how mathematics is used within science. With this in mind, let us look at the following passage from the Introduction to a standard textbook on Classical Mechanics:

> In fact, mechanics—and indeed all theoretical science—is a game of mathematical make-believe. We say: If the Earth were a homogeneous rigid ellipsoid acted on by such and such forces, how would it behave? Working out the answer to this mathematical question, we compare our results with observation. If there is agreement, we say we have chosen a good model; if disagreement, then the models or laws assumed are bad.
>
> (Synge & Griffith 1959, p. 5).

I take this to be a promising account of how mathematics actually is applied. First, *pure* mathematics tells us facts about a certain kind of abstract mathematical entity, an *ellipsoid*, or perhaps an ellipsoid embedded in abstract Euclidean 3-space. Pure mathematics tells us about the properties of this abstract entity: its topology and geometry. Next, a mathematical description of some concrete situation is introduced: this description postulates some *contingent connection* between this mathematical entity and the concrete system, say the Earth or the solar system. And given this conjectured Earth-ellipsoid relation, the mathematical facts about the ellipsoid somehow yield, or "translate into" or are "interpreted as giving", *non-mathematical facts* about the Earth. For example, a piece of pure mathematics will determine the sum of internal angles of a large "triangle" on the surface (say, the "triangle" formed by joining Dublin, Birmingham and Madrid with geodesics).[6]

I will give an illustration of how this translation might work later in my discussion of a certain simplified theory of time, which states simply that time is isomorphic to the continuum. For example, the density and continuity of the real numbers (under the usual total order <) translates into the density and continuity of the set of temporal instants (under the temporal "before" relation).

The authors continue:

> Let us now sum up the general procedure in theoretical mechanics in the following five steps:
>
> (1) A physical system is an object of curiosity; we wish to predict its behaviour under various circumstances (The system might be a pendulum, or a pair of stars attracting one another).
>
> (2) An *ideal* or *mathematical model* of the physical system is constructed mentally. (The pendulum is regarded as a rigid straight line and the stars are regarded as two particles).
>
> (3) Mathematical reasoning is applied to the mathematical model. (This means that differential or finite equations are set up and solved. Formulas are developed to give answers to interesting questions, such as those concerning the periodic time of the pendulum or the orbit of one star relative to the other).
>
> (4) The mathematical results are *interpreted physically* in terms of the physical problem.
>
> (5) The results are compared with the results of observation, if possible.
>
> (Synge & Griffith 1959, pp. 5-6. Emphasis added).

Much of this provokes philosophical puzzlement. There are at least two basic questions:

i. what is an *ideal* or *mathematical model* of a physical system?

[6] Some philosophers of mathematics, especially structuralists, suggest that we may say that the *connection* between concrete systems (e.g., the Earth) and their mathematical representations (e.g., the ellipsoid) is this:

> the concrete system *exemplifies* (at least approximately) the mathematical entity.

See, for example, Shapiro 1997. I discuss this important idea a little more below in Chapter 2.

ii. how are mathematical results (about the model) *interpreted physically*?

Unfortunately, what many mathematicians, even philosophically sensitive ones, say about the application of mathematics is sometimes not particularly helpful. An example is Stewart:

> Mathematics is about ideas; about how certain facts follow inevitably from others; about how certain structures automatically imply the occurrence of particular phenomena. It lets us build mental models of the world, and manipulate those models in ways that would be impossible in a real experiment.
>
> You want to know how far a javelin would travel in Jupiter's gravity? Plug in the equations and set g equal to 27 m sec^{-2}, and work it out. You don't have to go to Jupiter and throw one.
>
> Mathematics is about ideas per se, not just about specific realizations of those ideas ... its very abstraction gives it a unity and universality that would be lost if it were made more concrete and specialized.
>
> (Stewart 1987 (1996), pp. 319-320).

It is not difficult to criticize this account. What does it mean to say "that mathematics is about ideas"? Perhaps in the sense that trained mathematicians possess (and presumably create or discover) lots of ideas, like the idea of surjectivity or infinitude, which can be linguistically expressed and thereby communicated. But this is surely equally true of economists, biologists and children. Any science is "about ideas", in this sense. The question should be: what is so distinctive about *mathematical* ideas?

Is it true that mathematics is about "how certain facts follow inevitably from other ones"? This seems very much like an unargued conflation of mathematics with logic. The general applicability of logic is easily explained, for logic is concerned with how certain propositions follow automatically from others and these propositions can be about whatever you like (indeed, the implications hold irrespective of what they are about: an implication holds purely in virtue of logical form). Of course, there is no denying that mathematical reasoning is logical, but so is *any correct reasoning* (e.g., in economics). But is mathematics nothing more than logical reasoning? In particular, can the *truth* of the axioms of, say, set theory or analysis be established purely through logical reasoning?

If their truth is not established purely logically, perhaps they are merely arbitrary *stipulations* or (consistent) *conventions*[7]. In this case, two enormous puzzles appear:

i. why do we concentrate on just *these* axioms and not others?

ii. why are these axioms so *fruitful* in our mathematicized empirical theories?

If mathematics is about how "certain structures automatically imply the occurrence of certain phenomena", then that is merely a restatement of the problem of application and demands explanation. What is a *structure*? How is a structure *connected* to (related to) the phenomena whose occurrence it implies? Exemplification? What is that?

Suppose I want to know how far a javelin on Jupiter would travel. If mathematics were "purely verbal", or "purely logical", or "purely conventional", or "stipulative", why should these purely verbal mathematical equations help me determine the *physically correct* answer?

For another example, consider the following summary by Davis and Hersh:

> Mathematics has penetrated sociology, psychology, medicine and linguistics ... Why is this so? What gives mathematics its power? What makes it work?
>
> On very popular answer has been that God is a mathematician ... the universe expresses itself naturally in the language of mathematics ... Mathematics, in this view, has evolved precisely as a symbolic counterpart of the universe. It is no wonder, then, that mathematics works; that is exactly the reason for its existence. The universe has imposed mathematics upon humanity ...
>
> But there is another view of the matter. This opinion holds that applications of mathematics come by fiat. We create a variety of mathematical patterns or structures. We are then so delighted with what we have wrought, that we deliberately force various physical and social aspects of the universe into these patterns as best we can ... This view is related to the opinion that theories of applied mathematics are merely "mathematical models". The utility of a model is precisely its success in mimicking or predicting the behaviour of the universe. If a model is inadequate in some respect, one looks around for a better model or improved version ... [which model] we operate with is determined by such things

[7] An early but influential publication, Quine 1936 ('Truth by Convention') attacked these ideas. One of his arguments (related to Carroll's piece 'What the Tortoise said to Achilles') was that logical rules cannot consist of stipulations about what is to be true: for we should require *infinitely many* such stipulations. Of course, we can get all the (elementary) logical truths from a finite list of axioms, but then one needs to *use* the (sound & complete) *rules* of logic in order to perform the needed inferences.

However, the idea that the "meanings" of the logical constants $\neg$, $\wedge$, $\rightarrow$, $\exists$, etc., are "implicitly defined" *proof-theoretically* (by the introduction and elimination rules governing them) is still a popular, although controversial, one. See Dummett 1991.

> as simplicity, fruitfulness, etc. ... [models are] derived from prior mathematical experiences of a simple nature.
>
> ... Truth has abdicated and expediency reigns.
>
> (Davis & Hersh 1980 (1990), pp. 68-70).

Again, philosophical questions flood in. Do we really *create* mathematical structures? This sounds like idealism. Once "created", do these structures then go on to exist objectively? For example, is their existence spatio-temporal? If so, what are the physical properties (e.g., mass, optical reflectivity, spatio-temporal extent) of, say, the Cantor's second transfinite cardinal, $\aleph_1$? Presumably (of course) $\aleph_1$ is *not* a spatio-temporal entity. But, if this hint of constructivism-idealism were correct, then what generated, and what sustains, $\aleph_1$'s disembodied existence? Surely, we do not in any interesting sense "create" mathematical structures, any more than we "create" neutron stars. Either *there simply is no such thing* as $\aleph_1$ (and our talk using the symbol '$\aleph_1$' is a mere formal game or a useful fiction) or *there is such a thing as* $\aleph_1$, namely an abstract entity, which Cantor discovered, labelled and whose properties he investigated.

Furthermore, the notion that we "force physical aspects of the universe into a pattern" is one advocated by Galileo and most Platonists/realists: this is abstraction and idealization. But, in any case, how are such abstracta (like patterns) *related* to concreta, like physical systems and objects in the Universe? How are these patterns "derived from prior mathematical experiences of a simple nature"? Is this meant as a statement of mathematical structuralism? Maybe we somehow perceive simple (finitary) patterns objectively "exemplified" in the physical world, and then go on to postulate the existence of more complex, perhaps infinitary, patterns (like the continuum) in our mathematical investigations. Some of these may be "exemplified" (e.g., the continuum), some may not (e.g., models of axiomatic set theory with inaccessible cardinals).

I would suggest that Davis and Hersh are not keeping separate issues separate. They combine philosophical intuitions from *every* philosophy of mathematics in the space of six or seven sentences, rather than developing precise arguments about ontology, epistemology and semantics. In short, Davis and Hersh—like many mathematical writers tackling this difficult subject—take themselves to be advocating some particular position, while in reality what they actually adopt is almost unidentifiable: it could be idealism, structuralism, Platonism, or formalism. They seem to embrace everything that has been proposed. Having outlined this amalgam of everything, they then mischievously give it their favourite label (which happens to be "mathematical modelling").

Davis and Hersh conclude their book with the following passage:

> Mathematics is not the study of an ideal, pre-existing nontemporal reality. Neither is it a chess-like game with made up symbols and formulas. Rather it is part of human studies which is capable of achieving science-like consensus ...
>
> Mathematics does have a subject matter, and its statements are meaningful. Their meaning, however, is to be found in the shared understanding of human beings, not in an external nonhuman reality. In this respect, mathematics is similar to an ideology, a religion or an art form: it deals with human meanings, and is intelligible only within the context of culture. In other words, mathematics is a humanistic study. It is one of the humanities.
>
> (Davis & Hersh 1980 (1990), p. 410).

Even abstracting from the fact that I disagree with this, I cannot say that their "mathematics is part of culture" summary advances the debate. *Of course*, mathematics is a part of human culture. No-one disputes *that*[8]. Similarly, neutrino physics is a part of human culture, but neutrinos are not produced by human brains. Indeed, according to physics, zillions of solar neutrinos pass right through our brains every second, without any interaction. It would be ridiculous to say that neutrino physics deals with "human

[8] Gardner's review of *The Mathematical Experience* by Davis and Hersh is entitled 'How Not to Talk About Mathematics', and is reprinted in Gardner 1996:

> "For a mathematical realist a tree not only exists when nobody looks at it, but its branches have a "tree" pattern even when no graph theorist looks at it ... The existence of an external world, mathematically ordered, is taken for granted. I have yet to meet a mathematician willing to say that if the human race ceased to exist the moon would no longer be spherical. I suspect Davis and Hersh would not care to say this, yet the troubling thing about their book is that it does not make clear why." (p. 281)
>
> "... All that mathematician do is certainly part of culture for the simple reason that everything human beings do is part of culture ... Conceptualism in mathematics has its strongest appeal amongst anthropologists and sociologists who have a vested interest in making culture central. It is a language that also appeals to those historians, psychologists and philosophers who cannot bring themselves to talk about anything that transcends human experience." (p. 282).
>
> "... No mathematician hesitates to speak about "existence proofs" about objects even when they are nowhere modeled, or known to be modeled, by the external world. And most mathematicians, including the very greatest, think of such objects as independent of the human mind, though not of course existing in the same way Mars exists." (p. 284)

meanings". One is reminded of Mill's famous put-down: Davis and Hersh "must have made some progress in philosophy" to arrive at such views.

1.3 Applicability and Mathematical Idealism

Intuitionism, espoused by Brouwer, Weyl, Heyting and Dummett, represents a version of idealism (psychologism, conceptualism, empiricism or subjectivism) about mathematics and mathematicalia. Intuitionism was popular with a few mid-century empiricists, who thought of themselves (along with their positivist allies) as overcoming something derogatively called "metaphysics".

From an ontological point of view, intuitionism agrees with platonism in holding that mathematical entities *exist*, but suggests that mathematicalia are *mental* entities which somehow come into being as a result of human thought processes (or at least, through construction processes)!

Heyting put the view as follows,

> Intuitionist mathematics consists ... in mental constructions; a mathematical theorem expresses a purely empirical fact, namely the success of a certain construction. '2 + 2 = 3 + 1' must be read as an abbreviation for the statement: "I have effected the mental constructions indicated by '2 + 2' and "3 + 1" and I have found that they lead to the same result". ...
>
> The characteristic of mathematical thought is, that it does not convey truth about the external world, but is only concerned with mental constructions ...
>
> In fact, mathematics, from the intuitionistic point of view, is a study of certain functions of the human mind.
>
> (Heyting 1956 (1983), pp. 72-73).

The canonical objection to idealism is that it implies that if thinking (e.g., mathematical thinking) about some kind of things (the *F*s say) had not existed, then the *F*s themselves (the mathematicalia, such as π or Cantor's $\aleph_1$) would not have existed. The existence of *F*s is *ontologically dependent* upon the existence of thinking: *esse est concipi*. Furthermore, since thinking is often presumed to be limited to *finite* processes of construction, intuitionism expresses scepticism as to whether such a "completed" infinite entity as $\aleph_1$ exists. Irrespective of its critique of proof methods and its repudiation of "completed" infinities, all this makes intuitionism *ontologically* difficult to interpret.

For example, does π "gradually come into existence" as its successive digits are calculated (say, using Euler's series). The decimal expansion of π is now known to millions of digits (we let computers do this dirty work). But did π exist prior to this (partial) determination? Surely, it is more correct to say, along with Hardy, Penrose and countless others, that a computer evaluating this expansion is "uncovering something already there", that is, it is *discovering* what π is and always has been, and this fact would still have existed even if the computer had not been switched on.

It is difficult to see what the argument is for supposing that mathematicalia are mental entities. The fact that mathematicians think a lot does not imply that what they think *about* is itself mental. Presumably, idealism requires at least a plausible criterion whereby a mental entity x created in the mind of person A is the same as some mental entity y created in the mind of person B. But does it make sense to say that the *same* thing exists in two different minds? If not, then there will simply be many mental, purely "subjective versions" of each mathematical object (e.g., "π versions"), and no such objective thing as π itself. Thus, intuitionism threatens to make mathematics completely subjective.

Long ago, Frege made similar criticisms of the view that "numbers are ideas":

> If number were an idea, then arithmetic would be psychology. But arithmetic is no more psychology than, say, astronomy is. Astronomy is concerned, not with ideas of the planets, but with the planets themselves, and by the same token the objects of arithmetic are not ideas either. If the number two were an idea, then it would have straight away to be private to me only. Another man's idea is, ex vi termini, another idea. We should then have it might be many millions of twos on our hands. We should have to speak of my two and your two, of one two and all twos. If we accept latent or unconscious ideas, we should have unconscious twos among them, which would then return subsequently to consciousness. As new generations of children grew up, new generations of twos would continually be being born, and in the course of millennia these might evolve, for all we could tell, to such a pitch that two of them would make five. ...
>
> Weird and wonderful, as we see, are the results of taking seriously the suggestion that number is an idea. And we are driven to the conclusion that number is neither spatial and physical, like Mill's piles of pebbles and gingersnaps, nor yet subjective like ideas, but non-sensible and objective. Now objectivity cannot, of course, be based on any sense-impression, which as an

> affection of our mind is entirely subjective, but only, so far as I can see, on reason.
>
> It would be strange if the most exact of all the sciences had to seek support from psychology, which is still feeling its way none too surely.
>
> (Frege 1884, §27 (1980), pp. 37-38).

Hardy also argued against idealism in mathematics:

> A chair or a star is not in the least what it seems to be like; the more we think of it, the fuzzier its outlines become in the haze of sensation which surrounds it; but 2 or 317 has nothing to do with sensation, and its properties stand out the more clearly the more closely we scrutinize it. It may be that modern physics fits best into some framework of idealistic philosophy—I do not believe it, but there are eminent physicists who say so. Pure mathematics, on the other hand, seems to me a rock on which all idealism founders: 317 is prime, not because we think so or because our minds are shaped in one way or another, but *because it is so*, because mathematical reality is built that way.
>
> (Hardy 1940, p. 70).

However, the primary motivation of intuitionism is not ontological. It is epistemological, and attempts to supply a methodology whereby knowledge of, or access to, mathematicalia is obtained. In particular, this leads to an analysis of *truth* in terms of (or perhaps, a *repudiation* of truth in favour of) proof, and an analysis of proof in terms of construction. According to intuitionism, a mathematical theorem is not true until proved, and is not proved unless proved constructively.

The notion that truth is somehow reducible to proof is a version of so-called *internalist theories of truth*, which are discussed and dismissed below in Chapter 4. One basic problem with such theories (amongst their many problems) is generated by Gödel's First Incompleteness Theorem, which can be formulated (following Tarski) as follows. The class of truths of arithmetic[9] is not even recursively enumerable, and cannot be recursively axiomatized. No matter how powerful a formal system one chooses, not all truths, even of arithmetic, can be proved.

But constructive proof is much more restrictive than classical proof, for example abandoning some kinds of reasoning by *reductio*, impredicative reasoning, and requiring for any existence proof the construction of a specific instance. This leads to what is

[9] That is, the set of sentences in the language of arithmetic that hold in the intended structure $\mathfrak{N}$.

perhaps the canonical objection to the constructivist aspect of intuitionism. Constructivism involves a strong claim about what parts of classical mathematics "make sense", according to its analysis of what constitutes an acceptable or meaningful proof. The traditional cry of intuitionists in response to a non-constructive proof is "That's not mathematics: that's theology!"[10]. The problem, which is not negligible, is that most working mathematicians find such constructivistic demands intolerably restrictive. The history of the Axiom of Choice (and the acceptance of impredicative set theory, as codified within ZFC, as the "standard framework" for mathematics) is an instructive example. Mathematicians of constructivistic temperament, like Lebesgue and Borel, balked at its acceptability, until it was pointed out that their very own work had presupposed Choice all along; and furthermore, that this Axiom of Choice appears indispensable to the proof of a great deal of standard accepted mathematics[11].

The intuitionist at this stage may adopt an extreme response: *dogmatically* dropping (as "meaningless") non-constructive mathematics. Classical mathematicians may protest, but what grounds can they offer for the indispensability of their non-constructive reasoning?

At this stage, another powerful argument comes to light. Modern *science*, and especially modern mathematical physics, uses mathematics that is *non-constructive*[12]. If this is correct, the constructivist philosopher of mathematics left claiming that not only is classical mathematics theology (which is bad enough), but that mathematical physics is theology (which is intolerable)! He or she will be embarrassed to go and tell our modern

---

[10] The realist has an easy riposte: "That's not mathematics: that's solipsism!". Quine has quipped, "I have intuitions too, but they're not the intuitionists' intuitions" (see Ullian 1986).

[11] For a book length discussion, see G.H. Moore 1982, *Zermelo's Axiom of Choice: Its Origins, Development, and Influence*. Drake & Singh 1996 devote Chapter 7 to the Axiom of Choice. Famously, Choice is independent of the other axioms of ZF and is equivalent to several other axioms (modulo ZF): Zermelo's Well-Ordering Theorem, Zorn's Lemma, Hausdorff's Maximality Principle, the Tukey-Teichmuller Lemma. An axiom much studied by set-theorists is the Axiom of Determinacy which in fact contradicts Choice. Famously, the Choice is inconsistent with Quine's set theory, "New Foundations", NF.

[12] Here is an anecdote I once heard at Cambridge. A philosophically-inclined mathematician once explained the idea of "constructive proof" to the cosmologist Steven Hawking. He replied that almost all of his major work on singularity theorems (with Roger Penrose), black hole evaporation, and so on, was in fact *non-constructive*. Briefly, the major theorems are proved by supposing that a non-singular Einsteinian spacetime exists satisfying certain conditions, and deriving a contradiction. By *reductio*, then, for any such spacetime, *there exists* a singular point.

theoretical physicists (the likes of Hawking, Geroch, Thorne, Witten et al.) that their non-constructive inquiries into the mathematical structure of the physical world are meaningless and without substance!

For example, one wonders what Maxwell's Law for the electromagnetic field can mean for an idealist interpretation of mathematics. Maxwell's Law says, in a modern co-ordinate-free formulation,

> There are *continuous tensor-valued functions* $\boldsymbol{F}$ and $\boldsymbol{J}$, fields on physical space-time representing the electromagnetic field and the four-current density respectively, such that,
>
> for each space-time point e, $\mathbf{d}\boldsymbol{F}(e) = 0$ and $\delta^{*}\boldsymbol{F}(e) = \boldsymbol{J}(e)$.

How could an intuitionist reconstrue this clearly mathematicized physical law as a principle concerning what the mind may construct? Would its truth involve some (impossible) mental construction, wherein the mind somehow surveys *each space-time point e*, and "constructs" a pair of *continuous* tensor fields $\boldsymbol{F}$ and $\boldsymbol{J}$ on the space-time continuum, such that they represent the electromagnetic field and current density and satisfy Maxwell's equations? Presumably, in order to meet intuitionist standards, one would have somehow to *prove* Maxwell's Law (which is, after all, a mathematical statement, involving reference to functions and their real number values). But how would one prove it, let alone prove it constructively *ab initio*? Rather, we derive physical predictions and explanations from the law (concerning things like the deflection of electrons in e-m fields) and such predictions either contradict the law or partially support it. Of course, *we* do not construct these electromagnetic fields. They are already part of the furniture of the universe, correlated with the properties of space-time and the motion of electrons and photons. In short, it just isn't clear at all how intuitionism is going to treat mathematical assertions about the physical world, like *all the laws of physics*.

So, mathematical idealism faces prima facie difficulties concerning the *application* of mathematics. Mathematicalia, according to the idealist, are constructible mental entities. But why should these mental entities somehow describe external physical systems? In fact, a "full-blown idealist" like Kant or Fichte may refuse to think of even concreta, like trees or cups, as "external" to our cogitations: these entities are themselves mental entities constituted by the "*a priori* categories of pure reason", in combination with our sensory intuitions.

Field has argued similarly against the idealist philosophy of mathematics:

> ... on a limited idealist view, one that views mathematical entities as some sort of human construction but makes no claim about the physical world, the application of mathematics to the physical world may turn out to be a mystery. The danger, in other words, is that in order to explain the applicability of mind-dependent mathematical entities to the physical world, the idealist about mathematics may have to become a full-blown idealist, and hold that even things like electrons and dinosaurs are somehow 'human constructions'. If this danger were realized, I would regard that as a reductio ad absurdum of the idea that mathematical objects were human constructions.
>
> (Field 1989, Introduction, p. 27. Footnote 16).

In fact, Field repeatedly stresses the *obscurity* of mathematical idealism:

> I take mathematical idealism to be not only too obscure to assert, but also too obscure to deny. It may well be that my own view could be redescribed as a view according to which mathematical entities exist but are mind- or language-dependent.
>
> (Field 1988 (1989), p. 228. Footnote 2).

In what follows, I will follow Frege, Hardy and Field and simply set mathematical idealism aside. Any assumption concerning the *existence* of mathematicalia will be taken in the standard way, as an assumption of the existence of abstract *non-mental* entities, rather than of mental ones. Thus, in Chapter 3 below, the simple formula $\exists x Math(x)$ is introduced to say that *there are* mathematical entities. For us, this statement will count as an assertion of *platonism*.

Let us now turn to the meaningless marks of the formalist.

1.4 Applicability and Formalism

Formalism emerged in the nineteenth century, partly as a response to the discovery of consistent axiomatic systems of non-Euclidean geometry[13]. Formalism, roughly and briefly, is the view that mathematics is merely a *game*, played with *meaningless* or contentless *physical marks.*[14]

An early advocate of formalism in the nineteenth century was a German mathematician J. Thomae, who put the position like this:

> For the formalist, arithmetic is a game with signs, which are called empty. That means they have no other content (in the calculating game) than they are assigned by their behaviour with respect to certain rules of combination (rules of the game).
>
> (Thomae 1898, Introduction)[15].

In the early years of the twentieth century, the foremost proponent of a position related to formalism was Hilbert, whose important contributions to mathematics encompassed almost all its sub-branches: from axiomatic geometry, number theory and analysis, to applied mathematics and theoretical physics, and then later on to mathematical logic[16].

---

[13] The Collins Dictionary definition is this:

> **Formalism**: the philosophical doctrine that mathematical statements have no extrinsic meaning but that their symbols themselves, regarded as physical objects, exhibit a structure that has useful applications.
>
> (Collins Dictionary of Mathematics, p. 226)

Unfortunately, the second part of this definition is false. To say that a system S "exhibits a structure that ..." is to say that *there is a structure* Ω such that S exhibits Ω and ... This plainly *contradicts* formalism, which repudiates all abstract mathematical entities, including structures. Indeed, this latter anti-formalist clause is a statement of the rather attractive account of application advocated by *mathematical structuralism*, which is itself a version of platonism.

[14] But isn't a *game*, as opposed to its particular tokens, an abstract entity? Football, the "beautiful game", is an abstraction, unlike particular matches, which are events. And isn't an *expression*, say 'Tarski', as opposed to its concrete tokens (inscriptions and utterances), an abstract entity?

[15] Thomae 1898, *Elementare Theorie der analytische Functionen einer complexen Veränderlichen*, Introduction. Quoted in Frege 1903, *Grundgesetze*, Vol. ii, §§88-89. Requoted in translation by Kneale & Kneale 1962 (1988), p. 452.

[16] According to Abraham Pais, Hilbert attended a lecture in Göttingen in early 1915 given by Einstein, who outlined his latest thinking on General Relativity. Hilbert then wrote down the famous field equations for physical spacetime and they were published three weeks before Einstein's publication in July 1915. Indeed, perhaps the most profound laws in modern physics are sometimes called the "Einstein-Hilbert Equations", and the action functional from which they are derived using the usual Least Action Principle is commonly referred to as the "Hilbert Action".

Hilbert's formalism is somewhat difficult to interpret, but seems to be this: a tiny segment of mathematics *is* "contentful" and is, in fact, *true* in a certain non-mathematical interpretation. This part comprises those finitistic parts of arithmetic, such as quantifier-free arithmetical statements (i.e., *equations* $t_1 = t_2$, where t_1 and t_2 are terms which may contain free variables: e.g., $7 + 5 = 12$ or $x + y = y + x$). According to Hilbert, these are actually *truths* about certain *concrete entities*, which Hilbert took to be "physical strokes", like '| | |' and so on. The rest of mathematics—the so-called *ideal* part—is required merely to be a *consistent extension* of this contentful basis.

In particular, because mutually incompatible consistent extensions are possible, the notion of truth lapses and is replaced by *relativism*: a mathematical assertion φ may be **TRUE** relative to one extension M_1 (i.e., φ is provable in M_1), but **FALSE** relative to another M_2 (i.e., $\neg\varphi$ is provable in M_2).[17]

During the middle years of this century, formalism was an important and influential position in the philosophy of mathematics, advocated by many philosophers (especially those of strong empiricist, anti-metaphysical, inclination) and even some mathematicians, although formalism has consistently been unpopular with most working mathematicians who write about their subject[18].

Let us quote a more recent advocate of formalism, the mathematician and expositor of mathematics, E.T. Bell:

> Up until the early decades of the twentieth century it was quite commonly thought that mathematics has a peculiar kind of truth not shared by other human knowledge. For example, E. Everett (1794-1865) expressed the popular

[17] An example concerns the truth value of Cantor's Continuum Hypothesis in set theory, whether the cardinality of the continuum $\mathfrak{c}$ (i.e., $2^{\aleph_0} = |\,P(\omega)\,|$) is equal to the second transfinite cardinal $\aleph_1$. On this topic, Smullyan writes:

> There are those called formalists who regard the continuum hypothesis as neither true nor false, but entirely dependent upon which axiom system you take, since we can add either the continuum hypothesis or its negation to the axioms of set theory and have a consistent system in either case ... At the other extreme there are the so-called mathematical realists or Platonists—of which I am definitely one— who believe that the continuum hypothesis is either true or false, but we don't know which. We believe that we don't yet know enough about sets to answer the question, but this doesn't mean the question has no answer!
>
> (Smullyan 1993, pp. 248-249)

[18] Examples: Hersh & Davis, Ian Stewart, Roger Penrose, Keith Devlin.

> conception of mathematical truth as follows: "In pure mathematics we contemplate absolute truths, which existed in the Divine Mind before the morning stars sang together and which will continue to exist there, when the last of their radiant host shall have fallen from heaven" ...
>
> ... Although it would be easy to match this extravagance by many as wild from later writings of those who, like Everett, were not mathematicians by profession, it must be stated emphatically that only an inordinately stupid or conceited mathematician would now hold any such inflated estimate of his trade or of the "truths" he manufactures ...
>
> ... Against all this rhetoric that has been wafted like incense before the high altar of "Mathematical Truth", let us put the considered verdict of the last of the mathematical giants from the nineteenth century. Mathematics, according to D. Hilbert (1862-1943), is *nothing more than a game played according to certain simple rules with meaningless marks on paper* ... This is rather a comedown from the architecture of the universe, but it is the final dry flower of centuries of growth. The meaning of mathematics has nothing to do with the game, as such, and pure mathematicians pass outside their proper domain when they attempt to give their marks meanings.
>
> (Bell 1952, pp. 20-22. Emphasis added)[19].

Bell, following Hilbert, argues that mathematics is concerned solely with systems of meaningless postulates and with what meaningless theorems can be derived formally within such systems. He quotes approvingly some of the early "if-thenist" suggestions of Russell[20]. Bell remarks:

> A postulate is not necessarily 'self-evident', nor do we ask, "Is it true?". The postulate is *given: it is to be accepted without argument* ... Modern mathematics is concerned with playing the game according to the rules; others may inquire into the 'truth' of mathematical propositions, provided they think they know what they mean.
>
> (Bell 1952, p. 23. Emphasis added).

> Are the postulates then completely arbitrary? They are not, and the one stringent condition they must meet has wrecked more than one promising set and the whole edifice reared on it. *The postulates must never lead to an inconsistency* ...
>
> [And finally] one school of twentieth century mathematical philosophers discarded 'true' in favour of 'consistent'.
>
> (Bell 1952, pp. 27-28. Emphasis added).

---

[19] E.T. Bell, Professor of Mathematics at California Institute of Technology, published *The Queen of the Sciences* in 1931 and the *Handmaiden of the Sciences* in 1937. The book cited, Bell 1952, *Mathematics: Queen and Servant of Science*, is "a thorough revision and a very considerable amplification" of these two earlier volumes.

[20] See later, Chapter 6, for a refutation of formalistic if-thenism.

Actually, there was a *serious attempt* to construct a genuinely nominalistic account of the formal manipulation of *concrete* signs. Such a theory might be called "nominalistic syntax". The programme was outlined in Quine and Goodman 1947, 'Steps Toward a Constructive Nominalism'. We shall see below why this serious attempt to put the rather vague formalist ideas about meaningless marks on a proper footing was a failure. In order to get the system going, an infinity of abstract expression types is needed.

Were these recent realists, especially Frege and Gödel "inordinately stupid or conceited"? After all, it was Gödel whose Incompleteness Theorems demolished the central plank of Hilbert's formalist programme, the aim of finding *finitary* consistency proofs of infinitary mathematics. But Frege, Cantor, Gödel and Church were all realists, who thought that it makes perfect sense to inquire of a mathematical proposition whether it is *true* or not.

The formalist has no workable account of the *application of mathematics*. There is a common (but trivially fallacious) the view that mathematics is applied by taking mathematical symbols somehow to *refer* to *concrete physical entities*. This is trivially false. The symbol '$A_\mu(x)$' in electromagnetic theory refers to a *function*, not to some bizarre ghostlike physical presence. The function assigns to each physical space-time point a quadruple of real numbers. This function partially characterizes the (abstract) *state* of space-time at each point. The *state* of space-time at a point e is a *mathematical entity* which mathematically determines (via a differential equation) the behaviour (i.e., the geodesics) of electrons in its vicinity. To *deny* this is simply not to grasp the physics. So, the applicability of mathematics does not involve some mysterious and impossible *reinterpretation* of mathematical symbols. (A further simple reason is that there are *more* mathematicalia, including power sets etc., than physicalia). A more serious account must incorporate the idea that the application of mathematics in real mathematicized science involves the use of mixed predicates (and functors, like 'A_μ', '$\boldsymbol{R}_{\mu\nu}$' and so on), which express the appropriate *structure-preserving relations* between mathematicalia and concreta.[21]

So it is hard to see how formalism can even be consistent with the possibility of applying mathematics. The central tenet of (especially 19th century) formalism is that mathematics is the formal manipulation of *meaningless* combinations of signs (the formalist concedes that the logical particles are interpreted, otherwise the logical inter-relations of consistency and implication would not make sense).

It is difficult to know what to make of such an extreme position. In fact, formalism was in deep trouble in the 1890s as a result of insightful criticisms of Frege (again concerning application). Frege wrote in the *Grundgesetze*,

> Why can no application be made of a configuration of chess pieces? Obviously because it expresses no thought. If it did so and every chess move conforming to the rules corresponded to a transition from one thought to another, applications of chess would also be conceivable. Why can arithmetical equations be applied? Only because they express thoughts. How could we possibly apply an equation which expressed nothing and was nothing more than a group of figures, to be transformed into another group of figures in accordance with certain rules? Now *it is applicability alone which elevates arithmetic from a game to the rank of a science.*
>
> (Frege 1903, Vol ii. §91. Emphasis added)[22].

Frege's question 'How could we possibly apply an equation which expressed nothing?' was left unanswered (it is unanswerable). There can be no application of groups of figures which do not express thoughts, and are therefore incapable of being either true or false.

The message of formalism has to be that the application of mathematics is *impossible*: we simply cannot *use* (as opposed to *mention*) mathematical statements in our theorizing about the world. For to use a statement is to have a certain attitude to its content (usually

[21] To take an example, the standard differential equations of, say, accepted classical electromagnetic theory contain a quadruple of four-place functors (viz., '$A_\mu(x^0, x^1, x^2, x^3)$') and another quadruple of one-place functors ('$\varphi^\mu(e)$'). To be sure, the variable '*e*' ranges over spacetime points, but it is straightforwardly false to say that, given values of variables (that is, spacetime points as values of '*e*'), that either the terms '$\varphi^\mu(e)$' or the terms '$A^\mu(\varphi^0(e), \varphi^1(e), \varphi^2(e), \varphi^3(e))$' refer to something *concrete* in the world: the theory specifies that these terms refer to *real or complex numbers*. Equivalently, the theory says that the φ^μ *are* real numbers (which, e.g., can be algebraically manipulated).

[22] Translation taken from Geach & Black 1952 (1980), p. 167.

an assertive attitude). According to formalism, mathematical "forms of words" are not capable of being used assertively and therefore cannot be asserted (or used in any sense) in our empirical theorizing. But *this flatly contradicts the assertion of mathematical statements in the empirical application of mathematics*. An empirical scientist who freely considers (impure) functions from space-time to the real numbers (e.g., field functions like mass density or electromagnetic field) or functions from physical bodies to the real numbers (e.g., mass, length, temperature) does so by *using* and indeed, *asserting* certain interpreted sentences which refer to mathematicalia. Consequently, one surely cannot coherently maintain that the sentences containing terms that refer to these postulated functions are contentless.

Perhaps we need to be more explicit regarding the "use" of mathematics in science. How is mathematics used? Well, scientists sometimes *assert* mathematicized statements and laws purporting to describe Nature. Scientists sometimes *test* these laws. Scientists gather *evidence* for and against these laws. Scientists debate the *truth* and *approximate truth* of mathematicized laws of Nature. How could this even be consistent with the formalist claim that these assertions are not even meaningful?

As Maddy put it,

> The Platonist Gottlob Frege launched a fierce assault on early formalism, from many directions simultaneously, but the most penetrating arose from just this point. It isn't hard to see how various true statements of mathematics can help me determine how many bricks it will take to cover the back patio, but how can a meaningless string of symbols be any more relevant to the solution of real world problems than an arrangement of chess pieces?
>
> This is Frege's problem: what makes these meaningless strings of symbols useful in applications? Suppose, for example, that a physicist tests a hypothesis by using mathematics to derive an observational prediction. If the mathematical premiss involved is just a meaningless string of symbols, what reason is there to take that observation to be a consequence of the hypothesis? And if it is not a consequence, it can hardly provide a fair test. In other words, *if mathematics isn't true, we need an explanation of why it is all right to treat it as true when we use it in applications.*
>
> (Maddy 1990, p. 24. Emphasis added).

To sum up, it is incoherent to insist that mathematical statements are contentless groups of figures and yet at the same time to *use* them in applications. Formalism implies that the application of mathematics is impossible. But the application of mathematics is possible. Therefore formalism is false.

1.5 Illustration I. Measurement & Quantities

In science, we explain the observable behaviour of physical systems by referring to measurable properties of those systems. Such properties are sometimes called quantities or magnitudes. More exactly, a quantity is a function which systematically assigns certain mathematicalia (numbers, or sets of numbers, or vectors or whatever) to concrete systems.

For example, the sentence,

(1) the length-in-m of the Eiffel Tower > 100

expresses the fact that the length_m function assigns a *number* greater than 100 to a certain concretum, the Eiffel Tower.

The technical theory which explores the conceptual foundations of the mechanisms involved in such assignments is known as *Fundamental Measurement*[23]. Of prime importance in Fundamental Measurement Theory is the discovery (and proof) of interesting Representation and Uniqueness Theorems. A Representation Theorem explains how a structure Ω built of concreta (a "concrete structure", as it were) can be represented by an "abstract structure" Ω', built of abstracta, (e.g., the real numbers, under the usual ordering, with the usual algebraic operations). The notion of representation involved is that of a *homomorphism*: a structure-preserving mapping ρ from the "concrete structure" Ω into the abstract representing structure Ω'. The resulting homomorphism is called a *measurement scale*. The length_m function is just such a measurement scale.

The first explicitly formulated Representation Theorem was discovered and proved by Hölder. It shows how any ordering $\Omega = (D, \alpha)$ which satisfies certain constraints can be represented by the standard linear ordering of reals $\Omega' = (\mathbf{R}, <)$ by an injective map ρ: D

[23] The classic exposition is Krantz et al. 1971.

$\mapsto$ **R** such that, for any elements x, y of D, $(x, y) \in \alpha$ if and only if $\rho(x) < \rho(y)$. Here D may be an *impure* set of concreta and α may some ordering relation amongst these concreta (like 'is less massive than'). Then ρ is the *representing homomorphism* (measurement scale) which assigns real number values to the relata of α in a structure-preserving manner.

It almost goes without saying that for any given concretum x, the value $\rho(x)$ is an abstract entity, viz., a real number. Hence quantitative theorems using measurement scales are committed (through quantification, ultimately) to mathematicalia. Associated with Hölder's Theorem is a Uniqueness Theorem which says that if ρ_1 and ρ_2 are both such scales, then there is a real number a such that $\rho_2 = a\rho_1$. Now suppose we fix upon some particular concretum, x_0 say. Then a Uniqueness Theorem may allow us to prove that $\exists!\rho(\rho(x_0) = 1$. That is, the Uniqueness Theorem, plus the axiom '$\rho(x_0) = 1$', entails the uniqueness of such a ρ. We then say that x_0 is a *standard unit* for the scale ρ. Of course, there may be many such standard concrete units, which all occupy the same location in the ordering.

For example, consider the scale mass_{kg} (the mass-in-kilograms scale), ubiquitous in physical science. This measurement scale is a homomorphism with respect to the 'less massive than' relation. That is, for any concreta x and y, we have the theorem,

$$\mathit{Less\text{-}Mass}(x, y) \leftrightarrow (\mathit{mass}_{\text{kg}}(x) < \mathit{mass}_{\text{kg}}(y))$$ [24]

Measurement scales for other physical magnitudes, like temperature, pressure, force, temporal duration, and so on, are similarly introduced. For more details see Krantz et al. 1971. All of thcsc arc rcprcscntcd by mcasurcmcnt scales, which are structure-preserving homomorphisms from relations amongst concreta to some (usually the ordering) relation on real numbers.

---

[24] Notation: throughout I use the Arial font for formulas and symbols. Predicates, like *Less-Mass*, begin with a capital, while functors, like $\mathit{mass}_{\text{kg}}$, begin with a lower-case letter. I use such capital Arial symbols as 'T', 'M', 'P' and 'N' to refer to *theories* in any formalized language and the symbol 'L' to refer to such langauges. I sometimes use Fractur '$\mathfrak{I}$' for interpretations of a formalized language, and the symbol 'Ω' to range over structures or models, considered as mathematical entities rather than as interpretations. 'φ' almost always ranges over L-formulas. As a victim of Quine, I try not to confuse use and mention, but, like most mathematical authors, use many symbols *autonymously*, so $\leftrightarrow$ and $\exists$ are symbols.

Notice that it is the underlying Representation Theorem which says when a mathematical operation on values of some scale is "meaningful". For example, if we consider the mass scale, $mass_{kg}$, then *addition* of values of the $mass_{kg}$ function is meaningful, for we suppose that the usual concatenation (or aggregation) operation o on concreta is such that:

$$mass_{kg}(x) + mass_{kg}(y) = mass_{kg}(x \circ y)$$

On the other hand, the multiplication of masses is not meaningful: there is no (simple) physical operation ⊗ on concreta such that:

$$mass_{kg}(x) \times mass_{kg}(y) = mass_{kg}(x \otimes y)$$

Thus, we have an explanation of what a scientist means when he or she says that certain mathematical operations have no "physical interpretation". Adding energies or masses is meaningful, adding temperatures is not. Note: it doesn't follow from this that a "meaningless" term like '$T_1 + T_2$' cannot occur in a scientific law. The criterion of meaningfulness certainly does not rule out a theory in which "meaningless expressions" occur. E.g., in Newton's Law of Gravitation, the physically meaningless product $mass_{kg}(x) \times mass_{kg}(y)$ occurs, as well as the meaningless square $[dist_m(x, y)]^2$.

1.6 Illustration II. Geometry and Spacetime Theory

Modern physical geometry has moved far beyond (Cartesian) analytic Euclidean geometry. The problems of space, time and motion have always been at the very centre of philosophical and scientific discussion. A revolutionary new conception emerged between 1905 and 1915, in Einstein's work[25]. Modern Einsteinian relativistic space-time theory, *General Relativity* (henceforth, GR), provides a unified account of matter, energy, motion, space and time[26]. The traditional philosophical problems of time, change

[25] For a collection of Einstein's major papers and others see Einstein et al. 1952 (Sommerfeld (ed.)).

[26] The standard undergraduate text is Rindler 1979. The classic graduate texts are Weinberg's 1972 and Misner, Thorne & Wheeler 1975. The standard graduate texts are Wald 1984 and the monumental Choquet-Bruhat et al. 1982.

and trans-temporal "identity", substantivalism-versus-relationism, conventionality and simultaneity, and so on, must all be understood in terms of Einstein's work[27].

In Cartan's co-ordinate-free formulation, relativistic space-time theory contains theoretical principles like:

> If U is an open set of points in a region of space-time, then there exists a function φ from U to an open set in $\mathbf{R}^4$ such that (U, φ) is a co-ordinate chart.[28]

This statement is, of course, *contingent.* It says that space-time is locally diffeomorphic to $\mathbf{R}^4$. This might be false if superstring theory (or supergravity theory) is correct, according to which space-time is locally diffeomorphic to $\mathbf{R}^{10}$. In any case, from the definition of a co-ordinate chart, one can immediately *deduce* (given the existence of regions of space-time containing points) the existence of real numbers, for a quadruple of these is assigned by any co-ordinate chart to each space-time point. Thus, from the basic principles of space-time theory, we may infer:

i. the existence of mathematicalia: *sets* of space-time points, *quadruples* of *reals*, etc.

Of course, Einstein's theory is not *simply* an assertion that space-time is related to mathematicalia in a certain way. The theory is profoundly elegant, predictive and explanatory. Almost all workers in the field (called "relativists") think it is *true* (or so close to being true that any successor theory would have to preserve almost all of its content). Einstein's Field Equations for GR connect the space-time curvature (Ricci) tensor $\boldsymbol{R}_{ab}$ and metric $\boldsymbol{g}_{ab}$ with the energy-momentum tensor $\boldsymbol{T}_{ab}$. They are expressed as the tensor identity:

(*) For any space-time point e, $\boldsymbol{R}_{ab}(e) - \frac{1}{2}\boldsymbol{R}(e)\boldsymbol{g}_{ab}(e) = 8\pi G\, \boldsymbol{T}_{ab}(e)$

($\boldsymbol{R}$ is the Ricci scalar, $= \boldsymbol{g}^{ab}\boldsymbol{R}_{ab}$; G is Newton's constant)

[27] A major modern treatment of several philosophical problems connected to modern relativistic space-time theories is Friedman 1983. An even more advanced treatment is Earman 1995. A discussion of Zeno's Paradoxes and "infinite supertasks" in space-time is contained in Earman & Norton 1996.

[28] I.e., if e is a point in U, then the real numbers $\varphi^a(e)$ are the co-ordinates of e in (U, φ).

We can restate this universally quantified law more schematically as follows:

For each space-time point *e*, CURVATURE at *e* = ENERGY-MOMENTUM at *e*

On the question of a successor theory, considerable interest was created in 1984 when it was shown first that *closed* superstrings behave just like gravitons (the spin 2 quanta of the gravitational field) and second that superstrings can only "propagate consistently" (that is, with the proper cancellation of certain infinities called "anomalies") in a space-time that *satisfies Einstein's field equations*. The methodological principle is clear: any theory (supergravity, superstrings, Kaluza-Klein, etc.) that succeeds GR will have to explain it or reduce to it in some low-energy limit.

Of course, GR is an empirically testable theory. Given its central law above we can infer a spectacular range of empirical predictions. Some important consequences of Einstein's space-time theory (with auxiliaries, of course) are:

ii. the existence of space-time geodesics and null-geodesics;

iii. the time-like geodesic motion of small test bodies in curved space-time;

iv. the null-geodesic paths of light rays in curved space-time;

v. the precession of bound orbits;

vi. the gravitational red-shift of light;

vii. gravitational time dilation of clocks (anything that oscillates),

and so on.

Furthermore, one can show that Newton's NM "lives on a limiting case" of GR. This proceeds via the linear approximation, where we set $g_{\mu\nu} = \eta_{\mu\nu} + \gamma_{\mu\nu}$, where $\eta_{\mu\nu}$ is the flat Minkowski metric for flat space-time, and $\gamma_{\mu\nu}$ is a "small perturbation":

viii. the Einstein field equations (*) reduce (approximately) to:

$$\nabla^2\gamma_{00} \propto T_{00}$$

And this is Poisson's field-theoretic formulation of the law of Universal Gravitation, $\nabla^2\Phi \propto \rho$. The reduction shows that the time-time component of the metric γ_{00}

approximates the Newtonian "gravitational field", Φ. Einstein himself took this as a *central constraint* in his attempt (with Grossman) to find the correct field equations.

Indeed, within the linear approximation, it can be shown that:

ix: $\Box\gamma_{\mu\nu} = 0$

which is of course a *wave-equation*. So, GR predicts the existence of *gravitational waves* (wavelike fluctuations in the space-time metric $g_{\mu\nu}$). Indeed, many of the problems in quantizing GR (e.g., non-renormalizability) arise from trying to quantize this equation (ix), or rather, the perturbative series equation, in derivatives of $\gamma_{\mu\nu}$, of which (ix) is a limiting case.

In any case, irrespective of the more physical and testable consequences (ii)-(ix), the *Platonic* deductive consequence (i) reveals that modern space-time theory is "up to its neck" in explicit ontological commitment to mathematicalia. In ontological terms, GR quantifies over co-ordinate charts, real numbers, sets of points, tensor functions, and so on. Of course, GR is a *contingent theory* about the nature of physical space-time, which has been empirically corroborated, provides many unifying explanations, and which is looked upon (by experts) as one of the most profound inventions in the history of human ideas. But plainly its contingency does not prevent its quantification over, and thus its commitment to, an ontology of mathematicalia.

1.7 Illustration III. Syntax

Syntax treats the properties of linguistic expressions and their combinations. Syntactical theory quantifies over both concrete physical inscriptions (or *tokens*) and abstract symbols and expressions (or *types*). Indeed, tokens play an almost insignificant role in the theory "proper", and are only explicitly considered when *philosophers* ask what syntactic theory is really about.

The relation of "tokening" is somewhat similar to Platonic instantiation: 'inscription *x* is a token of symbol *y*' expresses a *relation* of concreta to abstracta. This might be clarified by introducing a formalization of syntax. Governing the 2-place predicate *Token* ('is a token of'), the monadic predicates *Inscr* ('is an inscription') and *Expr* ('is an expression'), we have axioms such as the following:

(1) $Inscr(x) \rightarrow \exists X(Expr(X) \wedge Token(x, X))$

which says that each inscription is a token of some expression.

Similarly, syntactical theory introduces an operation of (abstract) concatenation, often signified by ^, governed by the following closure axiom:

(2) $Expr(X) \wedge Expr(Y) \rightarrow Expr(X \mathbin{^\wedge} Y)$

The crucial point to recognize is that standard syntactical theory is developed Platonistically, for the following reasons. The operation (relation) of *concatenation*, $Concat(X, Y, Z)$:

Z is a concatenation of X and Y,

applies to expressions (that is, *types*) and is, roughly, the "abstract correlate" of a certain physical relation: *juxtaposition*, $Juxt(x, y, z)$:

z is a juxtaposition of x and y,

which applies to *tokens*.

Of central importance for Platonistic syntax is the existence/uniqueness theorem for concatenation:

(3) $Expr(X) \wedge Expr(Y) \rightarrow \exists! Z(Expr(Z) \wedge Concat(X, Y, Z))$

(For any pair of types, *there exists* a unique concatenation).

Then we can easily *define* the concatenation functor ^ thus:

$X \mathbin{^\wedge} Y =_{df} (\iota Z)Concat(X, Y, Z)$

The *application* of Platonistic syntax then rests on the "Representation Theorem":

(4) $Token(x, X) \wedge Token(y, Y) \rightarrow \forall z(Juxt(x, y, z) \rightarrow Token(z, X \mathbin{^\wedge} Y))$.

That is, if x and y are tokens of types X and Y, then any juxtaposition of x and y is a token of the concatenation of X and Y. Abstract concatenation of abstract types X and Y "represents" concrete juxtaposition of their respective tokens.

One makes good use of Platonistic syntax in standard proof theory and metalogic. For example, Gödel's Completeness Theorem for First-Order Logic says that that if a formula φ is a logical (semantic) consequence of some set Δ of formulas, then *there exists* a (finite) derivation Γ of φ from Δ. As Church argued [see below], this metatheorem is only true when a derivation Γ is construed as *abstractum*, i.e., as an arbitrarily long sequence of (abstract) expression concatenations. It is not true when derivations are reconstrued as concreta, i.e., as concrete *tokens*. The reason is simple. Only a small finite number of abstract proofs have ever been inscribed (as tokens). When we say that φ is derivable from Δ, we means that there exists a (Platonic) *sequence of concatenations* Γ of abstract symbols which constitutes a proof of φ from elements of Δ. We do not mean that someone has actually written down, uttered or even thought of that sequence Γ.

In short, any standard proof of the Completeness Theorem is *Platonistic*. It appeals (albeit implicitly) to principles from Platonistic syntax.

Fifty years ago, Quine and Goodman published a famous paper[29], in which they outlined certain philosophical principles and attempted to develop rigorously the theory of *nominalistic syntax*. Their paper begins with the announcement:

> We do not believe in abstract entities. Noone supposes that abstract entities—classes, relations, properties, etc.—exist in space-time; but we mean more than this. We renounce them altogether.
>
> (Quine & Goodman 1947 (in Goodman 1972), p. 173).

Roughly, Quine and Goodman assumed the existence of concrete inscriptions, invoked the notion of concrete juxtaposition, and attempted to derive proof theory. Church's reply to this paper appeared in a letter to Goodman:

> ... additional tasks that 'nominalistic' (better, finitistic) syntax might be asked to accomplish [:] the deduction theorem; ... the proof of the rules of substitution as derived rules ... Post's completeness theorem for the propositional calculus; the metatheorem that every quantifier-free theorem of first-order functional calculus

[29] Quine & Goodman 1947.

has a quantifier-free proof; the principal results about the Skolem normal form; Gödel's incompleteness theorems ...; Gödel's relative consistency proof for the axiom of choice and the generalized continuum hypothesis; the results of Fraenkel, Lindenbaum, and Mostowski concerning the independence of the axiom of choice.

(Church 1958b).[30]

Considerations such as these were instrumental in motivating Quine's early defection from nominalism[31]. Quine summed up in 1981:

> How far could one push elementary mathematics without thus reifying universals? Goodman and I explored this at one point. The formalist, we remarked, was already involved in universals in treating of expression types ... a formalism of tokens afforded considerable mileage but stopped short of full proof theory.
>
> (Quine 1981c (1981a), pp. 182-183).

One central problem for nominalistic syntax is simply that not enough (actual) tokens exist. This is why Church emphasized finiteness: standard proof theory is non-finitistic. Briefly, the following contingent assertion about tokens is empirically false:

(5) $Inscr(x) \wedge Inscr(y) \rightarrow \exists z(Inscr(z) \wedge Juxt(x, y, z))$

And because of this, in nominalistic syntax we *cannot define* a *nominalistic juxtaposition functor* 'o' with the right closure ("extendability") properties. If we had the above theorem (5), it would be easy:

$$x \circ y =_{df} (\iota z)(\exists Z Token(z, Z) \wedge Juxt(x, y, z)).$$

The basic advantage of Platonistic syntax, then, is its non-finitistic nature: abstract concatenations of distinct expression-types always exist and yield *new* expression-types. Given an initial symbol |, we get | |, and | | |, and so on: infinitely many (indeed, a model of arithmetic). If we formulate our syntactical theorems in terms of abstracta (i.e., types), then we can (implicitly) appeal to the properties of abstract concatenation to prove the

---

30 In a letter to Goodman on 1 Dec 1958. Quoted from Goodman 1972, pp. 153-154.

31 In fact, Quine had retracted this nominalistic credo in 1953, and probably had already defected by 1948. In the bibliography of his 1953 anthology, he added the comment:

> I should now prefer to treat that sentence as a hypothetical statement of conditions for the construction in hand.
>
> (Quine 1953a (1980), pp. 173-174)

That is, Quine conceded that nominalistic syntax is inadequate as a reconstruction of standard proof theory, with its quantification over abstract expression types.

standard proof-theoretical and metalogical theorems (that is, we automatically have the *existence* of an expression $X \wedge Y$, for arbitrary expressions X, Y). In nominalistic syntax, the required properties do not hold of concrete tokens. Given concrete inscriptions x and y, there might be no such thing as "$x \circ y$".

A "solution" is to introduce *modalist nominalistic syntax* with a modal existence theorem:

> for any inscriptions x, y, there *could have been* an inscription z such that $juxt(x, y, z)$

Or, formally, instead of (5), we have:

(6) $Inscr(x) \wedge Inscr(y) \rightarrow \Diamond\exists z(Inscr(z) \wedge Juxt(x, y, z))$

However, a nominalist would surely view the introduction of modality as a greater evil than admitting abstract objects. As Quine later put it,

> ... long ago, Goodman and I got what we could in the way of mathematics ... on the basis of a nominalist ontology and without assuming an infinite universe. We could not get enough to satisfy us. But we would not for a moment have considered enlisting the aid of the modalities. *The cure would in our view have been far worse than the disease.*
>
> (Quine 1986, p. 397. Emphasis added).

1.8 Illustration IV. Metalogic

Standard formulations of the formal science of mathematical logic[32] (sometimes called metalogic) invoke a rich variety of mathematicalia in both the syntactic (proof-theoretic) and semantic (model-theoretic) components: symbol types, arbitrarily long abstract concatenations and sequences thereof, sets and sequences of objects, functions, relations and structures.

[32] E.g., Robbin 1969, Mendelson 1987, Boolos & Jeffrey 1989, Machover 1996. See Hodges 1997 for a graduate level presentation of *model theory* and Takeuti 1975 for a standard presentation of *proof theory*. Boolos 1993 relates proof theory to systems of modal logic.

Machover 1996 puts this by saying that metalogic operates within an "ambient set-theory":

> As any other mathematical theory, our metatheory just starts from a launching pad of presuppositions: some underlying concepts, regarded as known, in terms of which further concepts of the theory are defined; and certain fundamental propositions, on the basis of which the theorems of our theory can be rigorously proved.
>
> Set theory—in the form of ZF or some similar codification—is certainly strong enough to underpin our study of logic. Indeed, the entire development ... can be read as occurring in set theory.
>
> (Machover 1996, p. 135]
>
> Those parts of our investigation that depend upon the concepts defined in Def 4.10 [viz., Tarskian notions of logical consequence and satisfiability] will generally presuppose the existence of infinite sets as objects, and must be viewed as taking place in an ambient theory that incorporates a sufficiently rich set theory.
>
> (Machover 1996, p. 157)

For a simple example, in standard metalogic the metatheorem:

(1) the schema $\exists x(F(x) \wedge G(x))$ is satisfiable,

is actually an abbreviation of an *existential* claim about structures:

(2) *there exists* a structure Ω such that $\Omega \models \exists x(F(x) \wedge G(x))$

Indeed, one can find simple formulas (Quine 1950 (1972) calls them "infinity schemas") satisfied *only* by *infinite* structures[33].

Likewise, a semantic theorem about logical consequence such as:

(3) $\forall x(P(x) \rightarrow \exists y Q(y)) \models \exists x P(x) \rightarrow \exists x Q(x)$

is an abbreviation of a universal claim about structures:

---

[33] An example is:

$$\forall x \exists y \forall z(F(x, y) \wedge \neg F(x, x) \wedge (F(x, y) \rightarrow F(x, z))).$$

See Quine 1952 (1974), p. 183. A simpler example (using just a constant and a function symbol) is the usual axiom for the successor function:

$$\forall x \forall y(s(x) \neq 0 \wedge (x \neq y \rightarrow s(x) \neq s(y))).$$

The reason is simple. If $f: X \rightarrow X - \{a\}$ is injective, then X must be infinite! (One proves by induction: if X has n elements, $a \in X$ and $f: X \rightarrow X - \{a\}$, then f is not injective).

(4) for any structure Ω, if $\Omega \models \forall x(P(x) \rightarrow \exists yQ(y))$, then

$\Omega \models \exists xP(x) \rightarrow \exists xQ(x)$

Now, structures are surely mathematicalia (set-theoretical entities), so the simple metalogical theorem that a certain schema is satisfiable implies the existence of mathematicalia.

Similarly, Gödel's Completeness Metatheorem for first-order logic says

(5) if $\Delta \models \varphi$ then $\Delta \vdash \varphi$

This refers to an arbitrary *set* Δ of premises (which may be infinite, e.g., the axioms of first order Peano Arithmetic, PA). But furthermore, the notation '$\vdash$' can be expanded in the standard way:

(6) if $\Delta \models \varphi$ then *there exists a derivation* Γ of φ from Δ

A derivation Γ in this sense must be an abstract entity: an arbitrarily long (but finite) sequence of sentence types such that every element of Γ is either an element of Δ or is a statement immediately derivable using a rule of inference from some earlier subset of elements in Γ. If one attempts to reconstrue (6) as a statement about *concrete tokens* then it is false: but only a nominalistic *fanatic* would try to argue that the first-order logic is incomplete because not enough tokens exist!

CHAPTER 2

The Structure of Mathematicized Theories

> Standard scientific theory supplies much of the information it supplies about physical entities only indirectly, by way of apparatus pertaining to supposed relationships of physical entities to supposed mathematical entities and supposed classifications of and relationships among the supposed mathematical entities themselves. As much of what science says about observable entities is 'theory-laden', so much of what science says about concrete entities (observable or theoretical) is 'abstraction-laden'.
>
> John P. Burgess and Gideon Rosen 1997, *A Subject With No Object*, p. 84.

2.1 The Logical Structure of Mathematicized Theories

Formally, it is convenient to represent a mathematicized theory of Nature in a manner that distinguishes between the concrete and mathematical. The benefits of doing so are three: logical, ontological and epistemological. Such a representation is achieved using a "standard two-sorted mathematicized language" L.

Let us begin with a rather loose definition:

Definition 1: Mathematicized Language

A language L is *mathematicized* if it contains the notation of mathematics (e.g., predicates from arithmetic, analysis, set theory, etc.).

Mathematicized theories are formulatable in some mathematicized language. Obviously, a sceptic about mathematicalia, a nominalist, can legitimately use a mathematicized language. Indeed, he or she must use a mathematicized language in order to express disbelief in mathematicalia, for he or she needs to assert statements like 'real numbers do not exist' (or, more generally, 'abstract entities do not exist').[34]

[34] This is why the view advocated by Carnap 1950, the alleged distinction between "external" and "internal" statements, is incorrect. Carnap held that, given the *language* of mathematics, the assertion 'numbers exist' is *analytic* and thus must be *accepted* by anyone who *uses* the language. But this *must* be wrong. For Hartry Field uses the language of mathematics and asserts 'numbers do not exist'. If Carnap were right, Field would

Next, we adopt the following definition:

Definition 2: Standard Mathematicized Language[35]

A standard mathematicized language L is a *two-sorted language* which contains:

i. *primary* variables: $x_1, x_2, \ldots$, ranging over concreta.

ii. *secondary* variables: $X_1, X_2, \ldots$, ranging over mathematicalia.

If L is a standard two-sorted mathematicized language for formalizing mathematicized scientific theories, its predicates can obviously be classified into three kinds, according to how they are correctly "saturated" by variables.

Primary predicates: predicates which have only argument places for primary variables: such predicates intuitively express relations between concreta [e.g., *hotter*, *taller*, and so on]. *Secondary predicates*: predicates which have only argument places for secondary variables: such predicates intuitively express relations between mathematicalia [e.g., *Equinumerous*, *Isomorphic*, *Larger*, and so on]. *Mixed predicates*: (2-or-more-place) predicates which have argument places for both primary and secondary predicates: such predicates intuitively express relations between concreta and mathematicalia [e.g., $\in$, $Mass_{kg}$, and so on].

Thus, corresponding to these kinds of predicate, we obtain atomic formulas of three kinds.

i. Primary formulas,

e.g., $Taller(x_1, x_2)$, $Hotter(x_1, x_2)$, etc.,

ii. Secondary formulas

e.g., $Equinumerous(X_1, X_2)$, $Isomorphic(X_1, X_2)$, etc.,

---

be contradicting himself! It follows that Carnap is wrong. If a nominalist asserts 'numbers do not exist' clearly it is incorrect to say that he or she is not *using* mathematical language. *How else* could the nominalist express his or her *disbelief* in mathematricalia. By ESP?

For further criticisms of the inadequacy of Carnap's view, see Quine 1951b and Quine 1954.

[35] In adopting this use of a two-sorted notation, I follow the conventions of Burgess & Rosen 1997, p. 69.

iii. Mixed formulas

e.g., $x \in X$, $Mass_{kg}(X, x)$, etc.

Any statement or axiom in L containing only primary (secondary) predicates will be called a primary (secondary) statement or axiom. A statement or axiom containing a mixed predicate will be called a mixed statement.

As an example, we may rewrite the axiom expressing the density of the reals:

(1) there is always a real number between any two given distinct reals

as the *secondary* axiom:

(2) $\forall X_1 \forall X_2 (X_1 < X_2 \rightarrow \exists X_3 (X_1 < X_3 < X_2))$

And we rewrite the axiom from measurement theory:

(3) every physical object has a mass-in-kg

as the *mixed* axiom:

(4) $\forall x \exists X Mass_{kg}(X, x)$

We will say that a mathematicized scientific theory is *standardly formulated* if it is formulated in some standard two-sorted mathematicized language L. (Note that if we adopt this two-sorted language form, then a formula like $Taller(X_1, X_2)$ is not well-formed. For *Taller* is a primary predicate and must be completed with primary variables, thus: $Taller(x_1, x_2)$).

In general, any mathematicized theory of Nature formulated in a standard two-sorted language L has, correspondingly, three kinds of axiom:

Axiom Type 1: Secondary Axioms

Such axioms always occur and their purpose is, loosely speaking, to specify the *abstract mathematical structure* Ω that the theory T is going to "apply" to the concrete physical world. For example, the structure may be the continuum (for space or time), or a symplectic manifold (for Hamiltonian mechanics) or a Hilbert Space (for quantum mechanics).

Axiom Type 2: Mixed (Representation) Axioms

Loosely speaking, again, such axioms specify how the structure characterized by the pure axioms is "exemplified in" concrete reality. *Almost all the laws of physics (differential equations) are in fact such mixed axioms.* Imagine a law which asserts the existence of a discrete linearly ordered sequence of temporal instants with an endpoint, satisfying the induction principle. A *concrete* ω-sequence, if you like. This law is a mixed law. In effect, it says that there is a concrete system which exemplifies the natural number structure (**N**, <). (Clearly, it is contingent that such a system exemplifies (**N**, <))

Axiom Type 3: Primary Axioms

Such axioms, if they occur at all, specify mundane facts (perhaps observable) about physical relations between concreta.

Below I shall provide two examples of mathematicized theories which are standardly formulated.

2.2 The Application of Arithmetic

Three of the basic (axiomatic) mathematical theories we are familiar with, Peano Arithmetic PA, Real Analysis RA and standard Zermelo-Fraenkel Set Theory ZFC, are not, as they stand, *applicable* mathematical theories. For example, the language in which PA is standardly formulated is one-sorted, with only first-order variables ranging over numbers. There is no place for quantification over non-mathematicalia or concreta, and no place for non-arithmetical predicates (that is, primary or mixed predicates).

Arithmetic made be made applicable by adding a *predicate-operator* #:

$\#F$

meaning, intuitively,

"the number of Fs".

The result of adding # and a stock of primary predicates (say, 'apostle', 'planet', 'teacup', 'moon', 'space-time point', etc.) provides a notation within which we may formulate assertions like,

(1) #(apostles) > #(planets)

(2) #(planets) = 9

(3) #(planets) + #(apostles) > #(moons of Earth)

Although these are mathematical assertions, they are not *pure* mathematical assertions. Indeed, each of them, if true, is only contingently true. It is only contingently true that the number of planets is equal to 9. More importantly, (1)-(3) are not *theorems* of applicable arithmetic.

So, what is applicable arithmetic? There is sense in which applicable arithmetic is a remarkably simple theory, reducible to a single axiom, known as "Hume's Principle":

HP: $\#F = \#G \leftrightarrow$ (there is a bijection from Fs to Gs)

In order to formulate this properly one needs either set theory or second-order logic. The usual second-order formulation is this. The clause "there is a bijection from Fs to Gs" means that there is a function f such that:

i.	$\forall x(F(x) \rightarrow G(f(x)))$	the f-image of any F is a G
ii.	$\forall x \forall y(x \neq y \rightarrow f(x) \neq f(y))$	f is an injection
iii.	$\forall y(G(y) \rightarrow \exists x(F(x) \wedge y = f(x)))$	f is a surjection from Fs to Gs

In short, the clause is to be rewritten:

$$\exists f(\forall x \forall y(F(x) \rightarrow G(f(x)) \wedge (x \neq y \rightarrow f(x) \neq f(y)) \wedge \forall y(G(y) \rightarrow \exists x(F(x) \wedge y = f(x)))$$

Abbreviate this second-order formula as

$$F \approx G$$

Then the second-order formulation of Hume's Principle is:

HP^2: $\#F = \#G \leftrightarrow F \approx G$

An interesting metatheorem (Boolos 1987, Wright 1983) is that, by introducing definitions for 0 and s it can be shown that $HP^2 \vdash PA^2$. That is, HP^2 is arguably a kind of logicist representation of arithmetic[36]. (Indeed, the second-order theory whose sole non-logical axiom is HP^2 is sometimes called "Frege Arithmetic", FA)

2.3 The Application of Analysis

The theory of real numbers (Real Analysis, RA) is the theory of what algebraicists call a *field*, a system of objects that can be added, subtracted, multiplied, and for which division is generally defined (except by 0)[37]. The field axioms can be formulated in a one-sorted language L_{RA} with the signature (0, 1, +, ×, <) and these axioms in L_{RA} hold in the class of ordered fields. (These axioms are given below in a footnote[38]). The important *completeness axiom scheme* for RA is:

---

[36] In *second-order* arithmetic, + and × are *explicitly definable* using just 0 and s. E.g.,

$$z = x + y \leftrightarrow \forall f([\forall w(f(w, 0) = w) \wedge \forall w_1 \forall w_2(f(w_1, s(w_2)) = s(f(w_1, w_2))] \rightarrow z = f(x, y)]$$

[37] See, e.g., Birkhoff & MacLane 1965, Chapter II, 'Rational Numbers and Fields'.

[38] The field axioms for the subsignature (0, 1, +, ×) are:

RA_1	$x + y = y + x$	commutativity of +
RA_2	$x \times y = y \times x$	commutativity of ×
RA_3	$x + (y + z) = (x + y) + z$	associativity of +
RA_4	$x \times (y \times z) = (x \times y) \times z$	associativity of ×
RA_5	$x + 0 = x$	identity element for +
RA_6	$x \times 1 = x$	identity element for ×
RA_7	$\exists y(x + y = 0)$	inverse for +
RA_8	$x \neq 0 \rightarrow \exists y(x \times y = 1)$	inverse for ×
RA_9	$x \times (y + z) = (x \times y) + (x \times z)$	× distributes over +
RA_{10}	$0 \neq 1$.	

The order axioms for < are:

RA_{11}	$(x < y \wedge y < z) \rightarrow x < z$	transitivity
RA_{12}	$(x < y) \rightarrow \neg(y < z)$	antisymmetric
RA_{13}	$(x < y) \vee (x = y) \vee (y < x)$	trichotomy
RA_{14}	$x < y \rightarrow (x + z < y + z)$	monotonicity of +
RA_{15}	$(x < y \wedge 0 < z) \rightarrow x \times z < y \times z$	monotonicity of ×

RA_{16} $[\exists x R(x) \wedge \exists y \neg R(y) \wedge \forall x \forall y (R(x) \wedge \neg R(y) \rightarrow x < y)] \rightarrow \exists z (\forall x (x < z \rightarrow R(x)) \wedge \forall y (z < y \rightarrow \neg R(y))$[39]

In set-theoretical language, this is usually written,

(1) For every non-empty subset bounded above, there is a least upper bound

If $(X, <)$ is a linear ordering, and satisfies (1), we say that it is *order-complete.* This order-completeness property of $(\mathbf{R}, <)$ distinguishes the reals from the rationals, $(\mathbf{Q}, <)$. There are bounded subsets of **Q** which lack a least upper bound (supremum)[40]. Can we just fill in the "gaps" in the dense ordering of the ratios by "postulation"? This idea provoked Russell's famous remark:

> From the habit of being influenced by spatial imagination, people have supposed that series *must* have limits in cases where it seems odd that they do not. Thus, perceiving that there was no *rational* limit to the ratios whose square is less than 2, they allowed themselves to "postulate" an *irrational* limit, which was to fill the Dedekind gap ...
>
> ... The method of "postulating" what we want has many advantages; they are the same as the advantages of theft over honest toil. Let us leave them to others and proceed with our honest toil.
>
> (Russell 1919, p. 71).

I pass over the two standard set-theoretical constructions, due to Dedekind (using "cuts") and Cantor (convergent Cauchy sequences) of the reals from the rationals[41].

As it stands, this axiomatic mathematical theory of real numbers RA is pure and not applicable. Its quantifiers and variables range over real numbers and it contains no non-mathematical predicates. An *extended* applicable theory may be obtained by introducing

---

From Mendelson 1987, pp. 76-77. (The axioms are those for the elementary theory of ordered fields).

[39] I take this formulation of order-completeness from Burgess and Rosen 1997, p. 77. (They call it the "continuity scheme").

[40] E.g., consider the subset $\mathbf{A} = \{z \in \mathbf{Q}: z^2 < {}^2/_1\}$. This set is bounded above (by ${}^2/_1$ for example). But take any rational upper bound, say ${}^n/_m$, where $({}^n/_m)^2$ is larger than any element of **A**. That is, $n^2 \geq 2m^2$. Equality is impossible by the standard proof in Euclid. So $n^2 > 2m^2$. Since n and m are positive whole numbers, it follows that $n > m$. But we can find *another* ratio ${}^p/_q$ smaller than ${}^n/_m$ such that $p^2 > 2q^2$. This is another upper bound. E.g., let $p = n + 1$, $q = m + 1$. Then, $pm = (n+1)m = mn + m$ and $qn = (m + 1)n = mn + n$. So, $qn > pm$. Thus, if $n^2 > 2m^2$, then there are whole p, q such that $p^2 > 2q^2$ and $pm < qn$. So, **A** does not have least upper bound, a supremum. (Intuitively, the required supremum for **A** is irrational, i.e., just $\sqrt{2}$).

[41] See Drake & Singh 1996, Chapter 6, 'Developing Mathematics Within ZFC' (especially Section 6.4).

a two-sorted notation: change all real number variables to secondary variables (X, Y, ...), add primary variables (x, y, ...), add mixed predicates such as,

i. "co-ordinate primitives", $Co\text{-}ord_i(X_1, X_2, ..., x)$, and

ii. "measurement primitives" like $Mass_{kg}(X, x)$,

and a fixed stock of primary (sometimes called "synthetic") predicates $P(x_1, ..., x_n)$.

There is nothing within applicable real analysis corresponding to Hume's Principle in applicable arithmetic. (Of course, we might want to *count* non-denumerable collections: Hume's Principle, extended to transfinite sets, is good enough for this, assigning a cardinal number $| A |$ to every concept or set A, no matter how transfinitely large[42]). Roughly, the simplest applications of analysis within science are based on *co-ordinate* or *measurement* notions, using reals as representations or "labels" of concreta (such as points in a line, or instants in time). Such is the purpose of the mixed co-ordinate predicates like,

$Co\text{-}ord_i(X, x)$ (real number) X is the co-ordinate of (concretum) x

and measurement primitives like,

$Mass_{kg}(X, x)$ (real number) X is the mass-in-kg of (concretum) x

It is thus possible using these mixed predicates $Co\text{-}ord_i$, $Mass_{kg}$ and so on, to write down formulas expressing *relations* between non-mathematicalia (values of primary variables) and real numbers. In other words, to *apply* analysis and formulate quantitative laws.

For example, Newton's Second Law is a mixed law:

(2) for any point mass x and any instant t, there are real numbers X_1, X_2, X_3 such that X_1 *is the mass-in-kg* of x at time t, X_2 *is the acceleration-in-ms*$^{-2}$ of x at time t, X_3 *is the total force-in-N* on x at time t, and $X_3 = X_1 \times X_2$.

---

[42] See Machover 1996, p. 36 or Drake & Singh 1996, p. 47. The naïve "definition", $| A | = | B | \leftrightarrow A \approx B$, which is Hume's Principle again, is incomplete. The definition of Cantor's "alephs" $\aleph_\alpha$ and the identification of cardinals with *initial ordinals* is carried out in Drake & Singh 1996, Section 6.7.

Since the mixed predicates, 'mass-in-kg', 'acceleration-in-ms^{-2}' and 'total force-in-N' satisfy the existence and uniqueness constraints discussed below (that is, they determine functions from concreta to the reals), this can be rewritten, à la Carnap, as,

(3) $\forall x, t$ [total-force-in-N(x, t) = mass-in-kg$(x, t) \times$ acc-in-ms$^{-2}(x, t)$]

Physicists usually write down the formula $F = ma$ as a simple mnemonic for (3), but (2) makes more explicit what a physicist *understands* by Newton's Second Law. (Actually, (2) only deals with one-dimensional motion: strictly speaking, forces and accelerations are 3-*vectors*, i.e., to be represented by triples of reals).

For the *philosophical* purposes of analysing the application of analysis (and of mathematics more generally), we need to consider formulas such as those below,

a. $\forall x \exists X\, Co\text{-}ord(X, x)$

b. $\forall x \forall X_1 \forall X_2 (Co\text{-}ord(X_1, x) \wedge Co\text{-}ord(X_2, x) \rightarrow X_1 = X_2)$

c. $\forall X \exists x\, Co\text{-}ord(X, x)$

d. $[Co\text{-}ord(X_1, x_1) \wedge \ldots \wedge Co\text{-}ord(X_n, x_n)] \rightarrow [P(x_1, \ldots, x_n) \leftrightarrow F(X_1, \ldots, X_n)]$

where F is a secondary predicate.

Formulas (a) - (c) are called *representation formulas* (e.g., Burgess & Rosen 1997), asserting that concreta are represented by real numbers in a certain way. Formulas (a) and (b) are *existence* and *uniqueness* assertions, ensuring that the mixed formula $Co\text{-}ord(X, x)$ may be replaced by the formula $X = co\text{-}ord(x)$, where $co\text{-}ord(x)$ is a function symbol. Then (c) and (d) become,

c′. $\forall X \exists x (X = co\text{-}ord(x))$

d′. $[P(x_1, \ldots, x_n) \leftrightarrow F(co\text{-}ord(x_1), \ldots, co\text{-}ord(x_n))]$

The formula (c′) is thus a very strong constraint, requiring that *co-ord* be a surjective map from the set of concreta to the set of mathematicalia. When (a) - (c) are in place, we have a very strong representation relation between concreta and mathematicalia (and then it follows from a theorem of Tarski, Mostowski and Robinson that the apparatus of analysis, and thus all reference to mathematicalia, may be eliminated[43]).

A formula such as (d) or (d′) is called a *Representation Theorem* for the synthetic predicate P. Any n-tuple of concreta (x_1, ..., x_n) to which P applies is such that the mathematical predicate F is true of their co-ordinates.

For example, consider the following theory of time, which we shall call Tim. The analytic formulation of Tim has axioms:

i. $\forall X \exists t(X = co\text{-}ord(t))$

ii. $\forall t_1 \forall t_2(t_1 \neq t_2 \rightarrow co\text{-}ord(t_1) \neq co\text{-}ord(t_2))$

iii. $Before(t_1, t_2) \leftrightarrow (co\text{-}ord(t_1) < co\text{-}ord(t_2))$

Axioms (i) and (ii) are strong *representation formulas* for *co-ord*, saying that the function it determines is a bijection. Axiom (iii) is the Representation Theorem for the primary predicate *Before*. Together, the axioms (i) - (iii) say that "the structure of time" under the *Before* relation is *isomorphic* to the reals under <. But it should be stressed that (i)-(iii) are not *theorems* of applicable RA. They are contingent extensions of RA.

2.4 The Application of Set Theory

Standard presentations of set theory[44], including axiomatic set theory, ignore *applications* to non-mathematical subject matter. The reason is that the classical

---

[43] See Burgess & Rosen 1997, Chapter I.B.

[44] See Machover 1996 or Drake & Singh 1996, who write of the "cumulative type structure":

> In fact, all of our subsequent work could be considered as based on the cumulative type structure, and all the axioms we give ... can be given intuitive justification in terms of this structure; we could say that they are intended to give a formal description, or theory, of this structure.
>
> (Drake & Singh 1996, p. 9).

The mathematical notion of set (associated with Cantor and Zermelo 1908) has been associated with the so-called *iterative* conception which is sometimes constrasted with Frege's and Russell's "logical" notion of a class (the extension of a predicate or a concept, and thus governed by the inconsistent naïve comprehension principle). For further philosophical explanations of this conception, see Boolos 1971, Parsons 1977, and Pollard 1990.

mathematics of numbers and structures may be developed within ZF without any assumption as to the existence of *urelements*. More exactly, none of the constructions (significantly, the construction of ordinals and their properties) presuppose the existence of urelements or individuals. However, without too much difficulty, set theory may be formulated as an *applicable* theory. For example, a system called ZFU (for Zermelo-Fraenkel set theory with urelements) may be formulated in a two-sorted language permitting quantification over individuals (non-sets).

The easiest way to do this is to augment the language $L_\in$ of ZF by adding a monadic predicate $U(x)$ meaning "*x* is an (non-mathematical) urelement", modifying the Axiom of Extensionality as follows:

(1) $\forall x \forall y([\exists z(z \in x) \wedge \forall z(z \in x \leftrightarrow z \in y)] \rightarrow x = y)$

and adding the following three axioms:

(2) $\forall x(\exists z(z \in x) \rightarrow \neg U(x))$

(3) $\exists x(\neg U(x) \wedge \forall z(z \notin x))$

(4) $\exists x \forall z(z \in x \leftrightarrow U(z))$

which assert that no set is an urelement, that there is a set with no elements (viz., $\varnothing$) which is not an urelement, and that *there is a set* of urelements.[45]

It now follows from the Axiom of Separation that, corresponding to *any* *n*-place predicate *P*, *there is a set* of *n*-tuples (of *urelements*) of which *P* is true (the restriction to *U* of the extension of that predicate). So, by expanding the language further by adding "urelement predicates" like *rabbit*, *mammal* and so on, one obtains *theorems* of ZFU like,

(5) $\exists z \forall x(x \in z \leftrightarrow rabbit(x))$

Obviously, we can specify that a new primitive predicate is an "urelement predicate" by adding an axiom,

[45] For this construction, see Chihara 1990, pp. 148-149. Mostowski 1939 discussed the appropriate modification of NBG set theory to include ur-elements.

(6) $\forall x_1 \ldots \forall x_n(P(x_1, \ldots, x_n) \rightarrow (U(x_1) \wedge \ldots \wedge U(x_n)))$

Within this theory one can go on to define a predicate operator *Ext* as follows:

(7) $Ext(P) =_{df} \{(x_1, \ldots, x_n): P(x_1, \ldots, x_n)\}$

where *P* is any urelement predicate. One can now formulate assertions such as,

(8) $Ext(rabbit) \subseteq Ext(mammal)$

which says that the set of rabbits is a subset of the set of mammals. Indeed, from (6), plus the modified Axiom of Extensionality, one may derive a weak form of Frege's Axiom V,

AxV: $\text{Ext}F = \text{Ext}G \leftrightarrow \forall \mathbf{x}(F(\mathbf{x}) \leftrightarrow G(\mathbf{x}))$

This is to be compared with Hume's Principle:

HP^2: $\#F = \#G \leftrightarrow F \approx G$

It will be of some importance later (when we discuss "if-thenism", for example) that a set-theoretical assertion like (8)—and many like it, about such abstract entities as the *set of rabbits* and the *set of mammals*—is certainly not a *theorem* of ZFU. Such assertions are true (if true at all) only *contingently.* A more interesting example is the contingent but still set-theoretical assertion that the physical "time structure" (*Tim*, *Before*) is isomorphic to $(\mathbf{R}, <)$, where *Tim* $=_{df}$ the set of temporal instants and *Before* $=_{df}$ the relation named by the predicate 'before'.

I want to quickly introduce something that will play an important role later. Begin with the original *applied* signature $L_{\in U} = (\in, U)$ and expand $L_{\in U}$ by adding the signature of arithmetic $(0, s, +, \times)$. Add as axioms:

i. $U(0)$

ii. $U(x) \rightarrow U(s(x))$

iii. $[U(x) \wedge U(y)] \rightarrow [U(x + y) \wedge U(x \times y)]$

plus the usual axioms for Peano Arithmetic, including the induction scheme:

$$[\Phi(0) \wedge \forall x(\Phi(x) \rightarrow \Phi(s(x))] \rightarrow \forall x(U(x) \rightarrow \Phi(x))$$

These axioms say that there is an "initial" urelement z, and another $s(z)$, and another and so on; and that if x and y are urelements then so are $x + y$ and $x \times y$. Furthermore, $x + s(y) = s(x + y)$, and so on. And finally, if F is true of z and true of $s(x)$ whenever true of x, then F is true of all urelements.

Suppose we forget about $\in$ and the set-theoretical axioms and consider the subtheory, call it PA^{U}. This theory is formally equivalent to Peano Arithmetic, even though in its intended interpretation z, $s(z)$, $s(s(z))$, etc., may be concrete physical objects. In short, PA^{U} might well be (part of) a *mathematics-free physical theory of concrete entities* (we shall see later that PA^{U} is a subtheory of Field's geometrical theory of space-time: it is the theory of a infinite discrete equally-spaced sequence of space-time points with an endpoint).

By Gödel's First Incompleteness Theorem, PA^{U} is incomplete. Furthermore, it is obviously consistent, for it has a model (e.g., the set of finite ordinals, ω). By arithmetization, its consistency can be formulated as a statement Con *expressible* within the sublanguage but, by Gödel's Second Theorem, Con is *unprovable* from the axioms of PA^{U}. Now suppose we *add* the set-theoretical axioms of ZFU. We apply the mathematics. Clearly, PA^{U} + ZFU can *prove* that PA^{U} is consistent (by proving the *existence of a model* of PA^{U}: indeed, one such model is just the set of urelements U, structured by s, + and $\times$) and thus can prove the formula Con[46]. What this means is that adding ZFU to PA^{U} is *not* a *conservative extension* of PA^{U}. We shall return to this point when we come to discuss (in Chapter 9) Field's deflationism about mathematics.

[46] There is one subtlety involved. To prove Con from PA^{U} + ZFU it is necessary to allow formulas containing $\in$ to appear in the induction scheme when the language and theory is expanded. This is perfectly natural, for the intention of the scheme is that it holds for *any* property P of urelements, including properties definable using $\in$. Indeed, if the intended structure of urelements is *isomorphic* to a Peano system, then induction must be taken in this second-order manner (first-order PA has non-standard models, even non-standard *countable* models). We return to this theme briefly in Chapter 5 (where I show that deflationary truth theories are conservative, even on this assumption about axiom schemes) and in Chapter 9, where we discuss Shapiro's 1983 argument that mathematics is not conservative.

2.5 Applicable Mathematics: Summary

Summarizing the previous three sections, the primary uses of mathematics within science fall into three categories:

i. Arithmetic

Counting (assigning numbers to concepts or to classes of concreta);

ii. Analysis

(a) *Measurement* (introducing certain real-valued functions from concreta to real numbers: e.g., measurement scales, real-valued classical fields on space-time);

(b) *Geometry* (introducing co-ordinate systems on space-time);

iii. Set Theory

(a) *Forming collections* of concreta (sets and relations-in-extension);

(b) *Exemplification* (assigning *structures* as "representations" of concrete systems).

These uses (i)-(iiia) have been explained. Use (iiib) deserves a slightly more extended treatment to which we now turn.

2.6 Exemplification: Structuralist Theory of Application

In the above uses of mathematics in science we included amongst the applications of set theory the notion of *exemplification*. Unfortunately, this issue is not at all well-understood. Briefly, the suggestion is that the use of set theory within science allows us to attribute *structure* to physical systems. Set theory (plus mereology) indicates how concrete physical systems are "carved up" into complex structures, which such systems may be said to exemplify.

In science we often talk of the *structure* of physical objects or systems. For example, we say things like:

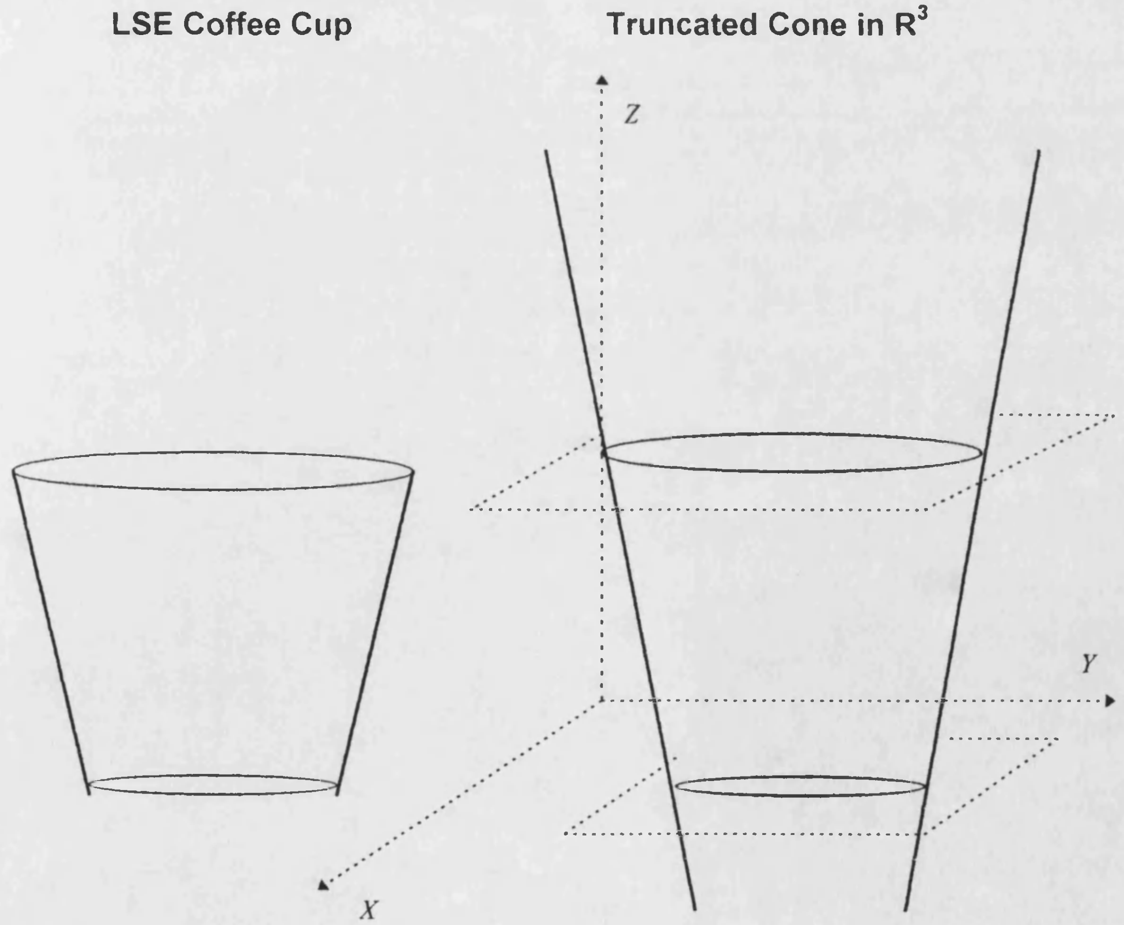
i. the *structure* of each (concrete) NaCl crystal is “Body-Centred Cubic”;

ii. the *structure* of space-time is a “differentiable manifold with a metric”;

iii. the *structure* of standard LSE coffee cups is a truncated cone in $\mathbf{R}^3$.

Similarly, we recognize of countless different types of vibrating system that their behaviour over time exemplifies the sinusoidal function $A\ sin(\omega t + \varphi)$. Things that vibrate are “simple harmonic oscillators”. And we may say that LSE coffee cups *exemplify* truncated cones.

This idea is highly intuitive, easily illustrated by the following diagram.



According to some proponents of mathematical structuralism, the notion of exemplification lies at the heart of explaining the very possibility of the application of mathematics in science. The *structuralist account of application*, advocated by Resnik and Shapiro[47], asserts that:

i. Pure mathematics characterizes and describes certain *structures*;

ii. Some of these structures may be *exemplified* by concrete objects or systems;

iii. The application of mathematics involves "translating" mathematical facts about an abstract structure Ω to contingent facts about the concrete system x which exemplifies the structures.

I am sympathetic to this (Platonist) analysis of application. But it is little more than a *sketch* of the beginnings of a theory of application. There seem to be at least three *prima facie* problems with this account:

i. What *is* "exemplification", this relation between a mathematical structure and a concrete object?

ii. *How do* we "translate" mathematical facts about the structure to contingent facts about the object?

iii. There are problems connected with the notions of *approximation* and *idealization* (as several philosophers of science have stressed).

For example, consider the following table:

"Real World" Physical System	**Idealized Mathematical Structure**
segment of the Colorado River	triangular prism, containing a vector field

[47] For Resnik's structuralism, see Resnik 1981, 1982, 1988. For Shapiro's structuralist analysis of application, see Shapiro 1983b. See also Shapiro 1997.

	satisfying the Navier-Stokes equation
football	a dodecahedron,
The Earth	an ellipsoid in $\mathbf{R}^3$
pendulum, spring, etc.	a simple harmonic oscillator.
enclosure of electromagnetic radiation	assembly of simple harmonic oscillators
crystal of NaCl	BCC lattice in $\mathbf{R}^3$

All of these are approximate idealizations. None of the real physical systems *exactly* exemplifies the mathematical structure in question. The ions in a real NaCl crystal are not points; there are always fractures and dislocations; and so on. Suffice it to say that, although many have tried, no-one has explained in any enlightening way how idealization works.

However, I suspect that we can begin to make sense of talk of exemplification itself, using ideas from set theory and *mereology*—the theory of the "part-whole" relation as applied to concrete physical entities. The basic idea is simple. With the resources of set theory and mereology one can define the *exemplification relation* between physical objects or systems and abstract structures. Roughly,

Definition 1: Exemplification w.r.t. Relations R_i

A physical object x *exemplifies* a structure Ω with respect to relations $R_1, \ldots, R_n$ if the impure structure $(D_x, R_1|_x, \ldots, R_n|_x)$ is isomorphic to Ω, where D_x is the set of parts of x and $R_i|_x$ is the restriction of R_i to D_x.

Definition 2: Exemplification Simpliciter

A physical object x *exemplifies* a structure Ω (simpliciter) if there are relations R_i such that x exemplifies Ω w.r.t. the R_i.

What then of the *translation* from the mathematics to the non-mathematical description? If a structure Ω has signature L, then associated with any such L-structure Ω

is the set of L-formulas that it satisfies. This is the *theory* of the structure, $\mathrm{Th}_L(\Omega)$. We may expect that, if x exemplifies Ω, then the theory of the structure $\mathrm{Th}_L(\Omega)$ will then be translatable into a true non-mathematical description T of the physical object x.

Two philosophically important theses might follow from this account of exemplification:

I. Anti-Essentialism:

Any physical object or system exemplifies *every* structure consistent with the cardinality of the set of its *parts*. To get "special" or "natural" structures one must single out, using a privileged vocabulary, *special* parts or special *relations* on the parts. At the very least, if x exemplifies a structure Ω and Ω is *definitionally equivalent* to another structure Ω^*, then x exemplifies Ω^*.

This insight is connected to the fact that a scientific or mathematical theory may be given many equivalent (intertranslatable) formulations, and is also related to the philosophical problems of "language dependence" that afflict confirmation theory (Goodman's "grue" paradox) and truthlikeness theory (Miller's reformulation paradoxes).

II. Ultra-Platonism

On reasonable assumptions concerning the nature of the physical world (e.g., concerning the cardinality of the set of physical individuals: say, space-time points and/or regions), there exist *transcendent structures*: a transcendent structure is a mathematical structure Ω such that no concrete physical object exemplifies Ω. That is, a transcendent structure is so "big" that it is not exemplified anywhere. If I am right, this is just a matter of *cardinality*: if the biggest physical cardinality is $\aleph_\alpha$, say, then any bigger mathematical structure will be transcendent.

2.7 Example I: Analytic Euclidean Geometry

An example of a theory incorporating mathematical analysis is *Analytic Euclidean Geometry* (AEG). Informally, analytic geometry presupposes the ordered field structure (**R**, *0*, *1*, +, ×, <) of the real numbers. Consequently, within an axiomatic treatment of analytic geometry, the pure axioms will be those for analysis.

Geometry is the theory of points, lines and regions. Intuitively, these are elements of the *physical world.* In this interpretation then, geometry is a part of *physics*. One may perhaps distance oneself from the *existence* of such points by talking of *idealization*: points are idealizations of small blobs of ink, and lines are idealizations of Socratic lines drawn in the sand, or of ink lines drawn with a pen. One may further distance oneself from any particular ontology by adopting a completely formal and axiomatic method, effectively the approach of Hilbert 1899. The axioms are set down and deductive consequences are drawn. (Such consequences hold irrespective of any intended interpretation or ontology).

On the axiomatic approach, the representation axioms (for co-ordinates) of AEG are:

$$\exists! X, Y\,(Co\text{-}ord_1(X, x) \wedge Co\text{-}ord_2(Y, x))$$

$$\exists! x(Co\text{-}ord_1(X, x) \wedge Co\text{-}ord_2(Y, x))$$

That is, each point x is represented by some unique real number co-ordinates X, Y. And all co-ordinate pairs (X, Y) represent some unique point x[48]. Any further axioms can thus be rephrased using function symbols $co\text{-}ord_1$ and $co\text{-}ord_2$, where $X = co\text{-}ord_i(x)$ is equivalent to $Co\text{-}ord_i(X, x)$.

The remaining axioms of analytic Euclidean plane geometry are then formulated using two further primary primitives:

i. $Bet(x, y, z)$, meaning intuitively "x lies between y and z", and

[48] These details about geometry are drawn from Burgess & Rosen 1997, Chapter II.A. One semi-popular presentation of formalized geometry is Tarski 1959.

ii. *Cong(x, y, z, w)*, meaning intuitively "the segment *xy* is just as long as the segment *zw*".

The first axiom is a representation theorem for betweenness, *Bet*:

$$Bet(x_1, x_2, x_3) \leftrightarrow \exists U[(0 \leq U \leq 1) \wedge (co\text{-}ord_1(x_2) = Uco\text{-}ord_1(x_1) + (1 - U)co\text{-}ord_1(x_3) \wedge (co\text{-}ord_2(x_2) = Uco\text{-}ord_2(x_1) + (1 - U)co\text{-}ord_2(x_3))]$$

In effect, x_2 lies between x_1 and x_3 just in case x_2, represented by (X_2, Y_2), lies on the "parametrized straight line" from x_1 (resp., X_1, Y_1) to x_3 (resp., X_3, Y_3). (See Burgess & Rosen 1997, p. 103).

The second axiom is a representation theorem for congruence, *Cong*:

$$X_i = co\text{-}ord_1(x_i) \wedge Y_i = co\text{-}ord_2(x_i) \rightarrow$$

$$Cong(x_1, x_2, x_3, x_4) \leftrightarrow (X_1 - X_2)^2 + (Y_1 - Y_2)^2 = (X_3 - X_4)^2 + (Y_3 - Y_4)^2$$

In effect, the segment x_1x_2 is congruent to the segment x_3x_4 just in case the "Euclidean" metric distance between (the co-ordinates representing) x_1 and x_2 is equal to that between x_3 and x_4. (See Burgess & Rosen, p. 113).

Equipped with these axioms plus the axioms for analysis one can then proceed to derive theorems about points (i.e., theorems saying how points are related by betweenness and congruence). Indeed, one may derive theorems which do not mention real numbers at all. Such theorems are called "synthetic theorems".

For example, we first define,

"*x*, *y* and *z* are collinear"

by the permutation formula,

$$Bet(x, y, z) \vee Bet(y, z, x) \vee Bet(z, x, y)$$

Equipped with just the first representation axiom for betweenness, plus the axioms of analysis, one can derive *Playfair's Postulate* as a theorem (see Burgess & Rosen 1997, p. 104):

for any x, y, z that are not collinear, there exists a point u such that no point is collinear both with x and y and with z and u;

and if v is any other such point, then z is collinear with u and v.

Burgess & Rosen 1997 mention several other theorems of elementary geometry deducible in this theory, commenting that this is a "miniature illustration of the usefulness of algebra and analysis in geometry (and mechanics)" [p. 104].

2.8 Example II: The Structure of Time

Consider the following simple theory $\mathsf{Tim_S}$ of the structure of time. In fact, a theory along these lines is believed to be *true* by physicists who study the structure of space-time (or rather, this theory is a subtheory of standard space-time theory). The theory talks of temporal instants and of the temporal "before-after" relation on these instants, using (primary) primitives $Inst(x)$ and $Before(x_1, x_2)$. Aside from any implicit set-theoretical assumptions, the theory contains two definitions and one mixed axiom:

Def$_1$: $Tim = \{x: Inst(x)\}$

Def$_2$: $Bef = \{(x_1, x_2): Before(x_1, x_2)\}$

Ax: (Tim, Bef) *is isomorphic to* $(\mathbf{R}, <)$

(This theory $\mathsf{Tim_S}$ includes some suppressed *set theory*, hence the subscript S).

Now $\mathsf{Tim_S}$ can be given a sort of "analytic reformulation", which we call Tim, that does not mention *sets* (or relations on sets). However, it introduces a co-ordinate notion, *co-ord*, functionally correlating instants with real numbers. Thus,

Axiom Type 1: secondary ["pure"] axioms

axioms for analysis

Axiom Type 2: mixed [representation] axioms

(Tim$_1$) $\forall x_1 \forall x_2 (x_1 \neq x_2 \rightarrow co\text{-}ord(x_1) \neq co\text{-}ord(x_2))$

(Tim$_2$) $\forall X \exists x(X = \textit{co-ord}(x))$

(Tim$_3$) $\textit{Before}(x_1, x_2) \leftrightarrow \textit{co-ord}(x_1) < \textit{co-ord}(x_2)$

Like analytic geometry, this theory has no primary axioms but it certainly has primary *theorems*. For example, it is a theorem of Tim that "density" holds of the temporal instants:

$$\forall x_1 \forall x_2(\textit{Before}(x_1, x_2) \rightarrow \exists x_3(\textit{Before}(x_1, x_3) \wedge \textit{Before}(x_3, x_2)))$$

Indeed, any pure mathematical theorem about real numbers expressed using <, can be "translated", via the representation axioms (Tim$_1$)-(Tim$_3$) into a *non-mathematical theorem about temporal instants*, expressed using *Before*.

Recall what we said earlier in Section 1.2 about the application of mathematics to the geometry of the Earth. The Earth, a physical system, is "represented" as an abstractum, an ellipsoid embedded in $\mathbf{R}^3$; then *mathematical facts* about the geometry or topology of this abstract ellipsoid can be "*translated*" into non-mathematical facts about the genuine physical geometry or topology of the Earth.

I suggest that the above analytically formulated theory Tim illustrates more accurately how this happens. For example, the density theorem about the real numbers:

(1) $\forall X_1 \forall X_2 (X_1 < X_2 \rightarrow \exists X_3(X_1 < X_3 \wedge X_3 < X_2))$

"translates", using the Representation Theorem, into the non-mathematical theorem:

(2) $\forall x_1 \forall x_2(\textit{Before}(x_1, x_2) \rightarrow \exists x_3(\textit{Before}(x_1, x_3) \wedge \textit{Before}(x_3, x_2)))$

Roughly, *any* abstract property of the system of reals expressible using < is translatable into a physical property of time, using *Before*. Again this is another illustration of the usefulness of analysis and algebra within physics.

We shall return to this simple mathematicized theory Tim later on for purposes of illustrating other arguments concerning the application of mathematics.

CHAPTER 3

The Quine-Putnam Argument

> Science would be hopelessly crippled without abstract objects. We quantify over them. In the harder sciences, numbers and other abstract objects bid fair to steal the show. Mathematics subsists on them, and serious hard science without serious mathematics is hard to imagine
>
> W.V. Quine 1995, *From Stimulus to Science*, p. 40.

> Surely one of the strongest reasons—if not the only reason—for taking mathematical truth seriously stems from the apparently indispensable role mathematical theories play in the very formulation of scientific descriptions of the material world around us.
>
> Geoffrey Hellman 1989, *Mathematics Without Numbers*, p. 94.

3.1 The Quine-Putnam Argument

Much of modern accepted science presupposes, and entails, the existence of mathematicalia. A provocative way to put this is to say that "science entails platonism". The originator of this argument was W.V. Quine, who has presented it several times over the years since the late 1940s. Several authors, of nominalistic inclination, have found this ontological conclusion, concerning the commitment of science to mathematicalia, somewhat upsetting and have struggled valiantly to demonstrate that this commitment is "only an appearance", or that science can "get by without mathematics". But, I shall argue, *such approaches do not work.* Much of science is irreducibly up to its neck in commitments to abstract mathematicalia, such as numbers, functions, sets, vectors, tensors, and so on.

Quine was the first philosophical author to propose a *criterion of ontological commitment.* Quine's proposal, sloganized in those ten immortal words,

To be is to be the value of a variable develops through several publications, going back to Quine 1939 and reaching its "canonical" form in Quine 1948:

> A theory T (perhaps formalized within an interpreted canonical notation L) is committed to those entities which must be taken as values of the bound variables of theorems of T if T is to be true.

From this, Quine quickly establishes (in his own words) that *classical mathematics is platonistic*: i.e., the theorems of classical mathematics are committed to an ontology of mathematicalia, via the occurrence of bound variables ranging over such entities as numbers and sets. Finally, he establishes that *accepted scientific theory is platonistic*: i.e., because mathematics is an "integral part" of virtually any serious scientific theory, much of accepted scientific theory is committed to an ontology of mathematicalia.

Before we examine this last claim, we should recognize that "what ontology *is to be accepted*" is not dictated by the criterion of ontological commitment itself:

> Now how are we to adjudicate between rival ontologies? Certainly the answer is not provided by the semantical formula "To be is to be the value of a variable"; ... we look to bound variables in connection with ontology not in order to know what there is, but in order to know what a given remark or doctrine, ours or someone else's, *says there is*;
>
> (Quine 1948 (1980), p. 15).

Quine's argument for actually *accepting* a platonistic ontology appears in the final pages:

> Our acceptance of an ontology is, I think, similar in principle to our acceptance of a scientific theory, say a system of physics: *we adopt, at least insofar as we are reasonable, the simplest conceptual scheme into which our disordered fragments of raw experience can be fitted and arranged.* Our ontology is determined once we have fixed upon the over-all conceptual scheme which is to accommodate science in its broadest sense; and the considerations which determine a reasonable construction of any part of that conceptual scheme, for example, the biological or physical part, are not different in kind from the considerations which determine a reasonable construction of the whole.
>
> (Quine 1948 (1980), pp. 16-17. Emphasis added).

Quine goes on to discuss the acceptability of a platonistic ontology of classes. He argues that we must accept them into the ontology of our over-all conceptual scheme:

> ... Now what of classes or attributes of physical objects, in turn? A platonistic ontology of this sort is, from the point of view of a strictly physicalistic conceptual scheme, as much a myth as that physicalistic conceptual scheme itself is for phenomenalism. This higher myth is a good and useful one, in turn, in so far as it simplifies our account of physics. Since *mathematics is an integral part* of this higher myth, the utility of this myth for physical science is evident enough.
>
> ... The analogy between the myth of mathematics and the myth of physics is, in some additional and perhaps fortuitous ways, strikingly close.
>
> ... Let us see *how, or to what degree, natural science may be rendered independent of platonistic mathematics*; but let us also pursue mathematics and delve into its platonistic foundations.
>
> (Quine 1948 (1980), pp. 17-19. Emphasis added).

Several philosophical themes are interleaved in these two passages. In the former, we see Quine's pragmatism: a pragmatic account of theory acceptance ("we adopt the simplest conceptual scheme into which our disordered fragments of raw experience can be fitted and arranged"), which might be contrasted, say, with an "evidence-driven" account, wherein we should adopt the "most confirmed" conceptual scheme[49]. Furthermore, while Quine's epistemological pragmatism is a much-debated topic, he has repeatedly stressed that his position is inconsistent with scientific anti-realism, scepticism or instrumentalism (that is, the claim that theories are *nothing more* than useful instruments for "saving the phenomena" of "raw experience").

However, the crucial point is that when we accept a standard scientific theory, we are thereby accepting an ontology of mathematicalia, for mathematics is an "integral part" of virtually any such system. The premise involved is the following thesis:

Integration Thesis

> Mathematical assumptions, axioms, etc. (i.e., axioms which quantify over mathematicalia), form an *integral part* of modern physical science.

The integration thesis is extremely important, especially in relation to possible objections to Quine's argument. In particular, it is a parody of the logical facts to think of

---

[49] This is complicated by the fact that a confirmation theorist might want to include simplicity considerations within the calculation of a theory's degree of confirmation (although this is clearly absent from Bayes' Rule). Use 'justification' as a neutral term. Then Quine is a justificationist who lays as much stress on *pragmatic* considerations (e.g., simplicity) as on *evidential* considerations (empirical support) in assessing the rational acceptability of a theory. See Quine 1963b.

a scientific theory T that uses mathematical axioms as a facile *logical conjunction*, say N $\cup$ M, where N represents the physical facts (without reference to mathematicalia) and M comprises some *pure* mathematical axioms. Indeed, the way in which mathematics is actually integrated is extremely complex: a typical scientific theory uses a whole host of *mixed predicates* (and *functors*) governed by sophisticated axioms. For example, mixed predicates like the following play a basic and prima facie indispensable role in formulating scientific theories:

x is an element of (the set) X

(the real number) X measures how massive x is

x is a token of type X

the quadruple of reals (X^0, X^1, X^2, X^3) measures the spatio-temporal location, relative to co-ordinate chart φ, of space-time point x

(the real number) X measures the value of the φ field at space-time point x

These are basic, and prima facie irreducible, mixed predicates, especially the latter two, which might be more conventionally written

$(x^0, x^1, x^2, x^3) = \varphi(e)$

$r = \varphi(e)$,

where e ranges over space-time points. Predicates such as these express the crucial *relations* between concreta and abstracta vital to the application of mathematics in science. For example, if X is a real number (say, 100) and x a concretum (say, an apple) and X measures how massive x is, then we are asserting a relation between X and x. Actually, the mass-measuring relation must be a relation-preserving relation: i.e., if the real number X_1 is assigned as a mass measure to x_1, and similarly X_2 assigned to x_2, then x_1 is less massive than x_2 just in case X_1 is less than X_2. Indeed, as we shall see later, it is precisely the problem of dealing adequately with the presence of these *mixed predicates* (which occur in virtually all important accepted scientific theories in physics) that frustrates the most sophisticated version of modern nominalism, Field's deflationism.

Furthermore, Quine makes it clear that he himself *accepts as true* standard physical theory, with its overall ontological commitments (to theoretical entities, mathematicalia and all) and considers it "a scientific error to believe otherwise". E.g.,

> As an empiricist, I continue to think of the conceptual scheme of science as a tool, ultimately, for predicting future experience in the light of past experience. Physical objects are conceptually imported into the scheme as convenient intermediaries—not by definition in terms of experience, but simply as irreducible posits comparable, epistemologically, to the gods of Homer. *For my part, I do, qua lay physicist, believe in the physical objects and not in Homer's gods; and I consider it a scientific error to believe otherwise.*
>
> (Quine 1951a (1980), p. 44. Emphasis added)

This talk of epistemological parity between 'homeric god' and 'physical object' can be misleading, perhaps suggesting that Quine thinks that objects are somehow "produced" by our theories. *Qua concepts*, there is parity between a mythical concept such as 'homeric god' and a theoretical predicate like 'electron'. They are pieces of language, incorporated within vastly complex theories, and ultimately tied to sensory experience. But *whether there exist entities* answering to these concepts is obviously a distinct and separate question. It is the question whether sentences such as '$\exists x$(x is an electron)' and '$\exists x$(x is a homeric god)' are true. Indeed, it is *this* question that divides realists and anti-realists. Quine adopts realism, insisting that the '$\exists x$(x is a electron)' is true and '$\exists x$(x is a homeric god)' is false: according to Quine, there *really are such things as electrons*, and there are no such things as homeric gods[50]. Thus,

> ... in likening the physicists' posits to the gods of Homer, in that essay [Quine 1948] and in "Two Dogmas", I was talking *epistemology and not metaphysics.* Posited objects can be real. As I wrote elsewhere, to call a posit a posit is not to patronize it.
>
> (Quine 1980, Foreword, p. viii. Emphasis added).

> I have given our objector his fair share of program time. He would like us to believe, I suppose, that the atoms and elementary particles and the sets and numbers and functions are unreal; mere heuristic fictions? Is that right? ... Who is to say, then, whether they exist or were invented? Have we here reached the limit of knowledge, the unanswerable question?
>
> I think not. If we subscribe to our physical theory and our mathematics, as indeed we do, then we thereby accept these particles and these mathematical

[50] To be a realist about *F*s is simply to assert that *there are F*s. In contrast, van Fraassen, for example, is *not* a realist about electrons. Van Fraassen will *not* assert that there are electrons. He claims to be agnostic.

> objects as real; it would be an empty gesture meanwhile to cross our fingers as if to indicate that what we are saying doesn't count.
>
> (Quine 1973 (1976), p. 65).

It is worth remarking that for Quine (and for Tarski, who in his 1936 emphasized the disquotational aspect of alethic concepts of truth, satisfaction and so on) the phrases 'accepts' and 'accepts as true' are simply interchangeable[51].

3.2 Recent Formulations

After Quine's lead, the argument mathematical realism based on science took a new and important turn with the appearance of *Philosophy of Logic* by Putnam in 1971:

> ... quantification over mathematical entities is indispensable for science ... therefore we should accept such [talk]; but this commits us to accepting the existence of the mathematical entities in question. This type of argument stems, of course, from Quine, who has for years stressed both the *indispensability* of [talk about] mathematical entities and the *intellectual dishonesty* of denying the existence of what one daily presupposes.
>
> (Putnam 1971 (1979), p. 347).

Putnam made the same Quinian point more polemically later, in Putnam 1975:

> ... mathematics and physics are integrated in such a way that it is not possible to be a realist with respect to physical theory and a nominalist with respect to mathematical theory.
>
> (Putnam 1975 (1979), p. 74).

Putnam's illustration involved the Law of Gravitation:

> ... one wants to say that the Law of Gravitation makes an objective statement about bodies—not just about sense data or meter readings. What is this statement? It is just that bodies behave in such a way that the quotient of two numbers associated with the bodies is equal to a third number associated with the bodies. But how can such a statement have any objective content at all if numbers and 'associations' (i.e., functions) are alike mere fictions? It is like trying to maintain that God does not exist and angels do not exist while maintaining at the same time that it is an objective fact that God has put an

[51] The equivalence of accepting T and accepting 'T is true' is closely connected with the disquotational aspect of truth, which is surely *constitutive* of the concept of truth (the axioms or theorems of the form ''*p*' is true iff *p*' of any disquotational truth theory are *analytic*). This need not be a deflationary view of truth, for such a view claims also that disquotation is *all there is* to truth. This is different. Indeed, I think I can *prove*, using Gödelian techniques, that the weak truth theory, consisting of "all" the disquotational axioms, is an incomplete theory of the concept of truth (see Chapter 5 below).

> angel in charge of each star and the angels in charge of each binary star were always created at the same time! If talk of numbers and 'associations' between masses, etc., and numbers is theology (in the pejorative sense), then the Law of Universal Gravitation is likewise theology.
>
> (Putnam 1975 (1979), pp. 74-75).

We may thus put together the combined argument of Quine and Putnam. It clearly comprises two main theses:

Thesis A:	**Science is platonistic (Quine 1948)**
Thesis B:	**Platonistic science is indispensable (Putnam 1971)**

This combined argument is sometimes called the *Quine-Putnam Indispensability Argument* for platonism or mathematical realism.

Many philosophers of mathematics, realists or not, accept Quine's Thesis A. An example is Field, who is a nominalist:

> After all, the theories that we use in explaining various facts about the world not only involve a commitment to electrons and neutrinos, *they involve a commitment to numbers and functions and the like*. (For instance, they say things like 'there is a bilinear differentiable function, the electromagnetic field, that assigns a number to each triple consisting of a space-time point and two vectors located at that point, and it obeys Maxwell's equations and the Lorentz force law'.) I think that this sort of argument for the existence of mathematical entities (the Quine-Putnam argument, I'll call it) is an extremely powerful one, at least *prima facie*.
>
> (Field 1989, Introduction, p. 17. Emphasis added).

Field's nominalist strategy is based on trying to overcome Thesis B, Putnam's indispensability component. In brief, Field proposes that we attempt to construct a non-mathematical *replacement* for platonistic science. Indeed, Field argues convincingly and correctly that,

> If one just advocates fictionalism [i.e., nominalism] about a portion of mathematics, without showing how that part of mathematics is dispensable in applications, then one is engaging in *intellectual doublethink*: one is merely taking back in one's philosophical moments *what one asserts in doing science*.
>
> (Field 1980, p. 2. Emphasis added).

Almost all other prominent philosophers of mathematics similarly endorse Quine's classification of science as platonistic. Examples are Lewis, Resnik, Burgess, Maddy and Shapiro[52]. For example, opening his critique of the central technical claim of Field's nominalism (the conservativeness claim), Shapiro writes:

> A common argument for platonism, due to Putnam and Quine, is based on the role of mathematics is science. The main premiss is a statement to the effect that mathematics forms an essential part of virtually every scientific theory. That is, scientific theories of the material world are formulated in mathematical terms and have variables ranging over abstract mathematical entities, such as numbers. It follows from widely accepted principles of ontological commitment that such theories presuppose the existence of these abstract entities.
>
> (Shapiro 1983a (1996), p. 225).

Having looked at these formulations, we will in the next section look at a semi-formal argument based on Quine's position, an argument which derives platonism from science.

3.3 A Semi-Formal Argument

In this section we prove two rather trivial metatheorems, the first of which we shall call the "Basic Metatheorem". In spite of its triviality, this theorem plays the central role in the argument for mathematical platonism. It says that if certain kinds of scientific theories are true, then there must be such things as mathematicalia. The Basic Metatheorem does not, of course, prove that mathematical platonism is true. It simply demonstrates that mathematical platonism is a *logical consequence* of certain other statements (namely, many theoretical statements from science).

---

[52] Perhaps the only major philosopher of mathematics who disputes Quine's Thesis A is Chihara. Chihara argues in his 1973 that Quine's criterion of ontological commitment may be misleading, and that it is not beyond dispute that treating the quantifiers of mathematical assertions as standard first-order quantifiers ranging over a range of (abstract) mathematical entities may not be the correct interpretation of mathematical statements. Chihara proposes (in his 1973 and his 1990) a *non-literal reinterpretation* of mathematical theories, using a certain modal quantifier, called a "constructibility quantifier", which Chihara claims not to carry the Quinian commitment to abstract mathematical entities. We return to Chihara's position in Chapters 6, 7 and 8.

It is crucial to emphasize that the statements that imply mathematical platonism do not *beg the question* in any interesting sense[53]. For careful examination of our overall accepted mathematicized scientific picture of the world, our "overall conceptual scheme", shows that the existence of mathematicalia (i.e., mathematical platonism) follows logically from a vast multitude of important statements *already accepted* by scientists, logicians and philosophers. In short, the semi-formal argument is an attempt to make explicit what was implicit all along: the scientific reification of abstract mathematical entities.

The second metatheorem shows why, given a mathematicized theory T, the result of simply "deleting" the axioms containing references to mathematical entities need not yield a suitable replacement for T. Call the result of this deletion T°, its *primary restriction.* It is possible to show that in some (important) cases, T° will actually lose some of T's deductive consequences about concrete entities. Technically, T is not a *deductively conservative extension* of its primary restriction T°.

Suppose that we convert our two-sorted notation L to a single-sorted notation by introducing (by a conservative expansion) the following predicate *Math* which, intuitively, means 'mathematical entity':

Definition 1: The One-Sorted Language, L_{Math}:

Let L be a standard two-sorted mathematicized language (introduced in Chapter 2). Let L_{Math} be the resulting of converting L to a one-sorted language as follows:

Add a monadic predicate *Math* governed by the following axioms: for any mixed or secondary formula $P(X, ..., x...)$,

$$\forall X(P(X_1, X_2, ..., X_n, x, ...) \rightarrow (Math(X_1) \wedge Math(X_2) \wedge \ldots))$$

[53] Of course, every valid deductive argument *does* beg the question in a trivial sense, for any logical consequence of a class of premises Σ is *already* an element of the consequence class of Σ. Properly speaking, begging the question is not a logical fallacy, but an argumentative one. One way of arguing that a person *P* ought to believe a statement φ *without begging the question*, is to show that, unbeknownst to *P*, φ already follows from *other statements which P already accepts*. Another way, as practised by mothers and policemen, is to "rub *P*'s face" in the observable evidence for φ.

is an axiom.

Now replace all secondary variables by primary ones.

It can be proved that the addition of the predicate *Math*, and the axioms governing it, constitutes a conservative expansion of any theory T in L. That is, no new theorems expressible in L are derivable.[54]

The definition of *Math* guarantees intuitively that if x is a set, then *Math* is true of x; and if x is a real number, then *Math* is true of x; and so on. In short, *Math* applies to mathematicalia. More exactly, if *Set* applies to x, then *Math* applies to x and if *Real Number* applies to x, then *Math* applies to x, and so on. Intuitively, *Math*(x) means 'x is a mathematical object'.

Definition 2: Existential Mathematicized Theory (EMT)

A theory T standardly formulated in our mathematicized language L_{Math} is an *existential mathematicized theory* (an EMT) if $T \vdash \exists x Math(x)$[55].

Now we can define mathematical platonism:

Definition 3: Mathematical Platonism

Let L_{Math} be the one-sorted mathematicized language. Then,

mathematical platonism $=_{df} \exists x Math(x)$.

[54] The proof of this fact requires only the lemma that if $\mathfrak{I}$ is a model of T in L, then there is an expansion $\mathfrak{I}_M$ which is a model of T $\cup$ {axioms for *Math*} in L_{Math}. This is not difficult to prove. Any interpretation $\mathfrak{I}$ of our two-sorted language L, has two domains, $D_1(\mathfrak{I})$ (over which the primary variables range) and $D_2(\mathfrak{I})$ (over which the secondary variables range). Now suppose $\mathfrak{I}$ is a model of T. It is easy to see that the L_{Math}-interpretation $\mathfrak{I}_{Math}$ obtained by simply assigning the whole of $D_2(\mathfrak{I})$ as the extension in $\mathfrak{I}$ of *Math*, yields an expansion of $\mathfrak{I}$ which is still a model of T and satisfies all the axioms for *Math*.

[55] I am going to analyse ontological commitment of an interpreted theory T by demanding that T implies $\exists x F(x)$, where F is true (in its intended interpretation) of certain entities (in our case, mathematicalia). This is slightly different from Quine's 1948 criterion. Indeed, this criterion using existential quantification was advocated by Church 1958a. It avoids problems associated with the fact that even a set of elementary logical truths carries commitment to *at least one object*, and a many-sorted extension of classical first-order logic requires one object in each of its variable domains. Thus, if x_i is a variable of sort i, then $\exists x_i(x_i = x_i)$ is a logical truth. *Pace* Carnap 1950, we do not want our account to permit the "creation" of new objects by simply introducing a new sort of variable. I owe this caveat to David Miller.

By Definition 3, mathematical platonism is simply the *existential thesis* that there are mathematicalia. Strictly speaking, this is consistent with mathematical idealism. But we gave reasons in Chapter 1 for not taking existential statements about mathematicalia as statements about "what can be constructed". In what sense can the electromagnetic field A_μ or the Riemann curvature tensor $R_{\mu\nu}{}^{\rho}{}_{\sigma}$ of space-time be "constructed"?. As I admitted earlier (echoing Frege, Hardy and Field) mathematical idealism is hard to make serious ontological sense of, especially in connection with the application of mathematics.

The point of introducing *Math* is only to obtain the requisite generality over mathematicalia. Without *Math* we should only be able to derive consequences like $\exists X(\mathit{Real\text{-}Number}(X))$, $\exists X(\mathit{Set}(X))$, and so on. Our soon-to-arrive philosophical theorem would then say that science is committed to the existence of real numbers or sets or ..., rather than that science is committed to the existence of mathematicalia in general.

The trivial metatheorem then follows:

Basic Metatheorem: Science Entails Platonism

Let T be any EMT formulated in L_{Math}. Then T $\cup$ {axioms for *Math*} implies mathematical platonism.

Proof (obviously trivial): Let T be any EMT formulated in L_{Math}. Then there is a monadic mixed or secondary predicate P such that T $\vdash \exists x P(x)$. Since for any such predicate P, $\forall x(P(x) \rightarrow \mathit{Math}(x))$ is an axiom, T $\cup$ {axioms for *Math*} $\vdash \exists x \mathit{Math}(x)$. Hence T implies mathematical platonism. ■

This is the Basic Metatheorem. Any EMT implies $\exists x \mathit{Math}(x)$. But doesn't $\exists x \mathit{Math}(x)$ express the existence of mathematicalia? If so, we have proved that any EMT implies mathematical platonism.

Of course, the implication of $\exists x \mathit{Math}(x)$ is very weak. For example, consider the theory below:

$$\exists X_1 \exists X_2 [X_1 \neq X_2 \wedge (X_1 \times X_1 = X_1) \wedge (X_1 \times X_2 = X_2) \wedge (X_2 \times X_1 = X_2) \wedge (X_2 \times X_2 = X_1)]$$

This theory implies the existence of just two mathematicalia: it is the theory C_2 of the cyclic Abelian group of order 2, more easily recognized by the group multiplication table:

×	*e*	*a*
e	*e*	*a*
a	*a*	*e*

Typically, however, interesting and important applications of mathematics within science will use either arithmetic, analysis or set theory, and thus will imply the existence of *infinitely* many mathematicalia: specifically, numbers (whole or real), or sets.

But all this is slightly tangential. Virtually every modern scientific physical theory is what I have called an EMT. So, the range of application of the above analysis is staggering. The theorem obviously also applies to chemistry, linguistics, psychology, etc.: modern science logically implies mathematical platonism. It seems surprising to some writers that mathematical science—the heritage of Pythagoras, Eudoxus, Archimedes, Galileo, Newton, Maxwell and Einstein—logically entails the existence of mathematicalia. But, unless there is something radically amiss in modern logic, it does.

3.4 The Trivial Indispensability Metatheorem

The second metatheorem concerns the following very naïve attempt to eliminate reference to, and thus ontological commitment to, mathematicalia. Let T in L be an axiomatic mathematicized theory of Nature. One might (very naïvely) think that the mathematical components of T are dispensable by simply deleting from the axioms of T any that refer to (quantify over) mathematicalia. If T is formulated in a two-sorted language as above, it is possible to do this, for the axioms that refer to mathematicalia are simply the *secondary* and *mixed* axioms. Let T° be the *primary restriction* of T, that is, the result of deleting the secondary and mixed axioms from T. Clearly, T° does not imply $\exists x Math(x)$. It is thus nominalistically acceptable.

One might hope that T° is an acceptable replacement for T. One obvious constraint on T°'s being acceptable is this: the procedure does not eliminate any of the *non-mathematical theorems* of T. If an axiomatic theory T has this property of having a primary restriction with all the same primary consequences, we then say that the theory T is a *deductively conservative extension* of the primary restriction T°.

Unfortunately, this simple-minded manoeuvre does not work. For we can easily show that there are simple examples of mathematicized theories T such that T is not a conservative extension of its primary restriction T°. To illustrate this simply, suppose that L is a standard two-sorted language containing the secondary primitives for real analysis, a primary predicate $Before(x, y)$, the mixed predicate $Co\text{-}ord(X, x)$ and a pair of primary constants a and b. A theory formulated in L is said to be *analytically formulated*: it is formulated with respect to the mathematics of real analysis.

Now consider the following theory T in L. Assume that T contains the usual pure axioms for analysis plus the mixed axioms:

T_1 $\quad \forall x \exists X(Co\text{-}ord(X, x))$

T_2 $\quad \forall X \exists x(Co\text{-}ord(X, x))$

T_3 $\quad Co\text{-}ord(X, x) \wedge Co\text{-}ord(Y, y) \rightarrow (X < Y \leftrightarrow Before(x, y))$

(this last axiom is sometimes called a "Representation Theorem"), and the primary axiom:

T_4. $\quad Before(a, b)$

From these axioms plus the secondary (density) axiom for real numbers we can derive the primary ("synthetic") theorem:

φ: $\quad \exists z(Before(a, z) \wedge Before(z, b))$.

Thus, T $\vdash \varphi$.[56] With this information we can prove the following metatheorem:

Metatheorem 2: Indispensability

the theory T in L is not a *deductively conservative extension* of T°.

Proof: We know that T $\vdash \varphi$. Now the sole primary axiom is T_4, so the primary restriction T° is just T_4. But it is obviously false that $T_4 \vdash \varphi$. So the theory T is not a deductively conservative extension of its primary restriction T°. ■

Thus, for a mathematicized theory, there may be non-mathematical assertions about primary entities (that is, assertions about concreta) that are not derivable in T° but are derivable in T. This means that simply taking the primary restriction of T does not generally yield even a minimally adequate replacement for T. More briefly, T may *not be dispensable in favour of* T°.

3.5 Philosophical Analysis of the Basic Metatheorem

From a logical point of view, the Basic Metatheorem is trivial, saying that certain kinds of precisely (i.e., standardly) formulated theories logically imply a certain statement, namely $\exists x Math(x)$. This consequence expresses the existence of mathematicalia.

The philosophical analysis of the Basic Metatheorem rests on two claims:

i. science contains EMTs (each of which logically implies $\exists x Math(x)$);

ii. $\exists x Math(x)$ says that there exist mathematicalia.

If we are not yet satisfied with the obvious interpretation of the Basic Metatheorem, the philosophical analysis will have to concentrate on two questions:

---

[56] The proof is this: let 'C' abbreviate 'is the co-ordinate of' and 'B' abbreviate 'is before'. Then, by axiom T_1, $\forall x \exists X C(X, x)$. And, axiom T_2, $\forall X \exists x C(X, x)$. By the representation theorem T_3, $C(X, x) \wedge C(Y, y) \rightarrow (X < Y \leftrightarrow B(x, y))$. By the density theorem, $\forall X_1 \forall X_2 \exists X_3 (X_1 < X_3 \wedge X_3 < X_2)$. Finally, by the primary axiom, $B(a, b)$. From T_1, $\exists X C(X, a)$ and $\exists X C(X, b)$. So, $C(X_1, a)$ and $C(X_2, b)$. Using the representation theorem, $X_1 < X_2 \leftrightarrow B(a, b)$. Hence, $X_1 < X_2$. From the density theorem, $\exists X_3 (X_1 < X_3 \wedge X_3 < X_2)$. Hence, $X_1 < X_3 \wedge X_3 < X_2$. Now, from axiom T_2, $\exists z C(X_3, z)$. Hence, $C(X_3, z)$. Finally, by the representation theorem again, $B(a, z) \wedge B(z, b)$. So there is a z which lies temporally between a and b. ■

Qi. Are there scientific theories whose logical formulations are EMTs?

Qii. Does $\exists x Math(x)$ really say that there exist mathematical entities?

It is surely undebatable that countless scientific theories have precise standard logical formulations which are EMTs. I have given some important examples already. So, almost trivially, Qi receives a positive answer.

Moreover, Qii concerns what the formula $\exists x Math(x)$ expresses, and this is a *semantical* problem: however, given the standard accounts we now possess of truth and ontological commitment, there is little hope of avoiding a positive answer to the question Qii. As a preliminary step in including these semantico-ontological assumptions, I wish to develop a more "philosophically-oriented" argument which, schematically, is:

Science + *Semantic Analysis* $\Rightarrow$ *Mathematical Platonism*

The premises are:

Science

there is a true scientific theory T whose standard logical regimentation T_{reg} is an EMT;

Semantic Analysis

i. Scientific theories, especially mathematicized theories, can be standardly formalized within some standard formalized language. So, for any theory T, if T is true, then its standard formalization T_{reg} is true,

ii. The truth theory for (semantics for), and ontological theory pertaining to, such standard formalizations is standard Tarskian semantic theory and the ontological theory is based on Quine's standard account of ontological commitment.

First, notice that the premise *Science* is *metalinguistic*: it talks of some true scientific theory T such that its standard logical regimentation T_{reg} is an EMT. Why is this metalinguistic statement equated with "Science"? Why not simply take some particular scientific theory, say General Relativity (GR), as a premise? There are two reasons.

First, we have been explicitly speaking metalinguistically from the outset. We have not argued thus:

i. there is a physical phenomenon X

ii. the best explanation for X assumes the existence of mathematical entities

Instead, we have semantically ascended and have been considering some already accepted mathematicized *theories* (which do in fact possess tremendous explanatory and predictive power) and discussing their ontological commitments. For example, we want to know what (a precise formalization of) GR logically implies there to be and this means that the conclusion we would be looking for would have the form,

(1) if GR is true, then there are *F*s.

Of course, we might simply look for a conclusion of the form,

(2) if [GR], then there are *F*s,

where '[GR]' is replaced by an explicit conjunction of the axioms of GR. However, operating explicitly with these axioms would complicate matters and the only fact of importance is that GR $\vdash \exists x Math(x)$. From this fact, plus the soundness of $\vdash$, we derive the conclusion that if GR is true then $\exists x Math(x)$ is true. The (disquotational) behaviour of the truth predicate is then invoked to derive the conclusion, that there are mathematicalia, from the intermediate conclusion that $\exists x Math(x)$ is true. In effect, we prove that if GR is true, then there are mathematicalia.

There is a second reason for the detour through talk of truth. The deductive consequences of a theory T are, in the presence of a standard definition or theory of truth, included in the deductive consequences of the metalinguistic assertion that T is true. This is why, when we establish that T $\vdash \varphi$, we legitimately conclude that if T is true then φ

must be true. Conversely, if we can prove, using purely logical means (that is, *no* non-logical assumptions), the *conditional* assertion that if T is true then φ is true, then T must imply φ. Indeed, it is tempting to say that T and 'T is true' are "conceptually" or "cognitively" equivalent. However, there are subtleties involved here, which I return to in Chapter 5. As a matter of fact, the metalinguistic assumption 'T is true' can, in the context of a standard *Tarskian* theory of truth, be deductively stronger than T itself. For example, the assumption that first-order Peano Arithmetic PA is *true*, along with standard Tarskian principles governing the concept of truth, permits the derivation of the consistency of PA (which is expressible as a closed formula Con_{PA} in the language of PA), while PA itself cannot, by Gödel's Second Incompleteness Theorem, derive its own consistency.

This kind of phenomenon has no importance here. We have a theory T and we know that T implies φ. We conclude, by soundness, that if T is true, then φ is true, and since φ in this case is the assertion 'there are mathematicalia', we infer that if T is true, then there are mathematicalia.

Call this true mathematicized scientific theory T. Next we isolate some specific premises that follow from *Analysis*. These other assumptions incorporate the Tarskian and Quinian components of the argument. We have the assumptions:

Science:	T is true and its regimentation T_{reg} is an EMT
Analysis$_1$	If T is true then its regimentation T_{reg} is true
Basic Metatheorem	if T_{reg} is an EMT then $T_{reg} \vdash \exists x Math(x)$
Analysis$_2$	if $T_{reg} \vdash \exists x Math(x)$, then if T_{reg} is *true* then *there exist* entities which *satisfy* *Math*
Analysis$_3$	Any entity which *satisfies* *Math* is a mathematical entity

Tarski's semantic theory of truth implies *Analysis*$_2$. As Quine later stressed, it connects the *truth* of any theory T having an certain existential implication $\exists x P(x)$ with the *existence* of entities *satisfying* the predicate *P*. Thus, if a theory T implies an existential quantification $\exists x Math(x)$, then if T is true then there must exist entities which satisfy *Math*. Since this predicate *Math* is introduced to abbreviate '*x* is a mathematical entity', it is a disquotational triviality to say that *Math* applies to mathematical entities. That is, the triviality is:

'*x* is a mathematical entity' is true of all, and only, mathematical entities.

Thus, $Analysis_3$ is a *trivial* theorem of disquotational semantics. (The air of triviality, or perhaps even *analyticity*, of disquotational axioms for theories of truth and satisfaction is examined more closely in Chapter 5 below. It can be shown that disquotation axioms are conservative).

The philosophical argument (still a deductive derivation) then proceeds as follows:

From *Science*, $Analysis_1$ and the *Basic Metatheorem*, we derive

(1) $T_{reg} \vdash \exists x Math(x)$

Hence, using $Analysis_2$,

(2) If T_{reg} is true then there exist entities which satisfy *Math*

But, from *Science* again,

(3) T_{reg} is true

From (2) and (3),

(4) There exist entities which satisfy *Math*

Finally, using $Analysis_3$,

(5) There exist mathematicalia

To summarize:

i. if there is a true scientific statement whose standard regimentation is an EMT; and

ii. the correct notion of truth for these regimentations is Tarskian; and

iii. the correct ontological analysis for these regimentations is Quinian,

iv. then there must exist mathematicalia.

Schematically, we have:

Science + ***Semantical Analysis***

implies

Mathematical Platonism

This then is our finished product. The existence of mathematicalia just follows from science (or its truth: almost the same thing[57]) plus plausible assumptions concerning how to semantically analyse theories. The existence of mathematicalia may seem an *unacceptable* conclusion, especially if one thinks that abstract entities represent some kind of sin against our "intuitions of reality". But it seems to me that if one wishes to deny the existence of mathematicalia, one must deny at least one of these premises. That is, one must argue that either the whole enterprise of developing mathematicized science must be wrong (!!) or the standard methods of logico-semantical analysis must be wrong (!!).

The assumptions built into *Analysis* are standard technical assumptions within applied modern logico-semantic theory and are very difficult to combat. Of course, scientists do not speak or write in formalized languages, like our standard two-sorted language or L_{Math}. But there are *canonical* ways of regimenting numerous scientific laws, especially the mathematicized ones of physics (mechanics, etc.). The further assumptions involve nothing more than platitudes from standard Tarskian truth theory (and, in a sense, Quine's analysis of ontological commitment is built into Tarski's theory of truth). These

[57] Modulo what I shall say below in Chapter 5 about deflationism about truth.

accounts of truth and ontological commitment are widely, almost universally, accepted by philosophers and logicians.

In the next section we collect together some obvious and important corollaries of this argument.

3.6 Corollaries

Corollary 1: Mathematics Entails Platonism

Quine clarified what it means for discourse to be ontologically committed to entities, obtaining the result that classical mathematics is committed to the existence of mathematicalia. That is, if the classical theorem '$\exists x$(x is a prime number greater than a million)' is *true*, then some value of 'x' must be a prime number. Indeed, this corollary is the first application to which the criterion of ontological commitment was put by Quine 1948:

> Classical mathematics, as the example of primes larger than a million clearly illustrates, is up to its neck in commitments to an ontology of abstract entities. Thus it is that the great mediaeval controversy over universals has flared up anew in the modern philosophy of mathematics. The issue is clearer now than of old, because we now have a more explicit standard whereby to decide what ontology a given theory or form of discourse is committed to.
>
> (Quine 1948 (1980), p. 13).

The point is not lost on coherent nominalists. It is summed up by Field thus:

> ... after all, *the existence of mathematical entities follows from the mathematical theory itself*, not just from the claim that the mathematical theory is true in the correspondence sense.
>
> (Field 1988 (1989), pp. 249-250. Emphasis added).

Some nominalists appear to dispute this obvious logical fact. I argue in Chapter 6 that if a nominalist denies that the existence of mathematicalia *follows* from mathematics, then he or she must be engaged in some kind of contradiction. What this kind of nominalist should say is better thought of as an example of "hermeneuticism". Namely, although the existence of mathematicalia follows from certain mathematical statements when *literally construed*, it does not follow from some "reconstrual".

For example, the existence of dogs follows logically from the assertion,

(1) There is a dog in front of me,

But the existence of dogs need not follow from some positivistic "reconstrual" of (1), say,

(2) An experience similar to previous "doggish" experiences is presently occurring in my mind.

Hermeneutic reconstruals of mathematics generate an important type of nominalist strategy, of which perhaps the most significant is Chihara's modal nominalism. I discuss the theories of "reconstrual" and their application for nominalistic interpretations of mathematics in Chapters 6, 7 and 8.

Corollary 2: Scientific Realism Entails Mathematical Realism

A scientific realist is a person who toes what Fine calls the "homely line" or the "core position":

> Then, it seems to me that both the realist and the anti-realist must toe what I have been calling the 'homely line'. That is, they must both accept the certified results of science as on a par with more homely and familiarly supported claims. ... Let us say, then, that both realist and anti-realist *accept the results of scientific investigations as 'true', on a par with more homely truths* (I realize that some anti-realists would rather use a different word, but no matter). And call this acceptance of scientific truths the 'core position'.
>
> (Fine 1984 (1996), p. 36. Emphasis added).

Actually, Fine goes on to argue that this core position is *neutral* between realism and anti-realism, and labels this position the "Natural Ontological Attitude", or "NOA". However, his further remarks about truth, reference and ontological commitment make this "neutralist" claim implausible:

> ... When NOA counsels us to accept the results of science as true, I take it that we are to treat truth in the usual referential way, so that a sentence (or statement) is true just in case the entities referred to stand in the referred to relations. Thus, NOA sanctions ordinary referential semantics, and *commits us, via truth, to the existence of the individuals, properties, relations, processes and so forth referred to by the scientific statements that we accept as true.*
>
> (Fine 1984 (1996), p. 38)

I cannot think of a more succinct statement of *standard realism*. (Perhaps: "we should accept certified scientific theories as true and we should treat truth in the standard Tarski manner"). In any case, there is nothing to please the anti-realist, sceptic or idealist in Fine's core position. For example, the *truth theory* (Tarski's) is inconsistent with relativism. And the proposal to *accept as true* standard certified scientific results is inconsistent with scepticism (e.g., van Fraassen 1976, 1980 and Laudan 1981) or instrumentalism (e.g., Cartwright 1983).

Musgrave agrees with this estimation of Fine's NOA:

> Fine thinks that NOA is a minimalist view which is neither realist not anti-realist. I think that NOA is a thoroughly realist view.
>
> (Musgrave 1989 (1996), p. 45)

Thus, a scientific realist believes or accepts some scientific theories (not all, obviously, but some: namely, those *certified as acceptable* by working scientists, like Chomsky, Penrose, Weinberg, Dawkins, Pinker, etc.). For example, a realist about General Relativity believes that space-time is a four-dimensional continuum *M*, at each point of which there exist a metric tensor g_{ab} and a matter tensor T_{ab} which jointly satisfy Einstein's Field Equations.

But numerous such theories contain mathematics as an integral part and logically imply the existence of mathematicalia. The argument that we have been considering clearly implies that any such scientific realist should likewise be a mathematical realist, and accept an ontology of mathematicalia:

> ... mathematics and physics are integrated in such a way that it is not possible to be a realist with respect to physical theory and a nominalist with respect to mathematical theory.
>
> (Putnam 1975 (1979), p. 74).

Of course, a realist with nominalist intuitions may at this point change his or her mind and become a "semi-realist": he or she renounces realism about (standard, accepted mathematicized) science, and adopts realism about some (promised) *mathematics-free replacement* for platonistic science. But the onus is clearly on him or her to actually show that such a mathematics-free replacement *exists* and is, indeed, a "good enough" replacement for standard platonistic science.

Corollary 3: Metalogic Entails Platonism

We sketched this corollary in Chapter 1, where we discussed the standard platonistic analyses of satisfiability and derivability. Standard metalogic has two components:

i. *Syntactical* (or Proof-Theoretic) component:

this deals with which formulas can (and cannot) be *derived* from which (sets of) formulas, using specified derivation rules;

ii. *Semantical* (or Model-Theoretic) Component:

this deals with which formulas (or sets of formulas) are, or are not, *satisfied* by various structures or interpretations.

According to the Quinian arguments presented thus far, standard metalogic is *doubly* platonistic. The syntactical component is platonistic through quantification over expression types (that is, sequences of symbol types). The semantics is similarly platonistic through quantification over structures, interpretations, sequences and valuations.

Corollary 4: Formalism Entails Platonism

The formalist refuses to accept (as true) object level talk about mathematicalia. Instead, the formalist semantically ascends to metalinguistic talk about *mathematical assertions* (the symbols and concatenations thereof): that is, to metamathematics. And, according to the formalist, *nothing more* need be said about mathematical statements than,

a. what derivation rules govern their manipulation,

and, of course, having adopted some system of such formal rules,

b. what results about derivability then follow.

In particular, having thus semantically ascended, the formalist refuses to attribute the standard *truth values* to these assertions (i.e., he or she does not permit the usual method of disquotational descent using the truth predicate, returning to object level assertions). Of course, the naïve formalist has no *explanation* for the mysterious fact that these meaningless symbol combinations play an integral part within science.

But suppose, *per impossibile*, that mathematics had never been applied. If metalinguistic formalism were to count as a version of nominalism, that is, a repudiation of abstracta, one would still need to be clear as to what statements and symbols are. Are statements and symbols concreta? But we saw above, however, that statements, symbols and expressions are types, and are thus abstracta. One naturally wonders whether the underlying syntactical theory of expressions could be reformulated as a kind of "nominalistic syntax", whose laws only quantified over concrete tokens. One wonders then whether a formalism of tokens is adequate to formalistic metamathematics itself. This is extremely dubious.

Quine and Goodman explored the limits of nominalistic syntax in 1947. In his letter to Goodman in 1958, Alonzo Church provided a lengthy list of the basic metatheorems of standard platonistic proof theory, emphasizing plain fact that they are not accounted for in the nominalistic syntax thus developed.

Hilbert's formalism contains the explicit suggestion of a non-literal interpretation of arithmetic in which numerals (or "strokes") refer to numerals. This might be looked upon as an ontological reduction of numbers to numerals. Technically, this involves finding a translation function—a *relative interpretation*—which maps every arithmetical axiom (and thus theorem) to a theorem of nominalistic syntax. We shall call any such relative interpretation a *token-interpretation* of arithmetic.

However, a token interpretation of arithmetic does not work. For arithmetic asserts that there are infinitely many natural numbers. But if numerals are "construed" as physical tokens, there will not be enough to identify with each distinct number. So the token-interpretation is not sound: some arithmetical theorems are translated as *falsehoods* about tokens: indeed, we expect on these grounds that the (translation of the) first-order induction schema will have false instances.

Thus, the ontological reduction of numbers to tokens does not work, as Quine readily conceded: "a formalism of tokens ... stopped short of full proof theory"[58]. This, remember, was precisely Church's objection to Quine & Goodman 1947. As we mentioned before, a "solution" to this is to adopt *modalism*, and construct a theory of what concrete tokens there might have been. But nominalism, as traditionally construed, ought to eschew modal talk of what *might have been the case* or talk of *unactualized possibilia*.

A *type-interpretation* of arithmetic, however, can be made to work:

i. numbers are identified with numeral *types*;

ii. addition is interpreted as (abstract) *concatenation*;

iii. 'larger than' is interpreted as 'is a longer concatenation than'.

In fact, Hilbert's notion of a syntactical interpretation of arithmetic undoubtedly played a suggestive role in stimulating Gödel's idea of gödel-numbering (that is, the reverse: an arithmetical interpretation of syntax).

The type-interpretation is undoubtedly sound, but only in *platonistic syntax*. Thus, it is inconsistent with nominalism. For types are themselves abstracta (construable as sequences of sets of equiform tokens). Abstract concatenation of types simply is not physical juxtaposition (it may, however, be adequately modelled as the set-theoretical operation of lengthening a sequence).

3.7 The Plight of the Nominalist

The nominalist must face the Quine-Putnam argument, and each of its corollaries, honestly without what Field calls "doublethink". The coherent nominalist cannot consistently accept any of the following:

---

[58] See Quine 1981c.

i. Classical Mathematics:

Classical mathematics is trivially committed, as Quine and others have stressed, to an ontology of mathematicalia;

ii. Science:

Many accepted scientific theories (as Quine, Putnam and Field and others have stressed) include mathematical assumptions about real numbers and perhaps other mathematicalia (vectors, charts, etc.) and are thus committed to an ontology of mathematicalia;

iii. Metalogic:

Standard metalogic assumes a platonistic ambient proof theory (quantifying over expression types, arbitrarily long derivations, etc.) and a platonistic ambient semantics (quantifying over structures, sequences, valuations, etc.);

iv. Metamathematical Formalism:

Even standard formalism requires a *platonistic* proof-theory, if one wishes to preserve standard metamathematical theorems about completeness, and so on.

The nominalist has an ontology of concreta—material individuals (and perhaps their mereological aggregates), perhaps spatio-temporal geometricalia (like physical points, lines, regions), and perhaps physical events (construed as concrete physical particulars occurring in concrete spatio-temporal regions)—and nothing else to work with: no *sets*, *functions*, *expression types*, *sequences*, *structures*, *states*, *quantities*, *fields*, etc. For all of these entities are abstracta (and may be construed as mathematicalia).

I am not suggesting that a workable, serious, and perhaps even rationally compelling, nominalism that eschews sets, functions, expressions, fields, etc., *cannot* be obtained. I am simply pointing out the monumental non-triviality of the task. Mathematical nominalism, if feasible at all, is only feasible after a great deal of work has been done.

3.8 Summary: Properties of Mathematicized Theories

From a philosophical point of view, a mathematicized theory T of Nature has at least six crucial properties, which become apparent when T is formalized in the manner sketched in Chapter 2. Indeed, each of these properties is illustrated by the simplified theory of time, Tim.

i. Mathematicality

This is trivial, but is worth emphasizing. A typical mathematicized theory of Nature T contains axioms (pure and mixed) containing bound variables ranging over mathematicalia (e.g., real numbers, sets, and so on).

ii. Platonistic

Again, rather trivially, $T \vdash \exists x Math(x)$. Hence, T is ontologically committed to mathematicalia. Note that theories like Tim and AEG do not just imply the existence of *one* mathematical entity: they imply the existence of *continuum many* real numbers! (It is not clear to me whether it actually makes sense to assert the existence of just one real number, but not the others).

Indeed, it is these two facts (i) and (ii) that illustrate the Quinian component (Thesis A: "science entails platonism") of the overall Quine-Putnam argument.

iii. Impurity

Typically, the laws of mathematical physics quantify over *both* mathematicalia and non-mathematicalia. Theories like Tim and AEG (analytic geometry) discussed in Chapter 2 illustrate this fact clearly: the mixed axioms of these mathematical theories are impure, expressing relations between concreta and mathematicalia.

iv. Contingency

One version of "mathematical truth" identifies it as *theoremhood* in an uninterpreted pure axiom system. This view is also called "deductivism". However, a mathematicized theory like Tim or AEG does not follow from any system of applicable mathematical axioms (say ZFU, with < suitably interpreted). In particular, the mixed (or representation) axioms are not theorems of mathematical analysis.

Moreover, the mixed (or representation) mathematical axioms of T are contingent. E.g., it is only contingently true (if it is true at all) that there is bijection from temporal instants to **R**. Even if mathematicalia have the mathematical properties they do by necessity, they do not necessarily have the relations to concreta posited by T. It is not a necessary trait of **R** that it is isomorphic to the set of temporal instants.

Facts (iii) and (iv) are very important. Assume the pure primitives of analysis (<, etc.) have been interpreted in the language of set theory (so the pure axioms of analysis come out as ZF theorems). Then there are interpretations of L which are models of the pure axioms (i.e., the axioms of applicable set theory ZFU) in which T is *false*, because the mixed axioms are. To put this in an enlightening way, it isn't "mathematically true" (although it is true) that the "instants" in real physical time are isomorphic to the real numbers.

v. Empirical Testability

Mathematicized theories like Tim and AEG are testable, falsifiable empirical theories. They are used implicitly in physics (e.g., space-time theory). In order to test them one derives non-mathematical (synthetic) theorems and attempts to determine their truth values, by empirical methods. It may not be *easy* to test the theories Tim and AEG (except perhaps by against our "intuitions" of space and time). At the very least, as Duhem and Quine stressed, one must add auxiliary assumptions before empirical consequences emerge. They become testable by adding further axioms linking the concreta (temporal instants, spatial points) to other more easily observable concreta (like moving bodies). Indeed, as a result of this, Euclidean geometry has been abandoned (if straight lines are light paths, then one can show that the geometry of light paths in the vicinity of large masses is non-Euclidean, as Einstein predicted). The continuity of time, however, as embodied in Tim, is still thought to be true.

vi. Simple Indispensability

The *primary restriction* T° of a two-sorted theory T is the theory whose axioms are just the primary axioms (non-mathematical axioms) of T. To see that (vi) is true with regard to Tim notice that the primary restriction Tim° of Tim is just the empty set, ∅! However, Tim implies:

$$\forall x \forall y \exists z(Before(x, y) \rightarrow Before(x, z) \wedge Before(z, y))$$

[that is, z lies temporally between x and y]

This theorem of Tim is expressed in the primary notation L° but is not a consequence of the empty set. So Tim is not conservative over Tim°.

Putnam's indispensability claim (Thesis B) is connected to fact (vi): one cannot naïvely *replace* Tim by Tim°, for the full theory Tim implies non-mathematical theorems (which we suppose are true) which Tim° does not. That is, Tim is *deductively indispensable* with respect to Tim°.

This is the precise sense in which the mathematicized parts of a scientific theory T may encode information about relations amongst concrete entities. If you like, the net information contained in T about these relations is "abstraction-laden".

CHAPTER 4

Theories of Truth

Logic, like any science, has as its business, the pursuit of truth. What are true are certain statements; and the pursuit of truth is the endeavour to sort out the true statements from the others, which are false.

W.V. Quine 1950 (1972), *Methods of Logic*, p. 1

There are philosophers[59] who stoutly maintain that 'true' said of logical and mathematical laws and 'true' said of weather predictions and suspects' confessions are *two usages of an ambiguous term* 'true'. There are philosophers who stoutly maintain 'exists' said of numbers, classes and the like and 'exists' said of material objects are two *usages of an ambiguous term* 'exists'. What mainly baffles me is the stoutness of their maintenance. *What can they possibly count as evidence*? Why not view 'true' as unambiguous but very general, and recognize the difference merely between logical laws and confessions? And correspondingly for existence?

W.V. Quine 1960, *Word and Object*, p. 131. Emphasis added.

4.1 The Role of Truth in the Quine-Putnam Argument

The concept of truth plays a role in the Quine-Putnam argument reconstructed in Chapter 3 because it is the central concept employed in semantico-ontological analysis of mathematicized scientific theories. The most obvious way to see this involves the observation that this argument involves the claim:

(1) any existential mathematicized theory (EMT) logically implies the existence of mathematicalia.

This claim amounts to the "Quine-Putnam conditional":

(2) if an EMT is true, then there exist mathematicalia.

---

[59] As examples of such philosophers, Quine cites Ryle 1949, p. 29; and Russell 1912, Chapter IX.

Demonstrating this involved showing that any such theory T (or rather, its standard formulation) logically implies the existential quantification $\exists x Math(x)$, and this latter statement is true only if there are mathematicalia.

Two possible objections to the Quine-Putnam argument then immediately arise:

i. Is the standard formulation of any such theory T the correct formulation of T?

ii. Does the concept of truth, in application to mathematicized scientific theories, behave like this?

A subsidiary objection, not to the claim (2) as such, but to its applicability, questions the antecedent:

iii. Are we entitled to suppose that any such theory really is true?

This objection involves a general scepticism about scientific theories, and involves an argument that it is never rationally permissible to accept the truth of a scientific theory. A position like this is known as *epistemic scepticism* (or *epistemic anti-realism*) and has recently been advocated by, amongst others, van Fraassen 1976, 1980 and Laudan 1981. This position will receive a short and unfavourable evaluation in Chapter 7, where it is discussed under the title of "instrumentalist nominalism".

Objection (i) focuses on the *formulation* of mathematicized theories. The suggestion is that it might be possible to reformulate, or reconstrue, such theories so that they no longer imply the existence of mathematicalia. Such suggestions receive discussion later in Chapters 6 - 9 where the attempt to reconstrue such theories is discussed under the title of "hermeneutic nominalism" and the attempt to reformulate such theories non-mathematically is discussed under the title of "deflationism".

So the remaining question concerns (ii): does the concept of truth really generate ontological consequences, like that of the existence of mathematicalia? To cut a long story short, the answer is yes. The axioms of any standard modern account of truth, including Tarski's theory and the deflationary theories of truth (see below), *imply* that if an existential quantification $\exists x P(x)$ is true, then there must exist entities which satisfy P. Even the axioms of a very weak theory of truth, which I shall call the "disquotation theory" DT, imply all the individual cases of this, such as,

if $\exists x(Set(x))$ is true then there are sets,

if $\exists x(Real\ Number(x))$ is true then there are real numbers,

etc.

(Indeed, the axioms of this theory of truth just are all the individuals cases!)

So the short story is that anyone who thinks that an existential mathematicized theory T is *true* must also think that there exist mathematicalia.

What, then, is the *long story* about truth?

4.2 Theories of Truth

By the early years of the twentieth century, philosophers of opposing schools (primarily, idealism and realism) found themselves advocating one of two quite distinct kinds of theory of truth.

i. Subjectivist (or Internalist or Anti-Realist or Epistemic) Theories of Truth

Such theories of truth are favoured by idealists, empiricists and pragmatists, and maintain that:

Truth is an "*internal*" *property* of a *system of beliefs* (or sentences).

Several variations on this theme of "truth as an internal property of representations" have been advocated:

i. the *coherence theory* (where coherence of a set of beliefs is a property presumably stronger than mere consistency),

ii. the *pragmatist theory* (a body of statements is true just in case it is *useful* to believe it),

iii. the *self-evidence theory* (some statements are true because they are self-evidently so).

Musgrave describes such theories of truth in a recent article:

> Anti-realist theories of truth identify it with *some internal feature of our beliefs* (their coherence, their usefulness, their self-evidence, their ultimate undisbelievability, or whatever)
>
> . . . Down the ages the chief motive for anti-realist truth theories has been anti-sceptical, to make *truth accessible to us*. By identifying truth with some internal features of beliefs, they make it something the believer is an authority on and can know for certain. Hence they are all called *subjective* truth theories
>
> (Musgrave 1989 (1996), p. 51)

Then there are the objectivist (externalist or realist) truth theories:

ii. Objectivist or Externalist or Realist Theories of Truth

Such theories are favoured by realists and maintain that:

Truth is (or involves) an *external relation* to (typically, *mind-independent*) *reality.*

The canonical example is the *correspondence theory*, articulated by Russell in the early years of the twentieth century. A statement or belief is true just in case it corresponds to a fact:

> When a belief is true, there is *another complex unity*, in which the relation which was one of the objects of the belief relates the other objects . . . On the other hand, when a belief is false, there is no such complex unity composed only of the objects of the belief . . . Thus, *a belief is true when it corresponds to a certain associated complex*, and false when it does not . . .
>
> This complex unity is called the *fact corresponding to the belief.*
>
> (Russell 1912, pp. 128-129. Emphasis added)

Very few philosophers nowadays seriously advocate a subjectivist theory of truth[60]. Indeed, there are several basic problems with subjectivist theories of truth. Primarily, it is possible to give examples of systems of statements or beliefs which first have the internal features (for a given thinker or community of such), that is, the system *is* either coherent, useful, or has actually been believed and which are straightforwardly false. For example, Newtonian Mechanics (henceforth, NM) is presumably coherent, was believed for a long time by generations of scientists, and remains eminently useful, but is not true. Unless one thinks that truth is *relative* somehow (so that NM was true *then*, or true *for those thinkers*, but is not true *now* or not true *for us*), a subjectivist theory of truth seems untenable.

Musgrave supplies a damning criticism of subjectivism about truth:

> An immediate consequence [of internalist/subjectivist truth theories] is *relativism about truth*. If A's belief that *P* possesses the internal feature and B's belief that *P* does not, then *P* is true for A and false for B. (Alternatively put, the 'laws of truth', contradiction and excluded middle, etc., fail).
>
> To avoid relativism and preserve the laws of truth, subjective truth theorists tend to go *ideal* and *to the long run*. For example, coherence theorists will say that something is true (by definition) if it 'coheres', not with your or my beliefs, but with the beliefs an *ideally rational inquirer* would have *in the long run*.
>
> Of course, such moves immediately threaten the anti-sceptical virtues (if virtues they be) of subjective truth theories. What God will be coherently believing at the end of time is just as inaccessible to you or me as is truth-as-correspondence (I think it is *more* inaccessible).
>
> To solve these problems, subjectivists tend to adopt a policy of *flipping back and forth*. When epistemological concerns are paramount, they stay subjective; when semantic concerns are paramount, they go ideal and to the long run. The resulting fandango is one of the least edifying sights in philosophy. Let us view it no more.
>
> (Musgrave 1989 (1996), p. 52, footnote. Emphasis added)

I shall now leave subjectivist theories of truth. Serious philosophical discussion about truth focuses on objectivist theories, of which there are roughly three kinds:

[60] As ever, there are exceptions to this philosophical consensus. Putnam 1981 and Ellis 1985 both advocate something they call "Internal Realism", according to which, truth is the perfected end product of the application of rational scientific method. (The truth is what rational humans shall be believing "in the end").

i. *The Correspondence Theory*[61],

ii. *Tarski's Theory of Truth,*

iii. *Deflationary Theories of Truth.*

We discuss Tarski's theory and deflationary theories more fully below (and in the next Chapter). It is easy to see that all of these truth theories imply the Quine-Putnam conditional (2) above. So it seems that any serious conception of truth under discussion will entail the *logical link* between the truth of standardly formulated scientific theories and the existence of mathematicalia.

For example, according to a correspondence theory, mathematical theories describe a realm of abstract mathematical facts, such as the (mind-independent) fact that no non-vanishing continuous vector field exists on the 2-sphere $\mathbf{S}^2$. According to Tarski's theory, an interpreted statement is true just in case it is satisfied by all sequences of objects. In particular, it can be proved within the truth theory that the statement 'there are $\aleph_0$ primes' is satisfied by all sequences (and is thus true) just in case there are infinitely many primes. Supposing the statement true, there must be infinitely many primes; and thus, since prime numbers are mathematicalia, there exist mathematicalia. Finally, according to the deflationary conception of truth, it is (analytically?) true that the statement 'there are $\aleph_0$ primes' is true just in case there are infinitely many primes. And the same preceding argument goes through.

4.3 Constituting the Alethic Concepts: T-Sentences

Russell's formulation of the correspondence theory requires that

(1) a sentence (belief, proposition) is true just in case it *corresponds* to a *fact*.

For example, the sentence 'Madrid is larger than Birmingham' is true. According to the correspondence theory, it is true *because* a fact "in the world"—the fact *that Madrid is*

[61] Kirkham 1992 (1995) and David 1994 both discuss the correspondence theory. I lack the space to discuss any of these issues.

larger than Birmingham—makes it so. In general, a sentence "corresponds" to a "state of affairs", which may either "hold" or not. Three problems arise immediately: what is a *state of affairs*? How do sentences *correspond* to such states of affairs? What is it for a state of affairs to *hold*? Some authors identify "states of affairs" with *propositions* (or conversely) and thus claim that a sentence corresponds to the proposition it expresses; a fact is then just a *true proposition*; finally a sentence is true just in case any proposition it corresponds to (expresses) is a fact[62]. In any case, the basic idea behind the correspondence theory is that it is a *fact* that *makes* a sentence or belief *true*. (A further realist notion is that this fact is typically mind-independent and language-independent).

An example suffices to show that the correspondence account can be significantly trimmed (some would say: "deflated" of naïve correspondence intuitions). Suppose we are wondering what would make a particular proposition—say, the proposition *that snow is white*—true. We instantiate the correspondence definition as follows:

(2) the proposition *that snow is white* is true just in case the proposition *that snow is white* corresponds to a fact.

But presumably,

(3) the proposition *that snow is white* corresponds to a fact just in case snow is white.

The proposition is made true by the "whiteness of snow", as it were.

A series of major thinkers noticed this connection: Frege 1892, Ramsey 1927 and finally (and most significantly) Tarski 1936. So,

(4) the proposition *that snow is white* is true if and only if snow is white

[62] On this approach, a syntactically well-formed sentence may not express any proposition, for logical reasons. It can be argued, for example, that the Liar Sentence "This sentence is not true" does not express a proposition, and in fact the Liar *Proposition* does not exist. If it did, it would be true if and only if not true. So it doesn't (c.f., Russell's barber who shaves all and only those who do not shave themselves).

This latter statement avoids talk of "correspondence" and "facts" and seems to be a perfectly acceptable (indeed, arguably *analytic*) fact about the truth conditions of the proposition in question. What could be simpler?

A similar piece of reasoning works as nicely for the sentence 'snow is white', yielding,

(5) the sentence 'snow is white' is true if and only if snow is white

One may attempt to schematize (4) as follows:

RS the proposition that *p* is true if and only if *p*

And one may similarly attempt to schematize (5) as follows:

DS the sentence '*p*' is true if and only if *p*

These schemes, RS (the "redundancy scheme") and DS (the "disquotation scheme") lie at the heart of all serious recent attempts to get to grips with the concept of truth. For example, the "totality" of statements like (4), all instances of RS, form the core of a certain kind of *deflationary theory of truth*: the *redundancy* or *minimalist* theory of truth. Similarly, statements like (5), all instances of DS, form the core of the *disquotational* theories of truth and, indeed, play the role of a basic *constraint* (Tarski's Convention T) on Tarski's semantical theory of truth.

Indeed, at the heart of almost all recent work on truth and reference lies a collection of *schematic principles*, which constrain and perhaps even constitute the alethic concept of truth. There are also similar schemes for names and predicates, governing the alethic concepts of designation and satisfaction respectively, as illustrated in the table below[63]:

Alethic Concept	**Scheme**	**Example**
designation, reference	'*n*' designates *n*, if *n* exists	'snow' designates snow, if snow exists
truth of, satisfaction	'*F*' is true of *x*, *y*, ... iff *Fxy*... '*F*' is satisfied by *x*, *y*, ... iff *Fxy*...	'loves' is true of *x*, *y* iff *x* loves *y* 'loves' is satisfied by *x*, *y* iff

63 This table is inspired by a similar one in Burgess & Rosen 1997, p. 57.

		x loves *y*
truth for sentences	the sentence '*p*' is true iff *p*	'snow is white' is true iff snow is white
truth for propositions	the proposition that *p* is true iff *p*	the proposition that snow is white is true iff snow is white

We may summarize the four constitutional schemes operative here as follows:

DS_{nam}: the name '*n*' designates *n* (if *n* exists)

DS_{pred}: the n-place predicate '*F*' is true of $x_1, \ldots, x_n$ iff $F(x_1, \ldots, x_n)$

DS: the sentence '*p*' is true iff *p*

RS: the proposition that *p* is true iff *p*

(As a purely terminological aside, in DS_{pred} we can replace 'is true of' by 'applies to' or 'is satisfied by', without significantly altering the discussion.)

A significant property of these schemes in discussions of semantic theory is their *prima facie obviousness* or *triviality*. For example, an instance of the disquotation scheme DS:

(6) 'snow is white' is true if and only if snow is white

is sometimes called a "*T-sentence*" and its obviousness seems beyond doubt[64].

On this topic, Davidson writes:

> ... a sentence like,
>
> 'snow is white' is true if and only if snow is white
>
> in itself provides a clue to what is unique to the semantical approach [to truth]. Of course, ..., *such sentences are neutral ground.* It is just for this reason that Tarski hopes that everyone can agree that an adequate theory or definition of truth must entail all sentences of this form.
>
> Davidson 1969 (1984), pp. 50-51. Emphasis added)
>
> Because T-sentences (as we may call them) are *so obviously true*, some philosophers have thought that the concept of truth, at least as applied to sentences, was trivial.

---

64 Anyone who has taught philosophy of logic to undergraduate students quickly learns that trying to convince students that these obvious "trivialities" are *important* is no mean task.

(Davidson 1973 (1984), p. 65. Emphasis added)

> The reason Convention T [see below] is acceptable as a criterion of theories [of truth] is that (1) *T-sentences are clearly true (pre-analytically)*—something we could recognize only if we already partly understood the predicate 'is true' ...
>
> (Davidson 1977 (1984), p. 218. Emphasis added)

However, we ought not to be too easily charmed by the obviousness of disquotational T-sentences. If we are not careful, they lead to the *semantical paradoxes*. The quickest method not involving indexicality (as in the usual Liar Sentence, 'This sentence is false') is the Grelling-Nelson Paradox of "heterological words".

A word or phrase is *autological* if it is true of itself. For example, 'word' is true of words; but 'word' is a word, and is thus autological. In contrast, 'red' is not red and 'herring' is not a herring, so neither 'red' nor 'herring' are autological. Another example is 'contains twenty five letters', which contains twenty five letters and is thus autological. Now, let a word or descriptive phrase be *heterological* if and only if it is *not* true of itself. So, 'red' and 'herring' are both heterological. Now, instantiate the scheme DS_{pred} with 'heterological':

(7) 'heterological' is true of x if and only if x is heterological.

So, in particular,

(8) 'heterological' is true of 'heterological' if and only if 'heterological' is heterological.

Thus, using the definition of 'heterological':

(9) 'heterological' is true of itself if and only if 'heterological' is not true of itself.

This is a contradiction.

In general, such paradoxes always arise when we predicate semantic concepts of phrases themselves containing semantical expressions. A simple evasion of the paradoxes is simply to avoid instantiating the schemes with expressions containing 'true' or 'true of' (or cognates). (Indeed, with this restriction in place, it is possible to give a simple *consistency proof* for the "totality" of such schemes: see next Chapter).

To summarize, there seems to be no question that semantical statements like,

(4) the proposition that snow is white is true if and only if snow is white

(5) the sentence 'snow is white' is true if and only if snow is white

(10) 'white' is true of x if and only if x is white

are, in some striking sense, obvious. While the semantic paradoxes alert us to the fact that we ought not to suppose *all* instances of the schemes "obvious", we can still ask *why* (suitably restricted) instances of disquotational and redundancy T-sentences are obvious. And in what does their obviousness consist? Are such statements perhaps analytic or conceptual truths? Does their truth derive solely from the meanings of the semantical words 'designate', 'denotes' and 'true' (or the other cognates for these concepts)? Are these instances *a priori*? Are they necessary in some sense? I shall return to these questions briefly in Chapter 5.

Associated with the schemes RS and DS are *Equivalence Principles*:

$\varphi \Leftrightarrow \ulcorner$the proposition that φ is true$\urcorner$

$\varphi \Leftrightarrow \ulcorner$the sentence '$\varphi$' is true$\urcorner$

For example,

'grass is blue' $\Leftrightarrow$ 'the proposition *that grass is blue* is true'

'grass is blue' $\Leftrightarrow$ 'the sentence 'grass is blue' is true'

One of the earliest invocations of this Equivalence Principle appears in Frege:

> One can indeed say 'The thought that 5 is a prime number is true'. But close examination shows that *nothing more has been said* than in the simple sentence '5 is a prime number'.
>
> (Frege 1892 (1980), p. 34. Emphasis added)

It is this claim, that to assert 'The proposition that p is true' is to assert *nothing more* than 'p', that underlies the deflationist view of truth, which I discuss below.

4.4 Redundancy, Disquotation and Deflationism

Deflationism suggests that the concept of truth is *redundant* or *dispensable.* The idea is that we need to "deflate" philosophical problems connected with truth (e.g., the correspondence intuition: truth involves a substantial language-world relation). The central idea of deflationism is that the Equivalence Principles and/or the equivalence schemes RS and DS express all there is to be said about truth.

The first such account of truth, the so-called *Redundancy Theory of Truth*, was first urged by Ramsey 1927, and later by Ayer 1936:

> There is really no separate problem of truth, merely a linguistic muddle.
>
> ... it is evident that 'It is true that Caesar was murdered' means no more than that Caesar was murdered, and that 'It is false that Caesar was murdered' means that Caesar was not murdered. They are phrases that we sometimes use for emphasis or for stylistic reasons, or to indicate the position occupied by the statement in our argument.
>
> (Ramsey 1927 (1978), p. 44).

> But when we come to consider what this famous question ["What is truth?"] actually entails, we find that it is not a question which gives rise to any genuine problem; and consequently that no theory can be required to deal with it.
>
> ... Reverting to the analysis of truth, we find that in all sentences of the form '*p* is true', the phrase 'is true' is logically superfluous. When, for example, one says that the proposition 'Queen Anne is dead' is true, all that one is saying is that Queen Anne is dead. ... Thus to say that a proposition is true is just to assert it, and to say that it is false is just to assert its contradictory. And this indicates that the terms 'true' and 'false' connote nothing, but function in the sentence simply as marks of assertion and denial.
>
> ... We conclude, then, that there is no problem of truth as it is ordinarily conceived. The traditional conception of truth as a 'real quality' or a 'real relation' is due, like most philosophical mistakes, to a failure to analyse sentences correctly.
>
> (Ayer 1936 (1971), pp. 116-119)[65].

[65] Shortly after Ramsey's and Ayer's advocacy of deflationism, however, Tarski presented his *magnum opus* on truth, Tarski 1936, in which he proved his *Indefinability Theorem*. One way to put this is to consider an interpreted language (L, A), where A is an L-structure. Then, truth for (L, A), i.e., the set of truths in A, is not *definable* in the theory $\mathsf{Th}_L(A)$, if the theory $\mathsf{Th}_L(A)$ is "strong enough". See below for further discussion. This in itself suggests that truth is not dispensable.

More recently, Horwich has defended a position closely related to Ramsey's original redundancy theory. Horwich call this the *Minimalist Theory of Truth.* Horwich claims that the totality of instances of the redundancy scheme RS yields the "whole truth about truth"[66]. Later I shall reconstruct this minimalist theory as a formalized theory. Indeed, it is the theory MT whose logical properties we discuss more fully below in the next Chapter[67].

Not all thinkers have appreciated (or agreed with) the Equivalence Principle (or the Redundancy Scheme). Consider the following passage from Spinoza's posthumous *On the Improvement of the Understanding*:

> ... *if anyone asserts*, for instance, *that Peter exists*, without knowing whether Peter really exists or not, *the assertion*, as far as its asserter is concerned, *is false*, or not true, *even though Peter actually does exist*. The assertion that Peter exists is true only with regard to him who knows for certain that Peter does exist.
>
> (Spinoza 1677 (1955), p. 26. Emphasis added).

(Seemingly, Spinoza didn't much improve the understanding of the concept of truth.[68])

The disquotation scheme DS appears first in Tarski 1936, where it is actually dismissed as slightly misleading, for problems having to do with the peculiarities of the *quotation operator*[69]:

> As a starting point certain sentences of a special kind present themselves which could serve as partial definitions of the truth of a sentence or more correctly as explanations of various concrete turns of speech of the type '*x* is a true sentence'. The general scheme of this kind can be depicted in the following way:
>
> (2) *x* is a true sentence if and only if *p*

---

[66] E.g., Horwich 1990 (p. 7), "... our thesis is that it is possible to explain *all* the facts involving truth on the basis of the minimal theory" and (p. 8), "... the minimalist *conception*: i.e., the thesis that our theory of truth should contain nothing more than instances of the equivalence scheme".

[67] To be a little more precise, Horwich thinks that MT should be thought of as the (non-linguistic) *proposition* expressed by any such totality of axioms.

[68] Spinoza is often cited as an early *coherence theorist* about truth.

[69] In particular, we cannot "quantify-in" using a referential quantifier into quotational contexts. E.g.,

$\exists x$(Quine wrote a book named '*x* and Object').

is problematic. However, careful examination shows that we can sensibly "quantify-in" using substitutional quantifiers. E.g.,

Σx(Quine wrote a book named 'x and Object').

is not problematic, and is in fact true, because 'Quine wrote a book called 'Word and Object'' is true.

> In order to obtain concrete definitions we substitute in the place of the symbol '*p*' in this scheme any sentence, and in the place of '*x*' any individual name of this sentence.
>
> (Tarski 1936 (1956), pp. 155-156).

Tarski then discusses one of his famous instances of this scheme:

> It might be thought that all we need to do is to substitute in (3):
>
> (3) 'it is snowing' is a true sentence if and only if it is snowing
>
> any sentential variable (i.e., a symbol for which any sentence may be substituted) in place of the expression 'it is snowing' which occurs there twice, and then to establish that the resulting formula holds for every value of the variable, and thus without difficulty reach a sentence which includes all assertions of type (3) as special cases:
>
> (5) for all *p*, '*p*' is a true sentence if and only if *p*
>
> ... we could try to generalize the formulation (5), for example, in the following way:
>
> (6) for all *x*, *x* is a true sentence if and only if, for a certain *p*, *x* is identical with '*p*' and *p*
>
> At first sight we should perhaps be inclined to regard (6) as a correct semantical definition of 'true sentence' ... Nevertheless the matter is not so simple. As soon as we begin to analyse the significance of the quotation-mark names which occur in (5) and (6) we encounter a series of difficulties and dangers.
>
> (Tarski 1936 (1956), p. 158-159).

It is worthwhile quickly looking at the problem Tarski located. Tarski explains some of the peculiarities of quotation-mark names:

> In any case, such functors are not extensional; there is no doubt that the sentence
>
> for all *p* and *q*, if *p* if and only if *q* then '*p*' is identical with '*q*'
>
> is in palpable contradiction to the customary way of using quotation marks. For this reason alone definition (6) would be unacceptable to anyone who wishes consistently to avoid intensional functors and is even of the opinion that a deeper analysis shows it to be impossible to give any precise meaning to such functors.
>
> (Tarski 1936 (1956), p. 161).

Actually, it seems to me that Tarski's scepticism here seems to be misplaced. I suggest that we can explain what a quotation function is as follows:

> A *quotation function* is an injective mapping *f* from a language L to another L* (*f* may be defined over any class of L-expressions, even including the comma, say) such that if ε is *any* L-expression then $f(\varepsilon)$ is an L*-term.

Thus, the mapping that maps any expression to the result of enclosing it in quotation marks is a quotation function. For example, such a function maps the sentence 'snow is white' to the term ''snow is white''. (It may map an expression to a term obtained by enclosing within French quotes '<<' and '>>', or Smullyan's quotes 'Q_1' and 'Q_2').

The *semantics* of a quotation function *f* is then explained as follows:

> An interpretation $\Im^*$ of L* is *quotation normal* (or *respects quotation*) with respect to *f* if and only if, for any L-expression ε, $\Im^*[f(\varepsilon)] = \varepsilon$.

Suppose we symbolize this operation as $<\ldots>: \varepsilon \mapsto <\varepsilon>$, where ε can be any L-expression. The expression $<\varepsilon>$ is an L*-term, perhaps formed from ε itself by prefixing and appending '<' and '>'.

Given the notion of a quotation-normal interpretation, it is easy to see then that the quotation operator $<\ldots>$ is not extensional in any such interpretation. A context $C(\varepsilon)$ of the expression ε is *extensional* with respect to an interpretation $\Im$ if and only if, for any expressions ε_1, ε_2, if $\Im[\varepsilon_1] = \Im[\varepsilon_2]$, then $\Im[C(\varepsilon_1)] = \Im[C(\varepsilon_2)]$. For example, let t_1 and t_2 be terms of a standard first-order language L and let $\varphi(x)$ be an L-formula with just x free. Suppose that $\Im$ is an L-interpretation and that $\Im[t_1] = \Im[t_2]$. It trivially follows that $\Im \models \varphi(t_1)$ iff $\Im \models \varphi(t_2)$. Thus, any formula $\varphi(t)$ is an extensional context of t, for any term t.

The requirement of extensionality is easily seen to be equivalent to the condition that the extension of $C(\varepsilon)$ be *a function* of the extension of ε. (Of course, this idea goes back to Frege 1892).

Now, let $\Im^*$ be a quotation normal interpretation of L* and let ε_1 and ε_2 be L-expressions. Even if ε_1 and ε_2 are *co-extensive* (that is, $\Im[\varepsilon_1] = \Im[\varepsilon_2]$), it need not follow that $\Im^*[<\varepsilon_1>] = \Im^*[<\varepsilon_2>]$. Indeed, if $\Im^*$ is quotation normal, then $\Im^*[<\varepsilon>] = \varepsilon$. So, if there are expressions ε_1 and ε_2 such that $\Im[\varepsilon_1] = \Im[\varepsilon_2]$ and $\varepsilon_1 \neq \varepsilon_2$, then $<\ldots.>$ is not extensional (in this interpretation $\mathfrak{I}^*$). This requires only that there are distinct expressions that have the same extension. This is trivial to satisfy (e.g., '1 + 0' and '1').

Nevertheless, we have supplied an analysis of a quotation mapping. It is a syntactical mapping from expressions to *terms*. Furthermore, we have defined what it means for an interpretation to *respect quotation* and this makes good sense, even though the quotation term $f(\varepsilon)$ containing ε is not an *extensional* context of ε.[70]

Returning to our main theme, a disquotational version of deflationism about truth is usually associated with Quine and Leeds. For example:

> To say that the statement 'Brutus killed Caesar' is true, or that 'The atomic weight of sodium is 23' is true, is in effect simply to say that Brutus killed Caesar, or that the atomic weight of sodium is 23.
>
> (Quine 1960, p. 24)

> ... Where the truth predicate has its utility is in just those places where, though still concerned with reality, we are impelled by certain technical complications to mention sentences. Here the truth predicate serves, as it were, to point through the sentence to the reality; it serves as a reminder that though sentences are mentioned, reality is still the whole point.
>
> What, then, are the places where, though still concerned with unlinguistic reality, we are moved to proceed indirectly and talk of sentences. The important places of this kind are places where we are seeking generality, and seeking it along certain oblique planes that we cannot sweep out by generalizing over objects.
>
> ... The truth predicate is a reminder that, despite a technical ascent to talk of sentences, our eye is on the world. This cancellatory force of the truth predicate is explicit in Tarski's paradigm:
>
> 'Snow is white' is true if and only if snow is white
>
> Quotation marks make all the difference between talking about words and talking about snow. ... By calling the sentence true we call snow white. *The truth predicate is a device of disquotation.*
>
> (Quine 1970 (1986), pp. 10-12. Emphasis added)

[70] It would be an interesting mathematical exercise to study "quotation-normal" interpretations of the language of arithmetic, where the quotation function is just the mapping that takes ε to its gödel numeral $\ulcorner \varepsilon \urcorner$. Part of the complication is that some numerals (numbers) are not gödel numerals (numbers). It follows from this that the recursive clause for evaluating the designation of $S(t)$ will not work straightforwardly, because t may be a gödel numeral but $S(t)$ may not be.

I'm not sure it is correct to regard Quine as a deflationist (or as a deflationary disquotationalist). First, he makes it clear that he thinks that unlinguistic reality plays a role in determining the truth value of sentences,

> No sentence is true but reality makes it so. The sentence 'Snow is white' is true, as Tarski taught us, if and only if *real* snow is *really* white.
>
> (Quine 1970 (1986), p. 10. Emphasis added).

However, the deflationist need not deny this. More importantly, when it comes to *defining* truth, Quine makes it clear that a more sophisticated procedure, transcending the totality of disquotational T-sentences, is necessary:

> Tarski's paradigm cannot be generalized to read:
>
> '*p*' is true if and only if *p*
>
> since quoting '*p*' produces a name only of the sixteenth letter of the alphabet, and no generality over sentences. The truth predicate in its general use, attachable to a quantifiable variable in the fashion '*x* is true', is eliminable by no facile paradigm. It can be defined, as Tarski shows, in a devious way, but only if some powerful apparatus is available.
>
> (Quine 1970 (1986), pp. 12-13).

(I explain this more powerful apparatus, *The Theory of Satisfaction*, invoking a metalinguistic predicate 'satisfies', below).

Someone may agree with the Equivalence Principle and yet *disagree* with deflationism. Consider Putnam, who is not a deflationist:

> ... the equivalence principle is *philosophically neutral*, and so is Tarski's work. On *any* theory of truth, 'Snow is white' is equivalent to "'Snow is white' is true'.
>
> (Putnam 1981, p. 129. Emphasis added).

Disquotational accounts of truth and reference have been widely discussed in the literature. For example, Field 1986, Field 1988, and two recent books, David 1994 and Kirkham 1992 (1995). David introduces his book with the quip:

> "What is truth?", asked Pilate. "Truth is disquotation", replied Quine.
>
> (David 1994, p. 3).

Field has done much to clarify the disquotational understanding of truth:

> One way of understanding the notion of truth is "disquotational". On this reading, a sentence of the form

> '...' is true
>
> is to be understood as cognitively equivalent (equivalent by logic plus the meaning of 'true) to the sentence appearing in the blanks, the sentence of which truth was predicated.
>
> (Field 1988 (1989), pp. 228-229).

> ... disquotational truth is a notion that applies primarily to sentences of one's own language ...
>
> (Field 1988 (1989), p. 244).

The idea is this. A competent speaker S of an *interpreted* language $\mathfrak{L}$ recognizes the *equivalence* of a statement φ with its corresponding truth-claim $Tr(<\varphi>)$. Indeed, if S knows that Tr^* is a disquotational truth predicate for an *alien* interpreted language $\mathfrak{L}^*$, then S is able to recognize *that* $Tr^*(<\varphi^*>)$ is equivalent to φ^*. But if S does not understand φ^* itself, then S similarly does not understand the truth-claim $Tr^*(<\varphi^*>)$, and conversely. To put it differently, S may know that Tr^* is a truth predicate (in an appropriate extension of $\mathfrak{L}^*$) for $\mathfrak{L}^*$, but still doesn't understand the truth claim $Tr^*(<\varphi>)$ because he or she doesn't understand the quoted sentence φ^*.

In more concrete terms, I know, for any German sentence φ^* not containing 'wahr', that $wahr(<\varphi^*>)$ is equivalent to φ^*, but of course I still needn't know what φ^* means. For example, if someone tells me that 'Blanca ist sehr schön' is a German sentence, then, equipped with my knowledge that 'wahr' is the truth predicate in German, I already know that the semantical German sentence,

(1) 'Blanca ist sehr schön' ist wahr

is equivalent to,

(2) Blanca ist sehr schön

(This is an empirical claim which could be checked experimentally: a fluent German speaker will assent to, or dissent to, both (1) and (2) in precisely the same circumstances).

Also, I know that if φ^* *means the same* as some statement φ of English that I understand, then, if I stipulate that Tr_G is to mean in English "true in German", then I know that I must accept $Tr_G(<\varphi^*>) \leftrightarrow \varphi$. But Tr_G, that is, 'true-in-German', is not a *disquotational* truth predicate for German sentences, for the *trivial* reason that φ^* is not a sentence of English, so $Tr_G(<\varphi^*>) \leftrightarrow \varphi^*$ is simply not a sentence of English. (Obviously, $Tr_G(<\varphi^*>) \leftrightarrow \varphi$ is not a *disquotational* sentence of English, for it *mentions* quotationally a German sentence φ^* and *uses* an English sentence φ to "represent" the truth condition of φ^*).

To return to the main theme, according to the deflationary conception of truth, the concept of truth is something like a purely logical device, which simply provides a way, in any language L, of:

i. "Disquoting" quotations of L-expressions (and "dispropositionalizing" *that*-constructions);

For the "disquotation effect", we are allowed to eliminate the truth predicate in predications containing quotations, by inferring from a truth-predication,

(1) 'snow is white' is true

the quoted sentence itself,

(2) snow is white

In fact, one might propose the following *introduction* and *elimination rules* for *Tr*:

Tr-Intro: $\varphi \vdash Tr(<\varphi>)$

Tr-Elim: $Tr(<\varphi>) \vdash \varphi$

(In a certain respect, then, the truth predicate behaves a little like a *logical concept*, with introduction and elimination rules, analogous to those for $\neg$ and $\wedge$).

However, another logical function of the concept of truth has been discussed by truth theorists:

ii. Expressing certain infinitely long conjunctions and disjunctions (which would otherwise require a logical device of substitutional quantification).

Perhaps the first author to mention this is Quine:

> We may affirm a single sentence by just uttering it, unaided by quotation or the truth predicate; but *if we want to affirm some infinite lot of sentences that we can demarcate only by talking about the sentences, then the truth predicate has its use.*
>
> (Quine 1970 (1986), pp. 10-12. Emphasis added)

Similarly, following Quine's lead, Leeds wrote:

> It is not surprising that we should have use for a predicate P with the property that " '...' is P" and "..." are always interdeducible. For we frequently find ourselves in a position to assert each sentence in a certain infinite set *z* (e.g., when all the members of *z* share a common form); lacking the means to formulate infinite conjunctions, we find it convenient to have a single sentence which is warranted precisely when each member of *z* is warranted. A predicate P with the property described allows us to construct such as sentence: $\forall x(x \in z \rightarrow P(x))$. Truth is thus a notion that we might reasonably want to have at hand, for expressing semantic ascent and descent, infinite conjunction and disjunction.
>
> (Leeds 1978, p. 121).

Leeds follows Quine in explaining the utility of the truth predicate as a logical device for "expressing semantic ascent and descent, infinite conjunction and disjunction". Is Leeds advocating deflationism? It is not clear. He does not explicitly say that there is *nothing more* to the concept of truth than this, and could, in fact, insist that such a logical device (of disquotation) is just the same thing as correspondence.

(There is an important analogy here between the claim that truth is a logical device for expressing infinite conjunctions and the idea that quantification is a logical device for expressing infinite conjunction: that is, when ordinary objectual quantification itself is construed substitutionally, it may (in some cases) be thought of as a logical device for expressing infinite conjunctions: $\forall x F(x) \Leftrightarrow F(a_1) \wedge F(a_2) \wedge \ldots$. However, full objectual quantification cannot be so construed, unless one supposes that there is a term for every object: a fully named universe. Since we usually impose strong constructive constraints on *linguistic entities* (indeed, usually, there will be at most denumerably many terms), this substitutional reinterpretation will not be sound for non-denumerable domains of discourse: in particular, those of analysis and set theory).

For an example of the expression of "infinitary disjunctions/conjunctions", (ii), the idea is that we would in principle be allowed to eliminate the truth predicate more generally, by replacing any open formula,

(3) x is true

by the infinitary disjunction,

(4) (x = 'snow is white' and snow is white) or (x = 'grass is green' and grass is green) or ...

But because natural languages do not contain such infinitary logical apparatus, the argument is that the truth predicate "serves this purpose" of "expressing infinitary conjunctions/disjunctions".

Indeed, if one imagines that the (meta-)language in question contains a substitutional quantifier (with respect to sentences of the base language), then we may restate (4) formally as a substitutional quantification:

(5) $\Sigma p(x = \langle p \rangle \wedge p)$

Technically, (5) is to be formalized within a metalanguage L^{+}_{sub} obtained by *expanding* the object language L by adding a *quotation function* and *substitutional quantification* with respect to sentences: that is, add *sentence variables* $p_1, p_2, \ldots$, and a *substitutional quantifier* Σp, with an implicit substitution class of L-sentences. Then, the formula $\Sigma p(x = \langle p \rangle \wedge p)$ defines $Tr(x)$. I discuss some properties of this kind of truth definition in the next Chapter.

4.5 Tarski's Convention T

Tarski's theory of truth is not a correspondence theory in the traditional Russellian sense. It does not explicitly express the idea that statements are "*made true*" by the "*facts*" to which they "*correspond*". Rather, Tarski intended his semantical conception of truth to be a clarification and elucidation of the intuitive idea that:

> ... a true sentence is one which states that the state of affairs is so and so, and the state of affairs indeed is so and so.
>
> (Tarski 1936 (1956), p. 155).

Indeed, Tarski alludes approvingly to Aristotle's formulation:

> To say of what is that it is not, or of what is not, that it is, is false, while to say of what is that it is, or of what is not, that it is not, is true
>
> (Aristotle, *Metaphysics*, 1011^b26).

Those who follow Tarski's line on a semantical definition of 'true sentence' sometimes put the basic idea like this:

> ... a sentence (or statement) is true just in case the entities referred to stand in the referred-to relations.
>
> (Musgrave 1989 (1996), p.48).

Lying at the heart of Tarski's thinking about truth is Convention T, a comprehensive criterion for what counts as a *truth predicate* for an object language L. Informally speaking, we have two languages,

i. an object language L , whose semantic properties are to be discussed

ii. a metalanguage ML, in which the metatheory is formulated.

The metalanguage ML is presumed to contain predicates which express the *syntax* of L. Thus, ML may contain predicates *Pred, Sen, Var, Term*, etc., meaning intuitively "predicate in L", "sentence in L", "variable in L", "term in L", etc. In particular, we shall assume that ML contains, for each L-formula expression, a *term* <φ> which means, intuitively, "the L-expression φ".[71]

This is how Tarski presented Convention T in his 1936.

> Using the symbol '*Tr*' to denote the class of all true sentences, the above postulate can be expressed in the following convention:
>
> **CONVENTION T**. A formally correct definition of the symbol '*Tr*', formulated in the metalanguage, will be called an *adequate definition of truth* if it has the following consequences:
>
> (α) all sentences which are obtained from the expression
>
> $x \in Tr$ if and only if p
>
> by substituting for the symbol '*x*' a structural-descriptive name of any sentence of the language in question and for the symbol '*p*' the expression which forms the translation of this sentence into the metalanguage;

[71] The syntax of a language L may sometimes be interpreted within (some theory in) L itself. If L is the language of arithmetic, this is the *arithmetization of syntax*. In the case of arithmetic, this is achieved via a *coding function* #, which assigns a gödel number #φ to each L-expression φ. One can find formulas in L which represent all the syntactical concepts of L. Thus, one can find formulas in L such as *Term*(*x*), *Form*(*x*) and so on, which are true of exactly the gödel numbers of the terms, formulas and so on in L.

(β) the sentence

for any *x*, if $x \in Tr$ then $x \in S$ [the set of object language sentences]

(in other words '$Tr \subseteq S$')

It should be noted that the second part of the above convention is not essential; so long as the metalanguage already has the symbol '*Tr*' which satisfies condition (α), it is easy to define a new symbol '*Tr*'' which also satisfies condition (β).

(Tarski 1936 (1956), pp. 187-188).

Then, consider the ML scheme:

(T) x is true in L if and only if p

Tarski argued that this scheme plays a basic role in constraining any truth theory or truth definition. For any L-sentence φ, replace 'x' by an ML name of φ (say $<\varphi>$) and replace 'p' by the *translation* of φ into ML (say, $\Gamma(\varphi)$). Then the result is what is sometimes called a *materially adequate* T-sentence.

For example, suppose ML is formalized fragment of English plus a predicate 'true in German', and L is a formalized fragment of German containing the sentence 'Der Schnee ist weiss'. Then corresponding to this German sentence we have the following materially adequate T-sentence:

(1) 'Der Schnee ist weiss' is true in German if and only if snow is white

which is materially adequate because the metalanguage sentence *used* (that is, 'snow is white') is the *correct metalanguage translation* of the object language sentence mentioned or quoted (that is, 'Der Schnee ist weiss'). This dependence upon the notion of *translation* is quite vital to the *general* Tarskian semantic conception of truth, and Tarski himself made this clear in his locus classicus:

As we know from §2, to every sentence of the language of the calculus of classes there corresponds in the metalanguage not only a name of this sentence of the structural-descriptive kind, but also a sentence *having the same meaning*.

... We take the scheme (2) ['*x* is a true sentence if and only if *p*'] and replace '*x*' in it by the name of the given sentence, and '*p*' by its *translation* into the metalanguage.

(Tarski 1936 (1956), p. 187. Emphasis added).

The requirement of a translation from object language L to metalanguage ML effectively requires that the object language L be interpreted relative to ML. We may say that the translation function Γ *interprets* L within ML.

This is one of the basic differences between a *general* Tarskian semantical theory of truth and *purely disquotational* theories. In the general Tarskian setting the metalanguage ML needn't be an extension of L. In general, to theorize about truth for some object language L (say, Spanish) in some metalanguage ML (say, English) we need to *translate* from object language to metalanguage. Suppose that Γ is a *translation function* from L to ML, mapping every L-formula to an ML-formula. That is, if φ is an L-formula, then Γ(φ) is an ML-formula. Then, if *P* is a predicate definable in ML, Tarski's suggestion is, in effect:

> *P* is a *truth predicate* for L within a metatheory MT in ML just in case, for each L-sentence φ, every ML-sentence $P(\langle\varphi\rangle) \leftrightarrow \Gamma(\varphi)$ is a theorem of MT.

This formulation emphasizes two important points:

i. the dependence upon a *translation function* Γ from object language L to metalanguage ML;

ii. a materially adequate T-theorem is required only for each L-sentence, not for each ML-sentence (in fact, the stronger demand may well lead to paradox).

Given my formulation above, it is easy to see that disquotational T-instances are obtained as a special case of Convention T. A disquotational truth theory is obtained in the case where the metalanguage ML is an extension L^+ of the object language L, and the terms <φ> are genuine quotations (obtained, say, in the French manner, by prefixing and suffixing the symbols "<<" and ">>" around φ). The crucial notion of *translation* is still operative but is transparent, because the translation function Γ from L to ML is simply the identity mapping 1.

We can say:

> *Tr* is a *disquotational truth predicate* for L within a metatheory MT in ML just in case, for each L-sentence φ, $MT \vdash Tr(<\varphi>) \leftrightarrow \varphi$[72].

Many authors have stressed that what Tarski showed how to define was "truth in L", rather than just plain "univocal truth". This observation has led some authors[73] to argue as follows:

i. the concepts *truth-in-English, truth-in-Spanish, truth-in-Glaswegian*, etc., all "have something in common";

ii. this something must be *truth*;

iii. but, Tarski doesn't tell us anything about *that.*

I am sceptical about this argument. After all, Convention T provides a *general criterion* for what it is for an ML predicate to be a *truth predicate for* L, for any sufficiently well-understood object language L for which there is a translation Γ from L to ML.

Putting it another way, suppose we consider formalized fragments of, say,

i. English L_1, Spanish, L_2, German L_3, Hopi L_4 and so on,

then we can introduce within the metalanguage ML a system of truth predicates:

ii. Tr_1, Tr_2, Tr_3 and so on

and (in the meta-metatheory!) translation functions

iii. Γ_1, Γ_2, Γ_3, and so on

---

[72] This corresponds to a generalization of the standard definition of a truth predicate in mathematical logic texts:

> A formula B(y) is called a truth predicate for T if for any sentence G of the language of T, $\vdash_T G \leftrightarrow B(\ulcorner G \urcorner)$.
>
> (Boolos & Jeffrey 1989, p. 180)

My formulation, however, follows Tarski in allowing the truth predicate to appear within the *metalanguage* ML and requires that $Tr(<\varphi>) \leftrightarrow \varphi$ be a theorem of the (stronger) *metatheory* in ML. If we collapse ML down to L and consider just the object level theory T, then we are asking whether truth in the object language can be defined within the object theory. Under standard conditions, this is impossible, as Tarski's Indefinability Theorem shows. However, truth *can* be defined within the stronger metatheory. For example, first-order arithmetical truth is definable in second-order arithmetic (see Boolos & Jeffrey 1989, Ch. 19).

[73] E.g., Blackburn 1983.

such that the "overall Tarskian theory of truth" Truth implies all the materially adequate T-theorems: that is,

Truth implies, for any (English) L_1 sentence φ, $Tr_1(<\varphi>) \leftrightarrow \Gamma_1(\varphi)$

Truth implies, for any (Spanish) L_2 sentence φ, $Tr_2(<\varphi>) \leftrightarrow \Gamma_2(\varphi)$

Truth implies, for any (German) L_3 sentence φ, $Tr_3(<\varphi>) \leftrightarrow \Gamma_3(\varphi)$

and so on.

Notice, by the way, that such an envisioned over-all "multiple Convention T satisfying" Tarskian truth theory Truth has a very powerful explanatory ability:

if φ_a is an (English) L_1-sentence and φ_b is a (Spanish) L_2-sentence and $\vdash \Gamma_1(\varphi_a) \leftrightarrow \Gamma_2(\varphi_b)$ then Truth implies $Tr_1(<\varphi_a>) \leftrightarrow Tr_2(<\varphi_b>)$

For example, Truth will *imply* theorems like:

'All ravens are black' is true-in-English iff 'Todos los cuervos son negros' is true-in-Spanish

In other words, because Convention T requires a *translation* from L to ML, the general Tarskian truth theory Truth possess the capacity to *link together* truth-in-L for variable 'L' (obviously and unfortunately, however, this variability in the choice of L is not allowed to include ML itself). Indeed, Tarski's theory makes it clear how the correlation of *truth-in-English* and *truth-in-Spanish* (and so on) depends upon the *correct translation* between (non-semantical assertions of) Spanish and English (and so on).

4.6 Tarski's Theory of Satisfaction

At the heart of Tarski's theory of truth lies the concept of *satisfaction.* Tarski provides first an informal account of the concept of satisfaction using examples analogous to,

a formula 'loves(x_i, x_j)' is satisfied by a sequence s if and only if s_i loves s_j

and then gives the following *recursive definition* of 'satisfies' for the language of class inclusion (I have changed the notation a little):

> DEFINITION 22: The sequence s *satisfies* the sentential function φ if and only if s is an *infinite sequence* (of classes) and φ is a *sentential function* and these satisfy one the following four conditions:
>
> (α) there exist natural numbers *k* and *l* such that φ = '$x_k \subseteq x_l$' and $s_k \subseteq s_l$;
>
> (β) there is a sentential function ψ such that $\varphi = \neg\psi$ and *s* does not satisfy the function ψ;
>
> (γ) there are sentential functions ψ and ξ such that $\varphi = \psi \vee \xi$ and *s* either satisfies ψ or satisfies ξ;
>
> (δ) there is a natural number *k* and a sentential function ψ such that $\varphi = \forall x_k \psi$ and every infinite sequence (of classes) which differs from *s* in at most the *k*-th place satisfies the function ψ.
>
> (Tarski 1936 (1956), p. 193).

Tarski adds in a footnote that this recursive definition of 'satisfies' may be converted to a direct definition. I will explain this below.

I prefer to treat 'satisfies' axiomatically and construct a *Theory of Satisfaction*, which I shall call TS. That is, one simply introduces into ML a two-place predicate 'satisfies' and takes as axioms for TS the "basis axioms", one for each primitive predicate *P* of L:

(α) for any sequence *s*, for any variables $v_i, \ldots, v_n$, the atomic formula $P(v_i, \ldots, v_n)$ is satisfied by *s* iff $R(s_i, \ldots, s_n)$

where '*P*' is replaced by a name of *P* and '*R*' is replaced by *P* itself.

and the "recursion axioms":

(β) for any sequence *s*, for any formula φ, $\neg\varphi$ is satisfied by s if and only if φ is not satisfied by *s*

(γ) for any sequence *s*, for any formulas φ, ψ, $\varphi \wedge \psi$ is satisfied by *s* if and only if φ is satisfied by *s* and ψ is satisfied by *s*.

(δ) for any sequence *s*, for any variable v_i, for any formula φ, $\forall v_i \varphi$ is satisfied by *s* if and only if, for any sequence *s** that differ from *s* in the i-th place at most, φ is satisfied by *s**

I formalize the axioms (α) - (δ) thus: a set of basis axioms,

(α) $\forall s \forall i, ..., n$: $(Sat(\langle P \rangle \wedge (v_i, ..., v_n), s) \leftrightarrow P(s_i, ..., s_n))$

one axiom for each atomic predicate *P*. Then,

(β) $\forall s \forall f$: $(Sat(neg(f), s) \leftrightarrow \neg Sat(f, s))$

(γ) $\forall s \forall f \forall g$: $(Sat(conj(f, g), s) \leftrightarrow (Sat(f, s) \wedge Sat(g, s)))$

(δ) $\forall s \forall f \forall i$: $(Sat(uniqu(f, v_i), s) \leftrightarrow \forall s^*(Id_i(s, s^*) \rightarrow Sat(f, s)))$

The conversion of this recursive definition of *Sat* to an explicit direct definition of *Sat* is explained in Quine 1970. One takes the conjunction of the axioms (α) - (δ), then rewrites any formula *Sat*(..., ---) as the formula (..., ---) ∈ *Z*, and takes the resulting formula with *Z* free to define a formula *SR*(*Z*), meaning intuitively '*Z* is the satisfaction relation'. Then, we can define *Sat* *explicitly* as follows:

$Sat(x, y) \leftrightarrow \exists Z(SR(Z) \wedge (x, y) \in Z)$

Tarski 1936 demonstrated two important metalogical facts about this theory of satisfaction TS:

i. The Definability of Truth:

One may construct, within TS in the metalanguage ML, a *definition of truth* for any standard first-order interpreted object language L. One introduces the above recursive characterization of the satisfaction predicate *Sat* and the explicit definition of *Tr* is then the ML-formula:

(D) $\forall x[Tr(x) \leftrightarrow (x \in Sen(L) \wedge \forall s(Seq(s) \rightarrow Sat(x, s)))]$

Tarski's definition of truth is given as Definition 23 of Tarski 1936:

> DEFINITION 23: *x* is a *true sentence*—in symbols $x \in Tr$—if and only if $x \in S$ and every infinite sequence (of classes) satisfies *x*.
>
> (Tarski 1936 (1956), p. 195).

Tarski sketches how it may be verified that from Definitions 22 and 23 (plus the metatheoretical axioms of syntax and set theory) one may derive all the T-theorems and thus, the definition (D) satisfies Convention T. Thus, TS $\vdash$ DT.

Incidentally, the derivation within TS of a disquotational "T-sentence" $Tr(<\varphi>) \leftrightarrow \varphi$ is no simple matter. Take any L-sentence φ. A syntactical theorem is obtainable in the metatheory of the form $<\varphi> = t$, where t is what Tarski calls a "structural-descriptive name" (t is a closed term obtained from the functors *exiqu*, *conj*, *neg*, etc.). Instantiate the definition of truth with this term t and then apply the recursive definitions to the right-hand side (thus "decomposing" t). In this way, one eventually obtains a theorem of form $Tr(t) \leftrightarrow \varphi°$, where $\varphi°$ is a formula, closely related to φ, but containing quantifications over sequences. This is hardly a *disquotational* truth condition for φ. However, one may prove in the ambient set theory the formula $\varphi° \leftrightarrow \varphi$. (It is not logically true!). So, finally, one may prove $Tr(<\varphi>) \leftrightarrow \varphi$. Note that the proof requires a certain amount of set theory, for the appropriate manipulations involving sequences.

Of equal significance to the *definability* of truth (for an object language L, in the suitably strong metatheory TS in the metalanguage ML) is Tarski's profound *Indefinability Theorem.*

ii. Tarski's Indefinability Theorem

Suppose that T in L is a "sufficiently strong" theory. This can be made more precise: the requirement is that Robinson Arithmetic Q can be interpreted within T. And T can to some extent express the syntax of L, by arithmetization. In this situation, an analogue of the "Diagonal Lemma" or "Fixed Point Theorem" holds in T. Then no predicate definable in T in L is a truth predicate for L in T. This is Tarski's Indefinability Theorem.

At the "guts" of the Indefinability Theorem (which is usually presented in the context of formalized arithmetic) lies the *Diagonal Lemma* or *Fixed Point Theorem.* Let L be the language of arithmetic. Let # be a gödel numbering, mapping each expression of L to a number. Let φ be any L-formula. Then the *diagonalization* of φ, $D(\varphi)$, is the formula

$\exists x(x = \ulcorner\varphi\urcorner \wedge \varphi)$. Suppose φ has just x free. Then $\mathfrak{N} \models D(\varphi)$ just in case $\mathfrak{N}^{n}{}_{\#\varphi} \models \varphi[n]$. (I.e., $\mathfrak{N} \models D(\varphi)$ just in case φ is true in $\mathfrak{N}$ of $\#\varphi$). The diagonal function (for #) is the function *diag* such that *diag*(*n*) is the gödel number of the diagonalization of the formula whose gödel number is *n*. I.e., $m = diag(n)$ iff $n = \#\varphi$ and $m = \#D(\varphi)$, for some φ. It is possible to show that *diag* is recursive. Since all recursive functions are representable in Q, there is a formula $diag(x, y)$ which represents *diag*. One can then prove the Diagonal Lemma:

> for any monadic formula $P(x)$, there is a "fixed point sentence" G such that Q $\vdash$ $P(\ulcorner G\urcorner) \leftrightarrow G$.

Now assume L contains a monadic predicate *Tr* such that T $\vdash Tr(\ulcorner\varphi\urcorner) \leftrightarrow \varphi$, for each closed formula φ. If this were possible, then *Tr* would be a truth predicate for L within T in L. However, just take *P* to be $\neg Tr$ and from the Diagonal Lemma, you immediately get a "liar sentence" L such that T $\vdash \neg Tr(\ulcorner L\urcorner) \leftrightarrow L$, contradicting the T-theorem $Tr(\ulcorner L\urcorner) \leftrightarrow L$.

[Actually, the Fixed Point Theorem also holds within any theory T which formalizes its own *syntax* and which can define the diagonal function *D*, where $D(\varphi)$ is $\exists x(x = <\varphi> \wedge \varphi)$. That is, the theory T is such that there exists an L-formula $Diag(x, y)$ such that, for any φ in L, T $\vdash \forall y(Diag(<\varphi>, y) \leftrightarrow y = <D(\varphi)>)$. The Fixed Point Theorem or Diagonal Lemma states that for any monadic L-formula $P(x)$ there is a closed L-formula φ such that T $\vdash P(<\varphi>) \leftrightarrow \varphi$. To prove this, suppose that $P(x)$ is any L-formula. Let *F* be the expression $\exists y(Diag(x, y) \wedge P(y))$, containing just x free. Let G be its diagonalization, the expression $\exists x(x = <F> \wedge \exists y(Diag(x, y) \wedge P(y)))$. Clearly, $G = D(F)$. Thus, by hypothesis, T $\vdash \forall y(Diag(<F>, y) \leftrightarrow y = <G>)$. But $G \Leftrightarrow \exists y(Diag(<F>, y) \wedge P(y))$. Thus, trivially, T $\vdash G \leftrightarrow \exists y(Diag(<F>, y) \wedge P(y))$. Hence, T $\vdash G \leftrightarrow \exists y(y = <G> \wedge P(y))$. So, T $\vdash G \leftrightarrow P(<G>)$.][74, 75]

---

[74] This proof is modelled on the proof of the diagonal lemma in Boolos & Jeffrey 1989, p. 173.

[75] For another application of the Fixed Point Theorem. Let T be a theory containing formal syntax in a language L with quotation. If T satisfies the Fixed Point Theorem, then we can prove:

> if T is consistent, then T cannot define the class of theorems in T.

Proof: First, T *defines* a class **A** of sentences just in case there is a formula $A(x)$ such that, In general, Tarski's Indefinability Theorem states that the concept of truth for an interpreted notation is not definable in any sufficiently strong theory in that notation. (Or in the set of true sentences in that notation). For example, the set of true sentences in first-order set theory is not definable in first-order set theory (i.e., the theory consisting of all the truths). Thus the set of first-order truths in $L_\in$ is not definable in ZFC or any (first-order) extension.

An important and philosophical consequential way of restating Gödel's Incompleteness Theorem involves the concept of truth and is due to Tarski. The class of truths in any theory containing elementary arithmetic is not *recursively enumerable.* For let T be $Th_L(\mathfrak{N})$ (i.e., first-order arithmetic). Suppose that T is r.e. Then, by Craig's Reaxiomatization Theorem, there would be a *recursive axiomatization* Δ such that T = DedCl(Δ). Thus, arithmetic would be axiomatizable, contradicting Gödel's First Incompleteness Theorem. So, T is not r.e.

Putnam made an epistemologically important comment on this fact:

> Even if all statements that can be *proved* are epistemologically *a priori* and conversely, the statements that can be proved from axioms which are evident to us can only be a *recursively enumerable set* (unless an infinite number of irreducibly different principles are at least potentially evident to the human mind, a supposition I find quite incredible). And Gödel's theorem can (in a version due, fundamentally, to Tarski) be expressed by the statement that the class of truths of just elementary number theory is not recursively enumerable
>
> (Putnam 1975 (1979), p. 63).

While no theory can adequately formulate its own truth theory (i.e., the disquotational truth theory for the language in which it is formulated), there are two more "positive" facts known about the definability of truth. The first is that (first-order arithmetical) truth for formulas of *bounded complexity* is definable in first-order arithmetic. More exactly, if

---

i. if $\varphi \in A$, then $T \vdash A(<\varphi>)$

ii. if $\varphi \notin A$, then $T \vdash \neg A(<\varphi>)$

Assume that T defines T. Then by the Fixed Point Theorem, there must be a formula φ such that $T \vdash \neg A(<\varphi>) \leftrightarrow \varphi$. Assume $\varphi \in T$. Then $T \vdash A(<\varphi>)$ and $T \vdash \neg A(<\varphi>)$. This is not possible, since T is consistent. So, $\varphi \notin T$. Thus, $T \vdash \neg A(<\varphi>)$ and $T \vdash \varphi$. Contradiction. ■

For example, PA cannot define (in this strong sense) the set of theorems of PA. (Even though one can formalize as a provability predicate the notion of proof-in-PA within PA).

V_n is the class of g.n.'s of true first-order formulas of logical complexity less than n, then V_n *is* definable in arithmetic[76]. (Indeed, this carries over to set theory). And if V is the class of g.n.'s of true *first-order* formulas in arithmetic, then V *is* definable in *second-order* arithmetic[77]. (But the class of true second-order formulas is not definable in second-order arithmetic).

Finally, some philosophers have argued that truth (and other semantic concepts, like reference and satisfaction) may somehow be "reduced" to physical notions. This doctrine is known as "Semantic Physicalism"[78]. The problem is that this would contradict Tarski's Theorem. For suppose that N is a "strong enough" physical theory in some interpreted language (L, A). It follows that the class of truths in the language of N (i.e., truths in the structure A) is not definable in N. In short, *physical truth is simply not a physical concept*. Analogously, *set theoretical truth is not a set-theoretical concept*. Analogously, *mathematical truth is not a mathematical concept*. If you like, truth is a "transcendent" concept, resisting reduction to either physics, or even physics *plus* mathematics. Note that the class of first-order truths in set theory is a countable set, and thus not particularly "big", in the set-theoretical sense. Rather it is its "*shape*" or structure which is beyond complete comprehension (computability, axiomatizability, provability, etc.). This illustrates the proverb that "size isn't important — it's shape that counts".

---

[76] See Boolos & Jeffrey 1989, Chapter 19, 'On Defining Arithmetical Truth'.

[77] *Ibid.*

[78] E.g., Field 1972:

> physicalism: the doctrine that chemical facts, biological facts, psychological facts and semantical facts are all explicable (in principle) in terms of physical facts.
>
> (Field 1972 (1980), p. 93).

CHAPTER 5

Deflationism and Tarski's Paradise

There is really no separate problem of truth, merely a linguistic muddle.

F.P. Ramsey 1927 (1978), p. 44.

... We conclude, then, that there is no problem of truth as it is ordinarily conceived. The traditional conception of truth as a 'real quality' or a 'real relation' is due, like most philosophical mistakes, to a failure to analyse sentences correctly.

A.J. Ayer 1936 (1946), pp. 116-119.

Truth is thus a notion that we might reasonably want to have at hand, for expressing semantic ascent and descent, infinite conjunction and disjunction.

Stephen Leeds 1978, p. 121.

... our thesis is that it is possible to explain *all* the facts involving truth on the basis of the minimal theory.

Paul Horwich 1990, p. 7.

... the minimalist *conception*: i.e., the thesis that our theory of truth should contain nothing more than instances of the equivalence scheme.

Paul Horwich 1990, p. 8.

5.1 Deflationism About Truth and Mathematics

In the previous Chapter we introduced the so-called deflationary conception of truth. I now want to suggest an *analogy* between deflationism about *truth* and deflationism about *mathematics*. Indeed, the deflationary view of truth has been widely discussed and is quite popular amongst philosophers. If I am right, deflationism about truth suffers from a problem quite analogous to a problem suffered by deflationism about mathematics, a programme initiated by Field 1980. The purpose of this Chapter then is to show that deflationism about truth is untenable.

What is *essential* to any deflationary view of truth? Many deflationists urge that,

i. The concept of truth is not *essential* to explanations

From this claim it surely follows that,

ii. Any assumptions involving the concept of truth are *eliminable*

or,

iii. Anything explained using truth-theoretic assumptions could be explained *without them*

This idea can be made quite precise. In technical terms, the claims (i) - (iii) may be reformulated more precisely as:

The Conservativeness Thesis

> The result of adding a truth theory to a collection of non-truth-theoretic assumptions always yields a *conservative extension.*

It is this claim that I shall take to form a basic component of the deflationary view. This idea is based on an analogy with Field's formulation of the dispensability of mathematics. Field claimed that,

> ... any inference from nominalistic premises to a nominalistic conclusion that can be made with the help of mathematics could be made (usually more long-windedly) without it.
>
> (Field 1980, p. x).

This is naturally formulated as the claim that adding a mathematical theory to a mathematics-free theory always yields a conservative extension. A deflationist about truth might put their view similarly. E.g.,

> Any inference from non-semantical premises to a non-semantical conclusion that can be made with the help of a truth theory could be made without it.

I discuss Field's conservativeness claim a more fully in Chapter 9 below.

However there is a further deflationary thesis which might also be laid alongside the conservativeness thesis. It is well articulated by Horwich 1990:

Completeness Thesis

> There is nothing more to truth than is given by the theory consisting of "all" instances of the redundancy scheme, 'the proposition that *p* is true if and only

if *p*'. Such a theory, the *Minimalist Theory*, is a *complete* theory of the concept of truth.

I shall argue that both theses are false.

5.2 The Conservativeness of Deflationary Theories

Let L be any standard formalized first-order language. L will be the "object level" or "base" language for which we shall examine some (also formalized) truth theories. (We shall use the symbol '$_n$L' to refer to the set of L-formulas with *n* free variables).

I will concentrate on three *deflationary theories of truth* for L:.

i. MT = the minimalist theory of truth for (propositions expressible in) L

ii. DT = the disquotational theory of truth for L.

iii. SDT = the substitutional disquotation theory of truth for L

The *minimalist theory* MT (c.f., Horwich 1990) is formulated in a metalanguage L^+, which is an expansion of L (so, $L \subset L^+$) obtained by adding a monadic predicate *Tr* and a term-forming operator $<...>$ which operates only on L-sentences to form L^+-terms. If $\varphi \in {}_0L$ (i.e., φ is an L-sentence), then $<\varphi>$ is an L^+-term (we assume that $<...>$ is injective: if φ_1 and φ_2 are distinct, then so are $<\varphi_1>$ and $<\varphi_2>$). Intuitively speaking, $<...>$ means "the proposition that . . .".

Then let Ramsey(φ) be the L^+-sentence $Tr(<\varphi>) \leftrightarrow \varphi$, where $\varphi \in {}_0L$. The *axioms* of MT are then all these L^+-sentences, Ramsey(φ). Thus, the axioms of MT are analogous to:

(1) the proposition that snow is white is true if and only if snow is white

The *disquotational theory* DT is formulated in a metalanguage L^+, which is an expansion of L (again, $L \subset L^+$) obtained by adding a monadic predicate *Tr* and a term-forming operator $<...>$. So, if ε is any L-expression, then $<\varepsilon>$ is an L^+-term (again, we assume injectiveness: if ε_1 and ε_2 are distinct, then so are $<\varepsilon_1>$ and $<\varepsilon_2>$). Intuitively speaking, <...> is the quotation operator. We will say that an interpretation $\Im^{+}$ of L^{+} is *quotation-normal* if, for any L-expression ε, $\Im^{+}[<\varepsilon>] = \varepsilon$. (Such quotation-normal interpretations just invert quotation).

Then let Tarski(φ) be the L^{+}-sentence $Tr(<\varphi>) \leftrightarrow \varphi$, where $\varphi \in {}_{0}L$. The *axioms* of DT are then all these L^{+}-sentences, Tarski(φ). Thus, the axioms of DT are analogous to:

(2) the sentence 'snow is white' is true if and only if snow is white

From a formal point of view, the truth theories MT and DT are identical, as I have hinted in the construction above.

The *substitutional disquotational theory* SDT is formulated in a strengthened metalanguage L^{+}_{sub}, containing the monadic predicate *Tr*, the term-forming operator <...> and a substitutional quantifier Σ (with a denumerable infinity of substitutional sentence variables, p_1, p_2, ...). Unlike MT and DT which are not finitely axiomatized, SDT consists simply of the axiom:

SDT: $\forall x(Tr(x) \leftrightarrow \Sigma p(x = <p> \wedge p))$.

It is possible to see that each of these theories is a *truth theory* for L, in the sense explicated by Tarski, that they satisfy Convention T. And thus, *Tr* is a *truth predicate* for L in MT, DT and SDT.

Consider the minimalist theory MT.[79] Every L^{+}-sentence Ramsey(φ) is an instance of RS and thus MT is "the totality of instances" of RS. By construction, MT is just the set of all these sentences. Similarly, the disquotation theory DT, in Tarski's terminology, satisfies Convention T, for (by construction) it derives every L^{+}-sentence Tarski(φ), =

[79] Some authors (e.g., David 1994, p. 107-110) express concern that theories of truth like MT and DT below are *infinitely* axiomatized and thus not "finitely statable". I simply do not get the problem. These authors *have* finitely defined the theory. (It is like saying we cannot consider the set of reals between 0 and 1, because we cannot finitely list them all). Almost any interesting theory is not finitely (first-order) axiomatizable. E.g., first-order PA with the infinity of induction axioms, first-order Real Analysis (RA) with an infinity of continuity axioms and ZFC with an infinity of separation and replacement axioms. Indeed, take almost any interesting infinite mathematical structure *A* (e.g., (**N**, <) or (**R**, <) and so on). Then the *theory of A*—the set of first-order sentences that hold in *A*—will not be finitely axiomatizable. (It may not even be recursively enumerable! E.g., first-order arithmetic). Nevertheless, the theory of *A* is defined as $\{\varphi \in {}_{0}L: A \models \varphi\}$, and *that's* finitely statable.

$Tr(<\varphi>) \leftrightarrow \varphi$. More precisely, using the terminology of textbooks, Tr in L^+ is a *truth predicate for* L within DT (remember, the truth theory DT is formulated in L^+). Finally, SDT also satisfies Convention T. The proof is this: begin with $\forall x(Tr(x) \leftrightarrow \Sigma p(x = <p> \wedge p))$, from which $Tr(<\varphi>) \leftrightarrow \Sigma p(<\varphi> = <p> \wedge p))$ follows. Within this substitutional logic, one may show that φ is derivable from $\Sigma p(<\varphi> = <p> \wedge p))$ (in particular, using an axiom scheme governing quotations, namely, $<\varphi> = <\psi> \rightarrow (\varphi \leftrightarrow \psi))$. So, we may derive the T-sentence $Tr(<\varphi>) \leftrightarrow \varphi$.

Now I want to present the first technical result. These deflationary theories of truth, MT, DT and SDT, are *conservative* over any ("non-semantical") theory T in L.

Theorem 1: The Deflationary Theories of Truth are Conservative

Let T in L be a theory. Let $\varphi \in {}_0L$. Then

(a) if $T \cup DT \vdash \varphi$, then $T \vdash \varphi$

(b) if $T \cup SDT \vdash \varphi$, then $T \vdash \varphi$

Proof: Part I: The proof requires a pair of "expansion lemmas":

(*) any model of T may be expanded to a model of $T \cup DT$

(**) any model of T may be expanded to a model of $T \cup SDT$

We prove (*) and (**) by constructing, using any L-interpretation $\mathfrak{I}$, expansions $\mathfrak{I}^+$ (for L^+) and $\mathfrak{I}^+_{sub}$ (for L^+_{sub}) such that $\mathfrak{I}^+ \models DT$ or $\mathfrak{I}^+_{sub} \models SDT$. Let $\mathfrak{I}$ be an L-interpretation.

Let $\mathfrak{I}^+$ be the L^+-interpretation which is an expansion of $\mathfrak{I}$ such that:

i. $dom(\mathfrak{I}^+) = dom(\mathfrak{I}) \cup \{\varphi: \varphi \in {}_0L\}$,

ii. $\mathfrak{I}^+[<\varepsilon>] = \varepsilon$, for any L-expression ε,

iii. $\mathfrak{I}^+[Tr] = \{\varphi \in {}_0L: \mathfrak{I} \models \varphi\}$

Let $\mathfrak{I}^+_{sub}$ be the L^+_{sub}-interpretation which is an expansion of $\mathfrak{I}$ such that:

i. $dom(\mathfrak{I}^+_{sub}) = dom(\mathfrak{I}) \cup \{\varphi: \varphi \in {}_0L\}$,

ii. subclass($\Im^{+}_{sub}$) = $\{\varphi : \varphi \in {}_0\mathrm{L}\}$,

iii. $\Im^{+}_{sub}[<\varepsilon>] = \varepsilon$, for any L-expression ε,

iv. $\Im^{+}_{sub}[Tr] = \{\varphi \in {}_0\mathrm{L} : \Im \models \varphi\}$.

Clearly, $\Im^{+} \models Tr(<\varphi>) \leftrightarrow \varphi$, for any $\varphi \in {}_0\mathrm{L}$. Hence, $\Im^{+} \models \mathrm{DT}$.

Also, $\Im^{+}_{sub} \models \mathrm{SDT}$ just in case $\Im^{+}_{sub} \models \forall x(Tr(x) \leftrightarrow \Sigma p(x = <p> \wedge p))$, just in case, for all $e \in \mathrm{dom}(\Im^{+}_{sub})$, $\Im^{+}_{sub}{}^{x}_{e}[Tr(x)] = \mathbf{T}$ iff $\Im^{+}_{sub}{}^{x}_{e}[\Sigma p(x = <p> \wedge p)] = \mathbf{T}$. This holds just in case, for all $e \in \mathrm{dom}(\Im^{+}_{sub})$, $e \in \Im^{+}_{sub}[Tr]$ iff, for some $\varphi \in {}_0\mathrm{L}$, $\Im^{+}_{sub}{}^{x}_{e}[x = <\varphi> \wedge \varphi)] = \mathbf{T}$. This holds just in case (for all $e \in \mathrm{dom}(\Im^{+}_{sub})$, $e = \varphi$, for some $\varphi \in {}_0\mathrm{L}$ and $\Im \models \varphi$) iff (for some $\varphi \in {}_0\mathrm{L}$, $\Im^{+}_{sub}{}^{x}_{e}[x = <\varphi>] = \mathbf{T}$ and $\Im^{+}_{sub}{}^{x}_{e}[\varphi] = \mathbf{T}$). This holds just in case, for all $\varphi \in {}_0\mathrm{L}$, $\Im \models \varphi$ iff $\Im \models \varphi$. And this holds come what may. That proves the expansion lemmas.

Part II: Using these expansion lemmas we prove that DT and SDT are conservative. We'll do it just for DT. Assume the negation of conservativeness: for some $\varphi \in {}_0\mathrm{L}$, $\mathrm{T} \cup \mathrm{DT} \vdash \varphi$ and not-($\mathrm{T} \vdash \varphi$). There must be a model $\Im$ of T such that $\Im \models \neg\varphi$. Furthermore, from the lemma, because $\Im \models \mathrm{T}$, there is an expansion $\Im^{+}$ such that $\Im^{+} \models \mathrm{T} \cup \mathrm{DT}$. Because $\Im^{+}$ is an expansion of $\Im$, it follows that $\Im^{+} \models \neg\varphi$. But $\Im^{+} \models \mathrm{T} \cup \mathrm{DT}$ and, by assumption, $\mathrm{T} \cup \mathrm{DT} \vdash \varphi$. So, $\Im^{+} \models \varphi$. Contradiction. ■

It is possible to provide a *proof-theoretic* demonstration for the results that the deflationary truth theories always yield conservative extensions. In particular, this establishes that such theories are conservative even when $Tr(x)$ is permitted to appear in any axiom schemes within T.

Proof: Let **A** be a finite set of closed L-formulas. And let $\mathrm{DT}_{\mathbf{A}} =_{df} \{Tr(<\varphi>) \leftrightarrow \varphi : \varphi \in {}_0\mathrm{L}$ and $\varphi \in \mathbf{A}\} = \mathrm{DT} \cap \mathbf{A}$. That is, a *finite* set of "T-sentences". Suppose that $\mathrm{T} \cup \mathrm{DT} \vdash \varphi$. Then, since proofs must be finite, there is a finite subset Δ of proper axioms of T and a finite set **A** of L-formulas, such that $\Delta \cup \mathrm{DT}_{\mathbf{A}} \vdash \varphi$. (Indeed, $\mathrm{DT}_{\mathbf{A}}$ is equivalent to a single formula).

Let $\varphi_1, \ldots, \varphi_n$ enumerate the elements of **A**. Now define an L-formula $A_n(x)$:

$((x = \langle\varphi_1\rangle \wedge \varphi_1) \vee \ldots \vee (x = \langle\varphi_n\rangle \wedge \varphi_n))$

The idea is to find a single formula Ψ, a *definition* of $Tr(x)$, such that $\Psi \vdash DT_A$. First, consider $A_1(x) \equiv: (x = \langle\varphi_1\rangle \wedge \varphi_1)$. We can show that: $\forall x(A_1(x) \leftrightarrow Tr(x)) \vdash (Tr(\langle\varphi_1\rangle) \leftrightarrow \varphi_1)$. That is, $\forall x((x = \langle\varphi\rangle \wedge \varphi) \leftrightarrow Tr(x)) \vdash Tr(\langle\varphi\rangle) \leftrightarrow \varphi$. Next, let Ψ_A be the formula $\forall x(A_n(x) \leftrightarrow Tr(x))$. Then, we can prove (by induction) that $\Psi_A \vdash DT_A$.

So, if $\Delta \cup DT_A \vdash \varphi$, then $\Delta \cup \{\Psi_A\} \vdash \varphi$. Thus, $\Delta \cup \{\Psi_A\} \vdash \varphi$. But Ψ_A is an *explicit definition* of $Tr(x)$ in the L-vocabulary; so adding Ψ_A yields a conservative extension.

Case 1: Suppose that $Tr(x)$ does not appear in any induction scheme in Δ. Since the definition Ψ_A is redundant, we can drop it. So, $\Delta \vdash \varphi$, and thus, $T \vdash \varphi$. Thus, DT is conservative over T.

Case 2. Suppose that $Tr(x)$ occurs in instances of schemes of T, say, $\Phi_1, \ldots, \Phi_k$, in Δ. We know that $\Delta \cup \{\Psi_A\} \vdash \varphi$. Let Δ^* be the result of replacing $Tr(x)$ by $A_n(x)$ in each scheme Φ_i in Δ. The results are simply new schemes in T. So, Δ^* is a set of axioms of T. Then, $\Delta^* \cup \{\Psi_A\} \vdash \varphi$. Again, $\Delta^* \vdash \varphi$. So, $T \vdash \varphi$. Thus, DT is conservative over T. ■

A similar proof-theoretic argument works to establish the conservativeness of SDT even when axiom schemes are in operation. Although SDT is an explicit definition of *Tr*, we cannot immediately conclude its conservativeness, because it is framed using apparatus (i.e., substitutional quantification w.r.t. L-sentences) which exceeds that of the object language. Still, let us suppose that $T \cup SDT \vdash \varphi$, where φ is an L-sentence. So, we have a proof from $T \cup SDT$ of φ. This is equivalent to a proof of φ from $T \cup \Pi p(Tr(\langle p\rangle) \leftrightarrow p)$. Now, again, any proof will contain only finitely many L-sentences, and we can suppose that all of these are instantiated from $\Pi p(Tr(\langle p\rangle) \leftrightarrow p)$. But each such instance $Tr(\langle\varphi\rangle) \leftrightarrow \varphi$ is just an axiom of DT. Thus, again, there must be a *finite* set A of L-sentences, such that $T \cup DT_A \vdash \varphi$. The above argument then goes through to show that $T \vdash \varphi$. Again, we conclude that SDT is conservative. ■

It is worth pointing out that the conservativeness results provide *consistency proofs* for the deflationary truth theories DT, MT and SDT. They show that by *restricting* the occurrence of the predicate *Tr* in L^+-formulas one obtains a consistent theory of truth (and thus avoids the semantic antinomies). I will not examine the proposals for further extending the truth theories so as to define truth for formulas themselves containing *Tr*. Tarski's 1936 proposal is to introduce a hierarchy of metalanguages L^+, L^{++}, . . ., and truth predicates Tr_0, Tr_1, . . . Instead, Kripke's 1975 proposal is to maintain a "univocal" truth predicate *Tr*, but relax the Principle of Bivalence for certain sensitive (ungrounded) formulas (such as "This sentence is false")[80].

Now, if all the above looks very abstract, then take note of the following corollaries.

Corollary 1.1 (the contentlessness principle):

No statement (in $_0L$) follows from a deflationary theory of truth unless it is a logical truth.[81].

Corollary 1.2 (the irrefutability principle):

No non-semantical contingent assertion (in $_0L$) could refute a deflationary theory of truth.

Corollary 1.3 (the consistency principle):

A deflationary theory of truth is consistent with any consistent (non-semantic) theory (in L).

One might suggest that these corollaries illustrate a kind of "analyticity" or "contentlessness" that deflationary theories of truth exhibit. Adding them adds nothing. And perhaps this is somehow connected to the idea that the deflationary truth theories illustrate the "redundancy" or "non-substantiality" of truth. Indeed, one might go further: *if* truth is non-substantial—as deflationists claim—*then* the theory of truth *should* be conservative. Roughly,

non-substantiality of truth ⇔ *conservativeness of truth theory.*

[80] See Kripke 1975. See Kirkham 1992 (1995) for discussion.

[81] Quick proof: from the conservativeness theorem, if $T \cup DT \vdash \varphi$, then $T \vdash \varphi$. Now let $T = \varnothing$. Hence, if $DT \vdash \varphi$, then $\varnothing \vdash \varphi$. That is, if $DT \vdash \varphi$, then φ is a logical truth (in L, of course). QED.

5.3 Non-Conservativeness of Tarski's Theory of Truth

Next I present the second main technical result, concerning Tarski's theory of satisfaction TS. The result is that TS not generally conservative over theories in the object language L. We shall now briefly treat this theory as a formalized theory. It is formulated in a more powerful metalanguage L^+, still an expansion of L, containing the following extra apparatus:

i. syntactical predicates and functors: *Form*(*x*), *Sen*(*x*), *var*(*i*), etc., with obvious intuitive meanings ("formula", "sentence", "the *i*th variable", etc.).

ii. syntactical functors: *neg*(*x*), *conj*(*x*, *y*) and *exiqu*(*x*, *i*), with obvious intuitive meanings ("the negation of", "the conjunction of", "the existential quantification w.r.t *i*th variable");

iii. the concatenation functor, ^;

iv. all the terms generated by a suitable quotation-operator <...> (that is, for any L-expression ε, <ε> is an L^+-term);

v. a two-place satisfaction predicate, *Sat*(*x*, *s*)

vi. the mathematical predicate, ∈.

We assume that within this set-theoretical metalanguage we can define three further notions:

vii. a one-place predicate, *Seq*(*s*), meaning "*s* is a sequence";

viii. a three-place predicate, *Id*(*s*, *s**, *i*), meaning "sequences *s* and *s** are identical except possibly at the *i*th place";

ix. a two-place functor *tm*(*s*, *i*), meaning "the *i*th term of the sequence *s*".

Aside from any axioms governing the syntax of the object language or the set theory, the axioms governing *Sat* in TS are as follows:

A. The *Basis Axioms* of TS

for each *n*-place L-primitive *P*,

$\forall i ... \forall n \forall s[Sat(\langle P\rangle \wedge var(i) \wedge ... \wedge var(n), s) \leftrightarrow P(tm(s, i), ..., tm(s, n))]$

is an axiom.

B. The *Recursion Axioms* of TS

$\forall f \forall s[Sat(neg(f), s) \leftrightarrow \neg Sat(f, s)]$

$\forall f \forall g \forall s[Sat(conj(f, g), s) \leftrightarrow (Sat(f, s) \wedge Sat(g, s))]$

$\forall i \forall f \forall s[Sat(exiqu(f, i), s) \leftrightarrow \exists s^*(Seq(s^*) \wedge Id(s, s^*, i) \wedge Sat(f, s^*))]$

And, finally,

C. The *Explicit Definition* of *Tr* in terms of *Sat*:

$\forall x[Tr(x) \leftrightarrow (Sen(x) \wedge \forall s(Seq(s) \rightarrow Sat(x, s))]$

Notice that the metatheory TS must be taken also to include,

i. some (suppressed) axioms governing formation rules for formulas of the object language L (e.g., if *P* is an *n*-place primitive predicate and x_i is the *i*th variable, then the concatenation of *P* with $x_i, ...x_n$ is an atomic formula; and so on),

ii. some (suppressed) set-theoretical axioms governing the properties of sequences (e.g., an axiom saying that *s* is a sequence if and only if *s* is a complete function on the natural numbers).

The full theory of satisfaction TS is then the union of these axioms (syntactical, semantical and set-theoretical). The importance of TS is that, as Tarski showed, it can be shown to derive DT and thus satisfies Tarski's Convention T: that is, TS $\vdash$ DT.

We shall prove, using Gödel's Second Incompleteness Theorem, that the metatheory TS is *not* in general conservative over theories in L. We shall prove this with respect to PA, first-order Peano Arithmetic:

Theorem 2: Non-Conservativeness of the Full Theory of Satisfaction

Let L be the language of arithmetic and let TS be the theory of satisfaction in L^+ for L. Then PA $\cup$ TS is not a conservative extension of PA

Proof Sketch[82]: We know from Gödel's Second Incompleteness Theorem that there is a sentence Con_{PA} expressible in L (expressing the syntactical consistency of PA) which is not derivable from the axioms of PA. However, in PA $\cup$ TS one may prove that PA is true (that is, the existence of a standard model of PA), and thus prove the consistency of PA. Hence, within PA $\cup$ TS one may derive Con_{PA}. ■

We can construct a proof-theoretic argument for this result using some standard facts about PA and some further facts, established by Tarski 1936, about what can be proved using TS. We know from Gödel's Second Incompleteness Theorem that:

I: there is a provability predicate $Prov_{PA}$ definable in L which satisfies the usual constraints on a provability predicate[83]. Then the assertion Con_{PA}, = $\neg Prov(\ulcorner 0 \neq 0 \urcorner)$[84], is expressible in L and "asserts" that PA is consistent.

II: not:-(PA $\vdash Con_{PA}$).

Let $PA^+ = PA \cup TS$. Let Tr_{PA} be the L^+-formula $\forall x(Prov_{PA}(x) \rightarrow Tr(x))$, which asserts (intuitively speaking) that PA is true (that is, anything provable in PA is true). Then we have two further important facts:

i. $TS \vdash Tr(\ulcorner \varphi \urcorner) \leftrightarrow \varphi$, for any sentence φ of L

ii. $PA^+ \vdash Tr_{PA}$

Fact (i) just states that TS satisfies Tarski's Convention T. That is, TS derives the materially adequate T-theorems. Equivalently, TS $\vdash$ DT.

---

[82] This fact is apparently well known to mathematical logicians. I am indebted to John Burgess for pointing this out (private communication).

[83] See Boolos & Jeffrey 1989, Chapter 16, "Provability Predicates and the Unprovability of Consistency".

[84] Instead of the L^+-quotation <$0 = 0$>, I have switched to the standard gödel numeral $\ulcorner 0 = 0 \urcorner$, which is in fact a term in the *base language* L. This makes no essential difference, for if the base theory T is strong enough (roughly, as strong as PA), then the syntax of L may be formalized (via arithmetization) within T.

Fact (ii) just states that PA^+ derives the conclusion that PA is true (that is, that every theorem of PA is true; that is, that if φ is provable in PA, then φ is true). In fact, this result is equivalent to Theorem 5 of Tarski 1936, which Tarski expresses as $Pr \subseteq Tr$. Informally, Tarski proves this by showing first that the metatheory can prove that each axiom of the object theory is true[85]. Next, one shows that the metatheory can prove that the rules of deductive inference preserve truth, so that anything derived from true assumptions is also true[86]. Thus, the metatheory can prove that all the theorems of the object theory are true.

From (ii), we infer, $PA^+ \vdash Prov_{PA}(\ulcorner 0 \neq 0 \urcorner) \rightarrow Tr(\ulcorner 0 \neq 0 \urcorner)$. And, using (i), $PA^+ \vdash Prov_{PA}(\ulcorner 0 \neq 0 \urcorner) \rightarrow 0 \neq 0$. Then, $PA^+ \vdash \neg(0 \neq 0) \rightarrow \neg Prov_{PA}(\ulcorner 0 \neq 0 \urcorner)$. But $PA^+ \vdash \neg(0 \neq 0)$. Thus $PA^+ \vdash Con_{PA}$. ■

To summarize, adding the theory TS of truth/satisfaction for the language of arithmetic to PA permits the deduction of the consistency of PA, which is an assertion not deducible within PA itself (assuming PA is consistent (which it is, of course)).

Consider the second-order case. If T is a *categorical second-order theory* (say axiomatic second-order Peano Arithmetic, PA^2), then any new theorems in $T \cup TS$ are L-assertions which already hold in the (effectively unique) model of T and are thus *semantically implied* by T. But Gödel's Incompleteness Theorem still applies, but now as an incompleteness of the second-order *deducibility* relation, $\vdash$. In the first-order case, Gödel's (1930) Completeness Theorem ensures that the proof-theoretic concept $\vdash$ is coextensive with the model-theoretic concept $\models$. But in the second-order case, Gödel's (1931) Incompleteness Theorem ensures that $\vdash$ is a proper subset of $\models$. E.g., Con_{PA} is not a *theorem* of PA^2 but is a *logical consequence* ($\models$) of PA^2. Indeed, the second-order formula $PA^2 \wedge \neg Con_{PA}$ is *consistent* but unsatisfiable (logically-false!). In any case, adding the satisfaction theory TS allows the deduction of some of these previously

[85] The axioms of the object theory T are also axioms of the metatheory $T \cup TS$. Note also that the metatheory contains an *explicit definition* of 'φ is an axiom of T'.

[86] Roughly, the metatheory $T \cup TS$ can prove:

if Δ is a set of true L-sentences and φ is a deductive consequence of Δ, then φ is true underivable logical implications (of course, T $\cup$ TS is not *complete*: it is an axiomatic theory within which PA may be interpreted, so it may be gödelized too!).

5.4 ω-Incompleteness, Truth Laws, Non-Standard Models

Above, we constructed our deflationary disquotational theory of truth DT simply as the set of all required theorems $Tr(<\varphi>) \leftrightarrow \varphi$. There is nothing wrong with this (if the procedure is consistent, of course: which it is!). DT is the set of all ML-formulas, $Tr(<\varphi>) \leftrightarrow \varphi$, for each L-sentence φ. Then clearly, by construction, *Tr* is a disquotational truth predicate for L within DT in L^+.

However, the deflationary theory DT is not (in the usual sense) a definition. A simple argument (using the Beth Definability Theorem) shows that DT cannot even implicitly define *Tr* [see below].

Furthermore, DT does not provide for unique eliminability of the truth predicate in any open formulas $Tr(x)$ of the metalanguage. To clarify this, using the disquotation theory DT we *can* eliminate the truth predicate from certain metalanguage statements, say $Tr(<\varphi_1>) \leftrightarrow \neg Tr(<\varphi_2>)$. We just get the "deflation" of this statement, namely $\varphi_1 \leftrightarrow \neg\varphi_2$. However, we cannot eliminate *Tr* from a metalanguage statement like $\forall x(Tr(x)) \vee \neg Tr(x))$ containing an open formula $Tr(x)$. This sentence, we might say, is *non-deflatable.* Of course, within Tarski's full theory of truth TS, we may eliminate $Tr(x)$ using the definition, leaving a formula containing *Sat*. Furthermore, if the object language L does not contain the membership predicate $\in$, then the recursive definition of *Sat* can made explicit, so we may also eliminate *Sat* (of course, by Tarski's Indefinability Theorem, this is impossible if one attempts to define satisfaction for the first-order language of set theory within first-order set theory).

The deflationary theory DT has four important logical properties:

i. DT has important non-standard models;

ii. DT cannot prove certain general laws of truth (like the law for negation);

iii. DT can prove "all the instances" of such general laws, so it has a property analogous to ω-incompleteness;

iv. DT does not *implicitly define* *Tr* (or "fix the extension" of *Tr*).

First we prove (i) by exhibiting a *non-standard model* of DT. That is, we exhibit an interpretation of the metalanguage L^+ which is a model of DT but in which the general "law of negation" fails. This proves (ii). Then we show that, despite DT's failure to prove this law, it still has the power to prove "all" of its instances, so it is analogous to certain incomplete axiomatizations of arithmetic (like Robinson Arithmetic Q). Finally, we prove (iv), which is a refutation of the oft-made claim that Tarski's Convention T *on its own* "fixes the extension" of 'true'.

i. Non-Standard Model $\Im^+$

Let $\Im$ be any L-interpretation. Let **Tr** be the theory of $\Im$ (the set of true L-sentences in $\Im$). A *standard* model of DT is just the expansion ($\Im$, **Tr**). Construct a non-standard model $\Im^+_{ns}$ by setting:

i. $\text{dom}(\Im^+_{ns}) = \text{dom}(\Im) \cup \{a, \varnothing\}$, where a is not an element of dom($\Im$).

ii. $\Im^+_{ns}[Sen] = \{\varphi: \varphi \in {}_0L\} \cup \{a\}$

ii. if $\varphi \in {}_nL$, for some n, then $\Im^+_{ns}[neg](\varphi) = \neg\varphi$

$\Im^+_{ns}[neg](a) = a$

if x is neither an L-formula nor a, then $\Im^+_{ns}[neg](x) = \varnothing$

iii. $\Im^+_{ns}[Tr] = \{\varphi \in {}_0L: \Im \models \varphi\} \cup \{a\}$

iv. $\Im^+_{ns}[<\varepsilon>] = \varepsilon$, for any L-expression ε.

Now $\Im^+_{ns}$ is surely a model of DT, because, for any L-sentence φ, $\Im^+_{ns} \models Tr(<\varphi>) \leftrightarrow \varphi$. Now call anything in the extension in $\Im^+_{ns}$ of *Sen* a SENTENCE. In particular, the object a is a SENTENCE. Call the extension in $\Im^+_{ns}$ of *neg* the NEGATION function. And call anything in the extension in $\Im^+_{ns}$ of *Tr* a TRUTH. If we assign any sentence φ to x, then the formula $Sen(x) \rightarrow (Tr(neg(x)) \leftrightarrow \neg Tr(x))$ comes out true in $\mathfrak{I}^{+}_{ns}$. But if we assign, say, the SENTENCE *a* to x, the formula comes out *false*. This is because both the object *a* and its NEGATION (i.e., *a* itself) are TRUE SENTENCES.

ii. DT does not imply the general law of negation

Now consider the law of negation, NEG, which is just the L^{+}-formula

$$\text{NEG:} \qquad \forall x(Sen(x) \rightarrow [Tr(neg(x)) \leftrightarrow \neg Tr(x)])$$

It is easy to check that this formula is false in $\mathfrak{I}^{+}_{ns}$, because, as we noted above, *a* and its NEGATION are TRUE SENTENCES.

iii. DT is "ω-incomplete"

Despite the fact that DT does not imply NEG, it still implies all the "instances" of NEG. That is, for any φ, $\text{DT} \vdash Tr(<\neg\varphi>) \leftrightarrow \neg Tr(<\varphi>)$. The proof is straightforward:

i. $\text{DT} \vdash Tr(<\neg\varphi>) \leftrightarrow \neg\varphi$

ii. $\text{DT} \vdash \neg Tr(<\varphi>) \leftrightarrow \neg\varphi$

iii. Hence, $\text{DT} \vdash Tr(<\neg\varphi>) \leftrightarrow \neg Tr(<\varphi>)$, for any $\varphi \in {}_0\text{L}$.

Now compare this with the notion of ω-incompleteness. A theory T in the language of arithmetic L is ω-incomplete if and only if there is a formula $P(x)$ such that,

i. for all natural numbers *n*, $\text{T} \vdash P(n)$

ii. not:-($\text{T} \vdash \forall x P(x)$)

Obviously, we may adopt an analogous notion for theories that talk about syntactical items, like closed formulas. That is, T is ω-incomplete when there is a formula $P(x)$ such that T always implies $P(<\varphi>)$, for each L-formula φ, but does not imply $\forall x P(x)$. We may summarize the findings of the last three sub-sections thus:

Theorem 3: DT (MT) is "ω-Incomplete".

CHAPTER 5. DEFLATIONISM AND TARSKI'S PARADISE

Proof: DT (MT) implies all the "instances" of NEG, but does not imply NEG. ■

Compare this with Robinson Arithmetic Q:

i. For any natural numbers n, m, Q $\vdash n + m = m + n$.

But Q does not prove the general law of commutativity for addition, namely,

ii. Q does not imply $\forall x \forall y(x + y = y + x)$.

Q is good at (finite) arithmetic, but bad at algebra.[87]

iv. DT and MT do not implicitly define Tr.

A number of authors writing about Tarski's theory have claimed that Convention T (in effect, the infinite list of T-sentences: i.e., the theories DT and MT) "fixes the extension of", or *implicitly defines*, 'true'. Examples of such authors are Quine[88], Haack[89] and Corcoran[90].

Certainly, if we set up two disquotation theories DT_1 and DT_2 governing truth predicates Tr_1 and Tr_2, then since $DT_i \vdash Tr_i(<\varphi>) \leftrightarrow \varphi$, we see that $DT_1 \cup DT_2 \vdash Tr_1(<\varphi>) \leftrightarrow Tr_2(<\varphi>)$, for each $\varphi \in {}_0L$. But it does *not* follow that Tr_1 and Tr_2 are coextensive in $DT_1 \cup DT_2$, as an argument using the Beth Definability Theorem shows:

Theorem 4: DT and MT do not implicitly define *Tr*.

Proof: Assume that (i), the language of DT, *excluding* the predicate *Tr*, contains only apparatus that may be interpreted in the underlying theory T in the object language L (so the syntax of L may be formalized within T in L); and (ii), the theory T in L is a

---

[87] For a detailed presentation of Q see Boolos & Jeffrey 1989, Chapter 14, "Representability in Q".

[88] See Quine 1953b, p. 136: "Nevertheless the [disquotation schemes] resemble definitions in this fundamental respect: *they leave no ambiguity as to the extensions, the ranges of applicability, of the verbs in question.*"

[89] See Haack 1978, p. 100: "The *point* of the (T)-schema is that, if it is accepted, it fixes not the intension or meaning but the *extension* of the term 'true'."

[90] See Corcoran 1997: "One of Tarski's major philosophical insights is that … all Tarskian biconditionals whose right hand side exhaust the sentences of a given formal language constitute an implicit definition of 'true' as applicable to sentences of that given formal language."

consistent extension of Q (so Tarski's Indefinability Theorem applies). Now, assume that DT *implicitly* defines Tr. Then, by the Beth Definability Theorem, there exists an *explicit* definition of Tr in DT, say $\forall x(Tr(x) \leftrightarrow \Psi(x))$. By assumption, Ψ may be translated into the object language. Hence, we should have an explicit definition of Tr in the object language. This contradicts Tarski's Indefinability Theorem. ■

The disquotation theory DT does, however, rather trivially fix the extension of Tr in "standard models". For suppose we add an extra predicate Tr^*, and construct the analogous theory DT*. Let $\mathfrak{I}$ be any L-interpretation. Let **Tr** be the theory of $\mathfrak{I}$ (the set of true L-sentences in $\mathfrak{I}$). A *standard* model of DT is just the expansion ($\mathfrak{I}$, **Tr**). Obviously, if $\mathfrak{I}^+$ is any standard model of DT $\cup$ DT*, then $\mathfrak{I}^+[Tr] = \mathfrak{I}^+[Tr^*]$. For the standard model of DT $\cup$ DT* is just ($\mathfrak{I}$, **Tr**, **Tr**).

5.5 Further Inadequacies of Deflationary Theories

The deflationary theories of truth are, I claim, incomplete accounts of the concept of truth. The Tarskian theory of satisfaction is a much more complete account of our conception of truth. For example Tarski's theory, unlike the deflationary theories DT and MT, yields all the usual theorems expected from a (classical) theory of truth:

for any closed formula φ, $\neg\varphi$ is true if and only if φ is not true

if Σ is a set of true closed formulas, then any deductive consequence of Σ is true.

for any set of closed formulas Σ, if Σ is true then Σ is consistent

... and so on

Consider the attempt to express 'T is true' within T. The statement 'T is true' means,

for any closed formula φ, if φ is provable in T, then φ is true.

Now if 'true' is not definable in T, then we have a problem. But we can make certain progress. Let T be a consistent axiomatizable extension of first-order Peano Arithmetic. Now let φ be any L-sentence, we want to express,

(1) if φ is provable in T then φ is true

Now a provability predicate, $Prov_T$, *is* expressible in T and we can express (1) in the metalanguage L^+ by,

(2) $Prov_T(\ulcorner\varphi\urcorner) \rightarrow Tr(\ulcorner\varphi\urcorner)$

But we can "disquote" the truth-predication and obtain,

(3) $Prov_T(\ulcorner\varphi\urcorner) \rightarrow \varphi$

Any such formula is called a "reflection principle"[91]. Let $Refl_T$ be the *set* of all these reflection principles in L. One might think of this infinite collection of sentences $Refl_T$ as expressing within T the truth of T. However, it is possible to show that, if T is a consistent axiomatizable extension of Peano Arithmetic, then not-(T $\vdash Refl_T$).[92] Actually, $Refl_T$ implies Con_T![93] So T $\cup$ $Refl_T$ is certainly not a conservative extension of T.

But more importantly, one may show that adding DT to T is insufficient to derive a certain metalanguage formula which expresses the "truth of T". First, define the "truth of T", thus, as an L^+-formula:

Tr_T: $\forall x(Prov_T(x) \rightarrow Tr(x))$

Now it is possible to prove the following theorem,

Theorem 5: T $\cup$ DT does not imply Tr_T.

Proof. Assume that T $\cup$ DT $\vdash Tr_T$. Then, T $\cup$ DT $\vdash \forall x(Prov_T(x) \rightarrow Tr(x))$. Then, T $\cup$ DT $\vdash Prov_T(\ulcorner 0 \neq 0\urcorner) \rightarrow Tr(\ulcorner 0 \neq 0\urcorner)$. Then, T $\cup$ DT $\vdash Con_T$. And this is impossible, because, by Gödel's Second Incompleteness Theorem, T does not imply Con_T and DT is conservative. The same holds for the other conservative deflationary truth theories MT and SDT. ∎

---

[91] See Boolos & Jeffrey 1989, p. 283.

[92] Use Löb's Theorem (see Boolos & Jeffrey 1989, p. 187), which says that, if T $\vdash Prov_T(\ulcorner\varphi\urcorner) \rightarrow \varphi$, then T $\vdash \varphi$. So if T implied $Refl_T$, it would have to imply *every* sentence φ, and T would be inconsistent.

[93] Again, use Löb's Theorem: First, $Refl_T \vdash Prov_T(\ulcorner\neg 0 = 0\urcorner) \rightarrow \neg 0 = 0$. Thus, $Refl_T \vdash 0 = 0 \rightarrow \neg Prov_T(\ulcorner\neg 0 = 0\urcorner)$. Thus, $Refl_T + 0 = 0 \vdash Con_T$. Hence, $Refl_T \vdash Con_T$.

Incidentally, it is worth quickly looking at how does this works for Q, which is finitely axiomatized. Of course, we may suppose that Q is a (finite) axiom in L, and then we can "express its truth" in L^+ as a single formula $Tr(\ulcorner Q \urcorner)$. Q also has provability predicates, say $Prov_Q$ and we can also "express it truth" as Tr_Q: the formula, $\forall x(Prov_Q(x) \rightarrow Tr(x))$.

What we can then show is that,

i. $Q \cup DT$ does not imply Tr_Q

ii. $Q \cup DT$ does not imply $Tr(\ulcorner Q \urcorner) \leftrightarrow Tr_Q$.

The proof of (i) is analogous to the above Theorem 5. The proof of (ii) resides in the fact that $DT \vdash Tr(\ulcorner Q \urcorner) \leftrightarrow Q$ (by construction!!). So, $Q \cup DT \vdash Tr(\ulcorner Q \urcorner)$. If $Q \cup DT$ implied $Tr(\ulcorner Q \urcorner) \leftrightarrow Tr_Q$, then it would also imply Tr_Q, which it doesn't by (i). So (ii) is proved.

This further strengthens the case for thinking that the deflationary truth theories really are weak and incomplete theories of truth.

5.6 Tarski's Theory and Gödel Sentences

The non-conservativeness result is very interesting. Let T in L be a consistent axiomatic extension of PA. Then let G_T be a Gödel sentence for T. Now, G_T "says that" G_T is not provable in T. So, G_T is true if and only if G_T is not provable in T. This can be formalized within T. Indeed, $T \vdash G_T \leftrightarrow \neg Prov(\ulcorner G_T \urcorner)$. There is such a sentence, by the Diagonal Lemma (or Fixed Point Theorem).

Theorem 6: Provability of Gödel Sentences

> Let T be a consistent axiomatic extension of PA. Let G_T be a fixed point of $\neg Prov_T$. Then $T \cup TS \vdash G_T$.

Proof: By a basic property of TS (Tarski 1936, Theorem 5), $T \cup TS \vdash Tr_T$. Thus, $T \cup TS \vdash Prov_T(\ulcorner \varphi \urcorner) \rightarrow Tr(\ulcorner \varphi \urcorner)$, for all $\varphi \in {}_0L$. And, $T \vdash G_T \leftrightarrow \neg Prov_T(\ulcorner G_T \urcorner)$. Thus, $T \cup$

$\mathsf{TS} \vdash \mathit{Prov}_\mathsf{T}(\ulcorner \mathsf{G_T} \urcorner) \rightarrow \mathit{Tr}(\mathsf{G_T})$. By "disquotation", $\mathsf{T} \cup \mathsf{TS} \vdash \mathit{Prov}_\mathsf{T}(\ulcorner \mathsf{G_T} \urcorner) \rightarrow \mathsf{G_T}$. Thus, by simple logic, $\mathsf{T} \cup \mathsf{TS} \vdash \mathsf{G_T}$. ■

We can certainly "recognize" that a Gödel sentence $\mathsf{G_T}$ for T is true (on the assumption that T *is* true), but our knowledge of its truth does *not* obtain from correct formal derivations *within* the theory to which it applies. For example, one way of recognizing the commutativity of addition on natural numbers (i.e., the truth of the formula $\forall x \forall y(x + y = y + x)$) is to assume that each axiom of PA is true and to derive $\forall x \forall y(x + y = y + x)$), using the induction scheme (twice), from these axioms of PA. But this does not work for a Gödel sentence $\mathsf{G_{PA}}$. For $\mathsf{G_{PA}}$ is not a consequence of PA (if PA is consistent, etc.).

How then do we "recognize the truth" of $\mathsf{G_T}$? According to an argument associated with Lucas 1961 and, more recently, Penrose 1989, this recognition involves some kind of non-computational "insight"[94]. Although I (like them) am inclined to *disagree* with the computational theory of mind, I think they are wrong on this matter, for:

> $\mathsf{G_T}$ is *deducible* from the strengthened theory: T *plus the standard Tarskian theory of truth* for the language of T.

Theorem 6 above can be deformalized as follows. We have the Fixed Point Theorem:

> Hyp_1. T implies that $\mathsf{G_T}$ is true if and only if $\mathsf{G_T}$ is not provable in T

plus the "Generalized Equivalence Principle":

> Hyp_2. T + TS implies that T is true

Then we proceed as follows:

(1) T + TS implies that, for any φ, if φ is provable in T then φ is true

(2) T + TS implies that if $\mathsf{G_T}$ is provable in T then $\mathsf{G_T}$ is true

(3) T + TS implies that if $\mathsf{G_T}$ is not true then $\mathsf{G_T}$ is true

(4) T + TS implies that $\mathsf{G_T}$ is true

(5) T + TS implies $\mathsf{G_T}$

[94] See Penrose 1989, Chapter 4, 'Truth, Proof and Insight'.

In any case, remember that a deflationary theory of truth cannot achieve such "insight" (i.e., reasoning). It is conservative, so T $\cup$ "deflationary theory" does *not* imply G_T. Indeed, Hyp_2 fails for DT. If I am right, our ability to recognize the truth of Gödel sentences involves a theory of truth (Tarski's) which *significantly transcends the deflationary theories.*

To summarize, an adequate theory of truth looks like it must be *non*-conservative. Indeed, it is bound to be non-conservative if it satisfies the generalized "Equivalence Principle" above. Tarski's theory does the job nicely. But the deflationary theories are conservative. So they are inadequate.

5.7 Tarski 1936 Revisited

Some of the technical material presented above appeared, in slightly different clothing, in Tarski's famous 1936 essay, 'The Concept of Truth in Formalized Languages'.

i. The Conservativeness of DT

In the later (often unread by philosophers) sections of Tarski 1936, we read:

> THEOREM III: if the class of all provable sentences of the metatheory is consistent and if we add to the metatheory the symbol 'Tr' as a new primitive sign, and all the theorems that are described in conditions (α) and (β) of the Convention T as new axioms [i.e., all the "disquotational T-sentences"], then the class of all provable sentences in the metatheory is consistent.
>
> (Tarski 1936 (1956), p. 256).

This theorem is certainly implied by the conservativeness of DT. For if T $\cup$ DT is a conservative extension of a consistent theory T, then T $\cup$ DT must also be consistent. Actually, Tarski's Theorem III implies conservativeness also. Theorem III says that, for *any* consistent theory T in L, T $\cup$ DT is consistent. So, if T $\cup$ DT is inconsistent, then so is T. Suppose that T $\cup$ DT $\vdash \varphi$, where φ is an L-sentence. Then, T $\cup$ $\{\neg\varphi\}$ $\cup$ DT is inconsistent. By Theorem III, T $\cup$ $\{\neg\varphi\}$ must be inconsistent. Thus, T $\vdash \varphi$.

ii. The "ω-Incompleteness" of DT

Shortly after Tarski's introduction and proof sketch of Theorem III, we read:

> The value of the result is considerably diminished by the fact that the axioms mentioned in Theorem III [i.e., the axioms of DT] have a very restricted deductive power. A theory of truth founded on them would be a *highly incomplete system*, which would lack the most important and most fruitful theorems.
>
> (Tarski 1936 (1956), p. 257).

To illustrate this, Tarski discusses the formula, $\neg Tr(x) \vee \neg Tr(neg(x))$. He first points out (in effect) that DT proves $\neg Tr(<\varphi>) \vee \neg Tr(neg(<\varphi>))$, for any L-sentence φ. He then points out that the universal closure of this formula, $\forall x(\neg Tr(x) \vee \neg Tr(neg(x)))$, which he calls the Law of Non-Contradiction, is not a theorem of DT. He discusses this, along with a proposed "Rule of Infinite Induction", or an ω-rule, on pages 257-261.

iii. DT Does Not Implicitly Define *Tr*

Having noticed the refutation, using the Beth Definability Theorem, of the claim that Convention T fixes the extension of 'true', I discovered that Tarski had already said pretty much the same:

> Thus it seems natural to require that the axioms of the theory of truth, together with the original axioms of the metatheory, should constitute a categorical system. It can be shown that this postulate coincides in the present case with another postulate, according to which *the axiom system of the theory of truth should unambiguously determine the extension of the symbol* 'Tr' which occurs in it, and in the following sense: if we introduce into the metatheory, alongside this symbol, another primitive sign, e.g., the symbol 'Tr'', and set up analogous axioms for it, then the statement 'Tr = Tr'' must be provable. *But this postulate cannot be satisfied.* For it is not difficult to prove that in the contrary case the concept of truth could be defined exclusively by means of terms belonging to the morphology of the language, which would be in palpable contradiction with Theorem I [Tarski's Indefinability Theorem].
>
> (Tarski 1936 (1956), p. 258. Emphasis added).

It is interesting that Tarski's proof, which he does not give explicitly, involves a similar argument to the Beth Definability Theorem, which was not in fact proved until later (Beth 1953).

CHAPTER 5. DEFLATIONISM AND TARSKI'S PARADISE

iv. The Non-Conservativeness of the Full Tarskian Theory of Truth

Finally (Tarski 1936, pp. 274-276), Tarski discusses what amounts to our *non-conservativeness result*. In particular, the provability of undecidable-in-T Gödel sentences in the overall truth-theoretic metatheory T $\cup$ TS. He writes:

> The definition of truth allows the consistency of a deductive science to be proved on the basis of a metatheory which is of higher order than the theory itself. On the other hand, it follows from Gödel's investigations that it is in general impossible to prove the consistency of a theory if the proof is sought on the basis of a metatheory of equal or lower order. Moreover, Gödel has given a method for constructing sentences which—assuming the theory concerned to be consistent—cannot be decided in any direction in this theory. *All sentences constructed according to Gödel's method possess the property that it can be established whether they are true or false on the basis of the metatheory of higher order having a correct definition of truth.* Consequently, it is possible to reach a decision regarding these sentences, i.e., they can be either proved or disproved.
>
> (Tarski 1936 (1956), p. 274. Emphasis added).

Tarski then discusses Gödel's method of obtaining a sentence G

> ... which satisfies the following condition: G is not provable if and only if *p*, where '*p*' represents the whole sentence G.

Tarski then goes on to show that this sentence G is "actually undecidable and at the same time true". He concludes,

> By establishing the truth of the sentence G we have eo ipso . . . also proved G itself in the metatheory. . . . the sentence G which is undecidable in the original theory becomes a decidable sentence in the enriched theory.
>
> (Tarski 1936 (1956), p. 276)

5.8 Conclusion: Deflating Deflationism

It seems to me that if the result of adding a higher level theory T_2 to some lower level theory T_1 yields new theorems, expressible in the language of the lower level theory T_1 but not derivable in T_1, and we have reasons for thinking that these extra theorems are themselves true, then T_2 could not be considered *dispensable* in favour of T_1 (or somehow redundant). The non-conservativeness results above show that adding the full theory of satisfaction for L to a theory T in L, need not yield a conservative extension. And we have reasons for thinking that the extra theorems are themselves true.

Part of the basic (not necessarily deflationist) idea about truth is that a particular statement φ and its "truth" $Tr(<\varphi>)$ are somehow "equivalent". I think this is correct (indexicals aside), and if a truth theory satisfies Convention T then it proves the equivalence. But we must go further. Any *adequate* theory of truth should be able to prove the generalized "equivalence" of a possibly infinitely axiomatized *theory* T and its "truth" Tr_T (that is, the metalanguage formula $\forall x(Prov_T(x) \rightarrow Tr(x))$. And Tarski's theory comes up trumps. It is possible to show that with Tarski's theory of satisfaction TS, one has,

i. $T \cup TS \vdash Tr_T$

However, the preceding arguments indicate that with the deflationary theories one has,

ii. $T \cup$ deflationary truth theory does not imply Tr_T

The "recognizability" of Gödel sentences further emphasizes these points. Indeed, the ability to "see" that G_T and Con_T are in fact *true* is a fundamental element of understanding the significance of Gödel's Incompleteness Theorems: *we* can "see" that G_T and Con_T are true, even though the consistent axiomatic theory T itself cannot prove them. There are *truths* that cannot be *proved*. Truth and proof come apart[95]. We have seen that Tarski's theory of truth helps explain this phenomenon. Suppose we accept a standard axiomatization of arithmetic (PA, say). It seems correct to say that we also (implicitly) accept its truth, and thus we surely then think that it is consistent. Tarski's Indefinability Theorem tells us that we cannot define a truth predicate for L in the language L of PA, but we can (using Gödelian techniques) express the consistency of PA in L. So, having *accepted* PA, we think it's true, and we seem to be committed to thinking it consistent. But the consistency of PA is not deducible from the axioms of PA,

[95] Boolos & Jeffrey 1989, p. 180: "And perhaps the most significant consequence of [Gödel's First Incompleteness Theorem] is what it says about the notions of *truth* (in the standard interpretation of the language of arithmetic) and *theoremhood*, or provability (in any particular formal theory): that *they are in no sense the same*." Music to Platonist ears: the Gödelian overture to realism.

by Gödel's Second Incompleteness Theorem. Nevertheless, the consistency of PA *is* a true statement if PA is consistent. How do we "know it"? What we have shown is that by adding a strong enough *theory of truth* (the theory of satisfaction for the language of arithmetic), we can deduce the truth of PA (i.e., Tr_{PA}) and hence the consistency of PA (i.e., Con_{PA}) from this truth-theoretic strengthening of PA. However, we have also shown that the deflationary theories of truth are powerless to achieve this deduction, for they are conservative (anything derivable with them is derivable without them).

To summarize:

(1) The deflationary theories DT, MT and SDT are conservative;

(2) DT and MT are "ω-incomplete";

(3) Neither DT nor MT implicitly defines 'true';

(4) The standard Tarskian theory of truth/satisfaction TS is non-conservative;

(5) $T \cup TS \vdash Tr_T$;

(6) $T \cup$ deflationary theory does not imply Tr_T;

(7) $T \cup TS$ implies Gödel sentences: $T \cup TS \vdash Con_T$; $T \cup TS \vdash G_T$.

To sum up, *if* "deflationism about truth" is construed as incorporating the following claims,

Conservativeness Claim (implicit in the idea of redundancy)

> The result of adding a truth theory to a collection of non-truth-theoretic assumptions yields a *conservative extension.*

Completeness Claim (explicit in Horwich's minimalism):

> Deflationary theories of truth constitute "all there is to truth",

then deflationism is false. Formalized theories based on the deflationary conception of truth are incomplete with respect to our prior grasp of the "truth about truth". If I am right, there is simply *more to truth* than is expressed by the deflationary truth theories.

In the introduction I hinted at a close analogy between the indispensability of *mathematics* and the indispensability of a substantial *theory of truth.* Field's deflationary programme aims to show that mathematical theories (like standard set theory) are convenient fictions: ultimately redundant and (in principle) dispensable.

As we shall see in Chapter 9, this deflationary programme for mathematics founders on the non-conservativeness of adding mathematics to "mathematics-free" nominalistic theories, like the quasi-Newtonian theory of the gravitational field in Euclidean space-time presented in Field 1980. I have argued that *Tarskian truth theory* is in some way analogous (N.B., a Tarskian truth theory contains set-theory). So I would like to conclude by suggesting that this analogy between the indispensability of (Tarskian) *theories of truth* and the indispensability of *mathematical theories* deserves more intensive investigation. In the meantime, *no-one will drive us out of Tarski's truth-theoretic paradise!!*

CHAPTER 6

Mathematical Truth

> It is my contention that two quite distinct kinds of concerns have separately motivated accounts of the nature of mathematical truth: (1) the concern for having a homogeneous semantical theory in which the semantics for the propositions of mathematics parallel the semantics for the rest of language, and (2) the concern that the account of mathematical truth mesh with a reasonable epistemology. It will be my general thesis that almost all accounts of the concept of mathematical truth can be identified with serving one or the other of these masters *at the expense of the other.*
>
> Paul Benacerraf 1973 (1983), 'Mathematical Truth', p. 403.

> And perhaps the most significant consequence of Theorem 6 [Gödel's First Incompleteness Theorem] is what it says about the notions of *truth* (in the standard interpretation of the language of arithmetic) and *theoremhood*, or provability (in any particular formal theory): that *they are in no sense the same.*
>
> George Boolos & Richard Jeffrey 1989, *Computability and Logic*, p. 180.

6.1 Semantic Monism and the Benacerraf Dilemma

We argued in Chapter 4 that subjectivist (epistemic, internalist or anti-realist) truth theories lead to an unacceptable relativism. I have suggested that disquotational T-sentences are (in some sense) analytic or trivially true, primarily due to the conservativeness of their "totality" (one needs to be careful to avoid semantic paradox, imposing a restriction that the truth predicate *Tr* may not appear in the formulas of which truth is predicated). However, we argued in Chapter 5 that such deflationary theories of truth, consisting simply of the (restricted) totality of such sentences, are inadequate as a theory of truth. Such theories are weak in several senses: in particular, they are demonstrably *conservative* over theories in the object language, while an adequate truth theory must be *non*-conservative in order to correctly represent *truth-theoretic reasoning*. E.g., an adequate truth theory should be able to prove that if an object level theory **T** is true, then it is consistent; but if the truth theory could do this, it would have to be non-conservative. Tarski's theory does in fact represent this reasoning correctly. (The deflationary truth theories DT and MT are weak in further senses: they do not imply the generalized "laws of truth", and do not implicitly define truth).

By elimination that leaves as candidate theories of truth:

i. Tarski-style truth theories,

a. A recursive definition of a referential relation (like satisfaction) for the recursively specifiable formulas of the object language,

b. Truth defined in terms of this referential relation,

c. Satisfy Convention T, by proving all the T-sentences as theorems;

ii. Russell-style correspondence theories,

Truth defined as a "correspondence relation" between each sentence and a "fact".

In this section, I want to examine a doctrine I shall call (*standard Tarskian*) *semantic monism* and its relation to the problem of "epistemic access" explained by Benacerraf 1973. Semantic monism is the doctrine that the standard semantic conception of truth and satisfaction (or some suitable generalization thereof to handle, say, indexicals) is applicable to all pieces of linguistic discourse. In particular, semantic monism entails that the standard semantic conception of truth applies readily to *mathematical discourse*, both pure and applied. We shall say that someone who advocates a non-semantic conception of "truth" for mathematical sentences is advocating a *non-standard conception of mathematical truth.*

Actually, unbridled semantic monism would be the thesis that we should apply one uniform semantic theory for all linguistic discourse: mathematics, natural science, psychology, ethics, aesthetics, poetry, etc. *Tarskian semantic monism* is the thesis that the theory of truth required is Tarski's (or a suitable extension, for indexicality, modality, etc.).

Consider the "logico-semantic structure" of the following sentences:

(1) There is a mountain between Manchester and Leeds

(2) There is a prime number between 10 and 12

The prima facie "logical regimentations" of these statements are respectively:

$(1)^r$ $\exists x(\text{mountain}(x) \wedge \text{between}(x, \text{Manchester, Leeds}))$

$(2)^r$ $\exists x(\text{prime-number}(x) \wedge \text{between}(x, 10, 12))$

Notice that both are instances of the formal schema:

(3) $\exists x(F(x) \wedge G(x, a, b))$

The semantic analysis of (1) and (2) requires that in both $(1)^r$ and $(2)^r$, the quantifier $\exists$ should be interpreted as a referential quantifier and that the predicates and variables be treated in the usual Tarskian way. Arbitrary n-place predicates are *true of* n-tuples of objects; the *extension* of an n-place predicate P is the class of n-tuples of which P is *true*; any valuation σ *satisfies* the sentence $P(t_i, t_j, ..., t_n)$ just in case $(\sigma[t_i], \sigma[t_j], ..., \sigma[t_n]) \in$ the extension of P.

The alleged problem with a standard Tarskian treatment of mathematical truth, explaining truth in terms of satisfaction (reference), is that it forces mathematical statements to be construed as *descriptive* of a realm of (presumably abstract) *mathematical objects* (i.e., the values of variables). Some authors refer to this as "objects-platonism". The problem, if there is one, with objects-platonism is that it might be difficult to see how facts about such abstract objects could be *known*. The *locus classicus* for this alleged epistemological problem is a celebrated article by Paul Benacerraf.

In his 1973, Benacerraf imposes two conditions upon an "acceptable over-all account of mathematical knowledge and truth":

> The first component of such an over-all view is more directly concerned with the concept of truth. We can state [the first component] as the requirement that there be an over-all theory of truth in terms of which it can be certified that the account of mathematical truth is indeed an account of mathematical *truth*. The account should imply truth conditions for mathematical propositions that are evidently conditions of their truth (and not simply, say, of their theoremhood in some formal system). This is not to deny that being a theorem of some system can be a truth condition for a given proposition or class of propositions. It is rather to require that any theory that proffers theoremhood as a condition of truth also *explain the connection between truth and theoremhood.*

> Paul Benacerraf 1973 (1983), 'Mathematical Truth', p. 408.
>
> Another way of putting this first requirement is to demand that any theory of mathematical truth be in conformity with a *general theory of truth*—a theory of truth theories, if you like—which certifies that the property of sentences that the account calls 'truth" is indeed truth . . . [This] amounts only to a plea that the semantical apparatus of mathematics be seen as part and parcel of that of the natural language in which it is done. . .
>
> I suggest that, if we are to meet this requirement, we shouldn't be satisfied with an account that fails to treat ['there are at least three large cities older than New York'] and ['there are at least three perfect numbers greater than 17'] in parallel fashion . . . *I take it we have only one such account: Tarski's, and that its essential feature is to define truth in terms of reference (or satisfaction) on the basis of a particular kind of syntactico-semantical analysis of the language.*
>
> (*op. cit.*, p. 408. Emphasis added)

Accounts of mathematical truth which violate the constraints imposed are simply not accounts of truth. According to semantic monism, the appropriate concept of truth for the sentences of a mathematical theory is the semantic conception of truth as explicated by Tarski 1936.

Now why should this create a problem? As we noted above, the application of Tarskian semantical theory to mathematics theories leads us to construe the statements of such theories as purporting to describe a realm of mathematicalia. (In particular, if any existential statements are true, then mathematicalia must exist).

Benacerraf goes on in his 1973 article to argue that the implication of the existence of (presumably abstract) mathematicalia, proceeding via the theory of truth, leads to the problem of our *knowledge* of such entities:

> ... accounts of truth that treat mathematical and nonmathematical discourse in relevantly similar ways do so at the cost of leaving it unintelligible how we can have any mathematical knowledge whatsoever;
>
> (Benacerraf 1973 (1983), pp. 403)

This is the Benacerraf puzzle of "epistemic access" to mathematical objects. If mathematics is a body of known truths about a realm of abstract mathematicalia, and knowledge acquisition is a matter of *causal interaction with our surroundings*, then how is knowledge of such abstracta possible? The puzzle is sometimes expressed as,

The Benacerraf Dilemma:

> Either our truth-theoretic *semantics* is deeply flawed or our "causal" *epistemology* is deeply incomplete.

This dilemma has been one of the most widely discussed problems in the philosophy of mathematics in the past twenty-five years. Although several philosophers of mathematics (notably Chihara, Kitcher and Field) take the dilemma as a strong argument against realism, it is noteworthy that Benacerraf himself does not advocate nominalism.

I shall try to briefly summarize the three most prominent proposals for resolving the problem of epistemic access along lines consistent with Tarski's theory and mathematical realism:

i. Gödel, Maddy and Mathematical Intuition

The first proposal derives from some remarks of Gödel 1944, 1947 and concentrates on a proposed epistemic faculty of "mathematical intuition", whereby mathematical truths are forced upon the mind of the mathematician:

> ... the assumption of [sets] is quite as legitimate as the assumption of physical bodies and there is quite as much reason to believe in their existence. They are in the same sense necessary to obtain a satisfactory system of mathematics as physical bodies are necessary for a satisfactory theory of our sense perceptions.
>
> (Gödel 1944 (1983), pp. 456-7).
>
> ... the objects of transfinite set theory ... clearly do not belong to the physical world and even their indirect connection with physical experience is very loose (owing primarily to the fact that set-theoretical concepts play only a minor role in the physical theories of today).
>
> But, despite their remoteness from sense experience, we do have something like a perception also of the objects of set theory, as is seen from the fact that *the axioms force themselves upon us as being true*. I don't see why we should have less confidence in this kind of perception, i.e., in mathematical intuition, than in sense experience, which induces us to build up physical theories and to expect that future sense experiences will agree with them and, moreover, to believe that a question not decidable now has meaning and may be decided in the future.
>
> (Gödel 1947+1964 (1983), pp. 483-4. Emphasis added)[96].

[96] Gödel 1947+1964 is a revision to Gödel 1947. It appears in full in Benacerraf & Putnam (eds.) 1983.

The most well-known advocate of a version of this proposal is Maddy 1980, 1990. Maddy does not argue that pure mathematicalia are themselves perceivable. Rather, she argues that it is possible to perceive *impure* mathematicalia, such as sets of (or structures composed of) concrete urelements. Suppose you open the fridge, and there before you is a carton of eggs. Standard set theory with urelements assert the existence of a *set* of these eggs. Now, you can see the *eggs* all right, for photons emitted by the fridge light are reflected from the eggs into your eyes, stimulating the rods and cones in your retinas, and sending electrical information to the brain, which results, presumably causally, in the brain somehow "assenting" to the propositional message "Eggs ahoy". According to Maddy, you can perceive the *set* of eggs, as well as the eggs. The idea is that *impure* mathematicalia (i.e., sets containing concrete urelements) have *physical properties* and are *causally active*.

I find this implausible. What, for example, is the *optical reflectivity* of this *set*? Just how does this *set* transfer energy and momentum to the photons whereby it is perceived? Standard physics doesn't attribute physical properties to the *set* of eggs, and it might be argued that Maddy's proposal is simply inconsistent with the sorts of physical mechanisms postulated by conventional physics.[97, 98]

ii. Neo-Fregeanism and the Theory of Abstraction

The second proposal involves from Frege's theory of abstraction[99]. The example Frege gives concerns the existence and properties of *directions* of lines. The "abstraction principle" is that, if A and B are lines, then,

the direction of A = the direction of B just in case A and B are parallel.

---

[97] To be fair, however, Maddy does discuss in detail psychological work on perception, and it does seem to be little more than a *truism* to say that human perceivers are able to perceive simple patterns and structures, concretely exemplified.

[98] On the Indispensability Argument, Maddy recognizes that mathematics is indispensable for science, but argues that the use of mathematics within science is typically "*idealizing*", and that, because we are not therefore committed to the literal truth of such idealized theories, the argument for realism is undercut. See Maddy 1992, 1996. I disagree. The continuity, and indeed, full manifold structure, of spacetime, for example, is *not* thought of as an idealization in serious physics. See Earman & Norton 1996.

[99] See Frege 1884, §§ 65-68.

Since parallelism $\|$ is an equivalence relation, this principle is in fact derivable if we set $dir(A) =_{df} [A]_{\|}$.

According to the Neo-Fregean position advocated by Wright 1983 and Hale 1987, we can know that there is such an abstract entity as *the number of eggs* (that is, the abstract object #(egg)), because statements such as,

> #(egg) = #(left-hand fingers) just in case (the eggs are 1-1 correlated with left-hand fingers)

are *analytically true*, and provide implicit definitions of the singular terms '#(egg)' and '#(left-hand fingers)', which designate the appropriate abstract objects. Such statements are instances of Hume's Principle, which we mentioned briefly in Chapter 2. Neo-Fregeans argue that we possess *a priori* and indefeasible knowledge of such "abstraction-introducing" principles. Obviously, such principles are comprehension principles, asserting the existence of certain entities. This would be more convincing if it weren't the case that the analogous principle for sets (or extensions) is actually inconsistent, by Russell's paradox. Instead, standard Zermelo-Fraenkel set theory proceeds in a more piece-meal fashion, asserting the existence of certain sets, and ways of generating "new" sets (e.g., via the power set operation).

iii. Quine's Holism

In his classic articles, "On What There Is" (1948) and "Two Dogmas of Empiricism" (1951), and later work, Quine set out a comprehensive epistemological position which might be called (covering all bases) *pragmatic conjectural naturalistic holism.*

Quine's position offers an epistemology which sharply distinguishes, as all realists should, between the ontology of a theory T and the ontology of the evidence e for that theory. Indeed, there need be no simple connection (and certainly no *causal* connection) between the entities reified by a theory or conjecture T and the objects reified in the observational evidence e for that theory. Even if evidence statements typically quantify over "middle-sized observable concrete bodies" and events, that is simply *no restriction* on what theoretical statements may quantify over (and such statements may count as justified or confirmed by their corroborating evidence). Trivial examples include conjectures about *future events* (e.g., the return of Halley's comet), and more sophisticated examples might involve hidden space-time singularities inside causally inaccessible black holes[100]. These entities and events are clearly not *causally responsible* for our beliefs about them. The same argument transfers to our conjectural knowledge of abstracta like numbers and classes[101].

It is worthwhile looking a little more closely at Quine's epistemology. On this position, theory-formation and concept-formation does not involve a naïve empiricistic copying into the mind of "sensations", which are supposed to be *caused by*, and to be representations of, objects in the local observable environment. Instead, Quine emphasizes a process of *cognitive growth*: language acquisition, concept formation, theory construction, and the full-blown hypothetico-deductive scientific method involves the successive incorporation of more and more sophisticated conceptual apparatus, beginning with holophrastic observation sentences ("Mama!", "Dog!"), tightly conditioned to stimulation, passing through the emergence of predication ("Milk is hot"), truth functions ("If there is milk, then there is food") and culminating, relatively early on, in the introduction of objectual quantification over abstractions.[102]

On this conjectural holistic view, mathematicalia (and abstract entities generally) are "on a par" with electrons and photons, in being hypothetical or theoretical entities postulated by evolving scientific conjecture. Indeed, because Quine rejects the notion of immediate knowledge of a "given" object, he views *all objects as theoretical*[103].

---

[100] No light signal emitted inside a black hole horizon can leave it. If there are such singularities and event horizons in spacetime, we certainly do not know about them by their "causing" us to know about them. This reminds me of a joke by Putnam: "What is the *dominant cause* of our beliefs about electrons? Textbooks, of course".

[101] Along similar lines, Hart 1979 argues that Quine's epistemology solved the Benacerraf problem *avant la lettre*.

[102] Quine's speculations are set out at length in Quine 1974 and more succinctly in Quine 1995.

[103] See Quine 1981b. N.B., this is not intended to be idealism or anything like that. That would be to confuse epistemology with metaphysics, as Quine stressed in his Foreword to Quine 1980. Properly speaking, it is the *concepts* 'electron' and 'dog' that are (epistemologically) on a par.

In fact, this is supported by a well-known experimental fact, discussed in any decent textbook on child development or cognitive growth[104]. An early conjecture (or discovery) of the developing child is the "object-concept"[105], the discovery that *objects still exist when not perceived*[106].

Quine, as always, argues there is no fundamental difference between high-level conjectures about theoretical entities like electrons and low-level childish conjectures about the behaviour of toys and Mama. The fact that high-level conjectures involve explanations of effects (like lightning, rusting and so on) by the postulation of atoms, electrons and molecular forces indicates no difference *in kind*[107]. The 10 month old child learns to "pad" his or her system with unobserved toys, while the 24 year old physicist "pads" his or her system of the world with atoms, quarks, space-time points, forces and force fields, numbers, functions, classes, and so on.

A further component of Quine's position is a repudiation of naïve empiricism in favour of *pragmatism.* Our primary drive in this activity is not solely evidential or based on "theory-neutral" observations[108]. Rather, our conceptual growth is driven by *pragmatic considerations* of simplicity, convenience, elegance, explanatory power, and so on. Quine discusses a mathematical example: we "round out" our theory of integers with a theory of interspersed ratios; and, again, we round out our "gappy" theory of ratios with a

---

104 E.g., Flavell 1985.

105 This is Piaget's terminology. It is sometimes called the "object-persistence" concept and sometimes even called the "object-independence" concept.

106 The experimental demonstration of this is (I am told) a common undergraduate psychology experiment. The child is presented with a toy, and one checks whether the child understands or expects that the toy *is still there* even if it is hidden from view behind an obstacle or screen. It transpires that one can experimentally demonstrate the normal child's conceptual transition from "Berkeleyan idealism" to "Lockean realism" in the first year of development. (So much for Berkeley's *a priori* philosophical proof of immaterialism)!

107 This is connected to the so-called "Piaget-Chomsky" debate in development theory. Chomsky argues that the development of mental faculties such as the object-concept (and, indeed, general language acquisition) is biological and innate, while the Piagetians are silent on the matter. In contrast, Chomsky points out that knowledge of high-level physical principles is not innate. This is why learning a language is "easy" and learning high-school physics is "hard" (Chomsky thinks that mathematics, at least the simple parts of counting and geometry, lies in between). It is a question of the way the brain is designed.

108 Quine argues in his 1951a and later articles that there can be no hope of observational reductionism, of a phenomenalistic reduction of science to an evidential foundation of "remembered similarity" between sense data, or anything like that. This is one of "Dogmas" of empiricism.

theory of densely interspersed irrationals; and finally we pad our theory of reals with a theory of complex numbers, obtaining full algebraic completeness: every polynomial has a root. Similarly, we interpolate a smooth curve between plotted data points, and extrapolate our curves into the unknown.

6.2 Arguments for Semantic Monism

Benacerraf is sharply aware that any non-standard conception of mathematical truth leaves it unexplained how to connect the obtaining of non-standard conditions (e.g., derivability in an uninterpreted axiom system) with the *truth* of mathematical assertions:

> ... whereas those [accounts of mathematical truth] which attribute to mathematical propositions the kind of truth conditions we can clearly know to obtain, do so at the expense of failing to connect these conditions with any analysis of the sentences which shows how the assigned conditions are conditions for their *truth*.
>
> (Benacerraf 1973 (1983), pp. 403-4).

Is there an argument for treating the sentences of mathematics and physics on a par, semantically? One argument is this. Tarski, following Frege and Russell, originally developed standard "referential semantics" for *mathematical languages*. Out of this arose model theory. The theories to which model theory were applied were, in the beginning, always mathematical theories, like arithmetic, analysis, algebra, geometry and set theory. Hence, the discovery of non-standard models of arithmetic and analysis, as well as non-standard models of set theory. For example, the Löwenheim-Skolem theorem tells us that first-order set theory ZFC has a non-standard denumerable model $(D, \in)$: such a model has a *denumerable* domain D, and hence could not contain "all sets" constructible from this domain, in the intuitive sense, since the power set of D, by Cantor's Theorem, is non-denumerable; and thus $\in$ could not mean membership (in the intuitive sense). Only *later* were Tarski's ideas were taken up as a more or less correct way to do semantical theory by philosophers of *language*, like Davidson and McDowell, and applied to non-mathematical discourse about trees and milk.[109]

[109] See, e.g., Davidson 1967 and other papers anthologized in Davidson 1984.

The argument then is that Tarski proposed his theory primarily to explicate the concept of truth for *mathematical discourse*[110], and only afterwards was this theory applied to non-mathematical discourse[111]. It would be odd to abandon this eminently workable and standard theory of truth for extremely dubious philosophical reasons.

In any case, it seems to me that a stronger argument for semantic monism is this. In Chapters 2 and 3 we emphasized that we need to recognize the presence of, and analyse the logico-semantic properties of, *mixed* predicates and axioms in scientific theories. Standard mathematicized physical theories are expressed in languages containing polyadic mixed predicates simultaneously true of *physical* and *abstract mathematical* entities.

Quine discusses exactly this aspect of applied mathematics in his 1986 "Reply to Charles Parsons":

> Another example of applied mathematics is the use of number in measurement. In terms of physical testing procedures we describe a Fahrenheit temperature function whose arguments are place-times and whose values are real numbers. Fahrenheit temperature is a class of pairs of pure real numbers and concrete place times. Similarly distance in metres is a class of triples, each comprising one pure real number and two concrete localities.
>
> Mathematical objects and concrete objects are thus in perpetual interplay, participating in the same triples and pairs. Mathematical vocabulary and empirical vocabulary are in perpetual interplay, participating in the same sentences. We see this already at the most primitive level of applied mathematics, when we say that there are fifty people in this room: the pure abstract number, fifty, is how many concrete people there are in this concrete room. I see pure mathematics as an integral part of our system of the world.
>
> (Quine 1986, p. 398).

E.g., the dyadic predicate of thermodynamic theory,

(the real number) *r* is the temperature-in-°C of (the concrete body) *x*

---

[110] Tarski's main fully worked-out example in his 1936 is the first-order language of classes, with $\subseteq$ as primitive. Definition 22 contains the recursive definition of satisfaction and Definition 23 contains the explicit definition of truth in terms of satisfaction.

[111] Tarski 1936 famous expressed sceptical doubts about extending his ideas to "colloquial language". These problems have been widely discussed by Davidson and his followers, who propose that a theory of truth for a natural language can function as a theory of meaning for that language. A discussion of Davidsonian truth-theoretical semantics would take us too far afield. A good exposition is Platts 1979. For opposition, see Blackburn 1983.

or, more concisely,

$T_{\circ C}(x, r)$,

or a mixed predicate used in theorizing about the nature of time,

(the real number) *r* is a co-ordinate of (the temporal instant) *t*

It would be odd to say the least to treat the primary variable '*t*', which ranges over physical bodies or regions of space-time, referentially and the secondary variable '*r*', ranging intuitively over real numbers, in some non-standard non-referential way.

Actually, there is a proposal in the literature to do something like this: to treat '*r*' as a *substitutional variable*, and treat any quantifier that binds it as a substitutional quantifier. This approach might be labelled "substitutionalism"[112]. It cannot work for mathematicized theories generally, mainly for the reason that it is inadequate to the needs of real analysis. By Cantor's diagonal argument, for any standard language L with denumerably many expressions, there are more real numbers than numerals in L for real numbers[113].

The important issue here concerns the dual application of standard Tarskian semantics to physics and mathematics. Perhaps there would be little cause for even regarding mathematical sentences as meaningful, let alone true, were it not for the applicability of mathematics. If mathematics were (*per impossibile*) entirely autonomous as an activity, then a severe version of formalism would be appropriate and, perhaps, quite attractive. Mathematics could then simply be regarded as the manipulation of meaningless physical symbols. The arrangement of such symbols might "echo" those of sentences in interpreted language, but there would be few constraints on such manipulation. "Consistency" of a formal system would simply mean (nothing more than) that, for any "sentence-like" symbol φ, the pair of sentence-like symbols φ and $\neg\varphi$ never turn up in such manipulations. Treating such sentence-like symbols as interpreted statements

[112] An approach like this, based on substitutional quantification, is perhaps implicit in some of the work of Lesniewski. It is an explicit proposal in Parsons 1971 and Gottlieb 1980.

[113] Non-standard languages may, of course, be studied, with a distinct constant a_r for every real *r*. Such languages are, of course, *abstract mathematical objects*, and thus of no discernible use to a nominalist.

would, on such a formalistic view, be perverse; and asking whether they are true or false would be equally perverse. In short, if mathematics were not applicable, then mathematics could be construed just as Hilbert and Wittgenstein thought it should: a "language game" with only internal, conventional rules.

But this supposition is, of course, are wrong. Mathematics *is* applicable. It remains puzzling that Hilbert himself never recognized this problem, for he made major contributions to theoretical physics[114].

Finally, there is a burden of proof argument that the onus is surely on the person who thinks that standard semantic theory should be *replaced* (by some non-standard semantic theory) to actually *provide* such a novel, non-referential, semantic theory. In particular, this theory must deal adequately with the troublesome mixed predicates and axioms. In the absence of a non-standard semantics, the nominalist who wants to argue that standard existential assertions of mathematics are *not* true faces a basic problem. As we have seen, the referential apparatus of satisfaction will force us to conclude that if these existential mathematical statements are true, then there must exist certain entities for these truths to be truths about. But we think of mathematicized statements of certain scientific theories like relativity and quantum theory as truthful (or at least approximate truthlike) descriptions of the natural world. Mathematical Platonism is just a consequence of our antecedently accepted natural science.

6.3 The Standard Conception of Mathematical Truth

The techniques that Tarski introduced for theorizing about truth (simpliciter) are perfectly applicable to any standard formalized notation L. The trick is to introduce a metalanguage ML which extends the object language L by adding a new satisfaction predicate *Sat* and the membership predicate $\in$, and which contains axioms (TS)

[114] Especially in the treatment of action functionals and their use in the derivation, via a Least Action Principle, of physical laws: the action for General Relativity is called the "Einstein-Hilbert Action". Note: his work on abstract Hilbert spaces predated its application by Weyl, Jordan, von Neumann and others within quantum mechanics.

formulated in ML that recursively define the new satisfaction predicate *Sat* with respect to the syntactic structure of object language expressions.[115]

Thus, suppose that we are formulating the semantics[116] of some mathematical theory T. Let L be the object language in which T is formulated. Let *Tr* be a truth predicate for L in L^+, some sufficiently powerful metalanguage which extends L. The standard theory of satisfaction TS in L^+ for L derives a closed formula $Tr(<\varphi>) \leftrightarrow \varphi$, for each sentence φ of L. Of course, we cannot require that *Sat* or *Tr* be *definable* (in T) in the object language L. But we know that for any of the usual mathematical languages (in particular first-order and higher order extensional languages), we can introduce a Tarskian theory of satisfaction for L, TS, which does contain an explicit definition of *Tr* and derives all the T-theorems (one for each object language sentence, not each L^+ sentence; otherwise, by the Diagonal Lemma we should derive a contradiction: typically *Tr* is not definable in the base theory T in the object language L)[117].

Now consider a thinker contemplating mathematical assertions. Suppose he or she accepts $Tr(<\varphi>)$, where φ is some existential mathematical statement. For example, like almost anyone, he or she will say,

(1) the theorem 'there exist infinitely many prime numbers' is *true*.

[115] In general, it will also be necessary to introduce within ML new predicates and functors, like *Sen*(*x*), *Form*(*x*), *neg*(*x*), *conj*(*x*, *y*), *var*(*x*, *i*), etc., governed by appropriate axioms, for discussing the syntax of L.

[116] By 'semantics' I mean the theory of *truth simpliciter*, not the theory of "truth in a model". To be sure, model-theoretical semantics is important. But to say that T has a model (is satisfiable) is not to say that T is true. Ptolemaic Astronomy, Newtonian Mechanics, and aromatherapy all have models, but are not true.

[117] This is even possible when the object language is *set-theoretical* and the axioms governing $\in$ are, say, ZFC. In this case, Tarski's Indefinability Theorem ensures that one cannot convert the recursive definition of *Sat* to a direct definition (one can do this when $\in$ does not appear in the object language). But even so, the theory of satisfaction TS is well-defined in the metalanguage, and it derives every T-sentence for the object language as required by Convention T (i.e., formulas not containing *Sat* or *Tr*). In the set-theoretical case, Tarski's Theorem makes it clear that "satisfaction is untranslatable foreign language" (Quine 1970, p. 45). This is a very strong reason for not confusing *truth simpliciter* with the relational notion of "truth in a model". The relation, "*x* is true in model *M*" is fully definable in ZFC. But no set-theory can define the intended interpretation of the language of set theory: the interpretation in which the formula $x_i \in x_j$ is satisfied by a sequence s iff $s_i \in s_j$.

But if he or she accepts this truth claim, then he or she will be forced to accept φ itself, for we assume that he or she also accepts the Tarski T-theorem $Tr(\langle\varphi\rangle) \leftrightarrow \varphi$. So he or she will accept,

(2) there exist infinitely many prime numbers

Now φ itself will usually possess direct ontological commitment to abstract entities. That is, (2) implies,

(3) there are mathematicalia

Finally, suppose that this person claims to be nominalist. By assumption, he or she will reject the existential claim $\exists x Math(x)$. Consequently he or she must reject φ (since φ implies $\exists x Math(x)$). Thus he or she accepts,

(4) there are no mathematicalia

This is clearly a contradiction.

Now there is a version of nominalism, strongly hinted at by Chihara 1990, according to which,

> We *should* accept the truth claim $Tr(\langle\varphi\rangle)$, when φ is, say, a theorem of standard mathematics (e.g., Euclid's Theorem of the infinity of primes).

For example, Chihara writes:

> This is just the sort of development one would expect on the hypothesis that mathematics is a *system of truths* and mathematicians are attempting to arrive at *truths*.
>
> (Chihara 1990, p 171. Emphasis added)

> Many of these mathematical beliefs have been checked and rechecked countless times, and in countless ways, by both sophisticated and elementary methods . . . We thus have some strong reasons supporting the belief that *mathematics is a body of truths*.
>
> (Chihara 1990, pp. 172-173. Emphasis added)

We shall discuss this briefly below, as the "hermeneutic" conception of mathematical truth. (We shall discuss Chihara's overall modal nominalism later, in Chapter 8).

According to this conception of mathematical truth, we should accept:

(5) Euclid's Theorem is true

But, as nominalists, we should *not* accept:

(6) there are infinitely many prime numbers

for this statement logically implies,

(7) there are numbers.

which nominalists insist we cannot accept.

But it is surely a triviality that we all should accept the following instance of the disquotation schema:

(8) 'there are infinitely many primes' is true if and only if there are infinitely many primes

Now clearly, this sort of nominalist is suggesting that we accept (5) but we should reject (7). The problem is that (7) is a logical consequence of (5) and (8). And the disquotational T-sentence (8) is a triviality. Thus, such a nominalist position is logically inconsistent. How can anyone suggest we should accept *A* and *B*, which imply *C*, but reject *C*?

It seems to me that a coherent nominalist must reject (5), the truth claim concerning Euclid's Theorem. Thus, someone like Chihara simply cannot argue that mathematics is a body of truths and also say that the disquotational truth conditions of its theorems do not hold.

One (trivial) way a nominalist can reject the truth claim is simply by replacing the predicate 'true' (which has been tightly constrained by the disquotation axioms and Tarski's full theory of satisfaction) by some new unexplained phrase, '*mathematically true*'. That is, the nominalist proposes the quite radical idea that the standard Tarskian semantical conception of truth is somehow "inappropriate" for mathematical statements. According to this nominalist idea, no existential mathematical theorem φ is true (in the standard classical sense). Instead, any existential mathematical theorem φ is merely "mathematically true", although not true. Let us call this new predicate of mathematical sentences *M-Tr*, to be read, intuitively, "mathematically true" (and whose *meaning* is yet to be specified!!).

Such a Tarski-repudiating nominalist must then suggest that we have to break the Tarskian disquotational link between $M\text{-}Tr(<\varphi>)$ and φ. That is, the nominalist rejects (some of) the troublesome equivalences of the form:

$$M\text{-}Tr(<\varphi>) \leftrightarrow \varphi$$

Indeed, the whole motivation for this kind of nominalist is to achieve a coherent situation wherein he or she could consistently accept $M\text{-}Tr(<\varphi>)$ without accepting φ. But this means that *M-Tr* has lost its claim to being a *truth predicate.* This nominalist strategy involves introducing a non-classical conception of "mathematical truth", radically different in its logical behaviour from truth:

i. for truth simpliciter, we must accept every instance of:

$$Tr(<\varphi>) \leftrightarrow \varphi$$

Rejecting these equivalences amounts either to simply assigning a *deviant meaning* to the word 'true', or to misunderstanding the concept of truth. For the concept of truth is conceptually grounded in the disquotational T-equivalences. Furthermore, the totality of such instances (that is, the theory DT) is really quite harmless: it is conservative, and so *alone* it implies nothing but logical truths (in the base language), and even if added to another base theory implies nothing new.

ii. for "mathematical truth" we are free to accept some instances of,

$$M\text{-}Tr(<\varphi>) \wedge \neg\varphi,$$

Quite clearly, the nominalist must urge that $M\text{-}Tr \neq Tr$.

However, we need to ask what this new predicate *M-Tr* *means.* There are two accounts of "mathematical truth", different from and incompatible with the standard Tarskian semantic conception of truth.

6.4 Non-Standard Mathematical Truth I: If-Thenism

One proposal for analysing "mathematical truth" is that it is nothing more than derivability in some axiom system M.[118] This proposal is popular with those minimalists who assume that mathematics is *nothing more* than *theorem proving* (or the deductive study of *uninterpreted* formal axiom systems)[119]. Indeed, this proposal is closely connected to a version of formalism known as *if-thenism*. An early statement of if-thenism appears in Russell,

> Pure mathematics is the class of all propositions of the form '*p* implies *q*', where *p* and *q* are propositions containing one or more variables, the same in the two propositions, and neither '*p*' nor '*q*' contains any constants except logical constants.
>
> (Russell 1903, p. 3)

To this suggestion, Quine commented:

> . . . all that is left to the mathematician, for him to be right or wrong about, is whether various of his uninterpreted sentence schemata follow logically from his uninterpreted axiom schemata. All that is left to him is elementary logic, the first-order predicate calculus.
>
> (Quine 1978a (1981), p. 149).

For example, the following would presumably be a proposition of such "pure mathematics":

'$F(x)$' implies '$\exists y(F(y) \wedge y = x)$'[120].

A slight generalization of this idea is that what a mathematician discovers when he or she "discovers the truth" of some statement φ is simply that,

φ is formally derivable from the mathematical axioms M[121]

---

[118] But which one? The reply "Any consistent extension of quantifier-free arithmetic" will not do. For there are incompatible systems M_1 and M_2 which are both consistent extensions. Since they are incompatible, at least one must be false (simpliciter). Why do we prefer the axiom system which implies that the ratios are dense? There is certainly an "axiom system" containing the primitive 'x is a rational number' which implies the theorem '$\forall x_{Rat}\exists y_{Rat}\neg\exists z_{Rat}(y \neq x \wedge z$ is between x and $y)$', which says that the ratios are not dense. A realist will say that the ratios really are dense, and that this axiom system is *false*. The usual axioms are *truths* about the ratios or about the "rational number structure" (i.e., any countable dense linear ordering without endpoints: all of these are isomorphic, by Cantor's "back-and-forth" proof).

[119] As Paul Erdös once put it, "a mathematician is a machine for turning coffee into theorems".

[120] For *us*, of course, this is just a *metalogical* theorem.

That is, the existence of a formal derivation Γ of φ from M.[122] In this sense, the "assertibility" of the statement φ does not consist in, or derive from, its truth. Rather, when a mathematician occasionally "asserts" φ, what he or she "really" means to assert is rather the (validity of the) conditional M → φ. For he or she knows, by a chain of metalogical reasoning, that this conditional is true (indeed, logically true) just in case M ⊢ φ, and this is what he or she has proved.

The central advantages of this proposal concerning "mathematical truth" are these:

i. *ontological innocence*;

ii. *epistemological innocence*.

For example, the statement

(1) There is a prime number between 100 and 1000

could be "mathematically true" even though there were no prime numbers. The statement (1) is "mathematically true" because it is *derivable* from the axioms of first-

[121] We must suppose, if our purposes are nominalistic, that M is a single axiom. For we want to say that the sentence φ is derivable from the sentence M. We cannot (at this stage) suppose that φ is derivable from M, which is permitted to be a *set* of axioms, for that would contravene the nominalistic repudiation of mathematicalia. However, three points must be made:

First, if M is a (possibly infinite) recursively specifiable set of first-order axioms, and M ⊢ φ, then the very notion of proof (i.e., its finitistic nature) ensures that there is a *single axiom* M* such that M* ⊢ φ (that is, there is a finite subset of M axioms which derives φ, and this may be conjoined into a single axiom M*). Of course, if the "axioms" of M are, say, *all the truths of first-order arithmetic*, then no such M* exists.

Second, if M is a recursively specifiable set (possibly infinite) of axioms, M = {φ in L: *F*(φ)}, we can explain M ⊢ φ as follows: φ is derivable from *some of the F*s, and this can be expressed (in the metatheory) using Boolos' monadic second-order plural quantification (thus not referring to some *subset* of *F*s).

Third, we can represent any axiom scheme in, say PA or ZFC, as a single *second-order axiom*. Thus, by moving to a second-order formulation, we can express the antecedents for our naïve if-thenist conditionals as a single (second-order) axiom.

[122] Again, the nominalist will have problems with the status of Γ. Is this a mathematical object, a certain sequence of formulas, as standard proof theory asserts? Or is it an actual *inscription*, as the nominalist requires? If the nominalist construes 'M ⊢ φ' to mean that a proof φ has actually been inscribed, then this will simply falsify some of the standard principles of proof theory (e.g., The Deduction Theorem: if M, φ ⊢ χ then M ⊢ φ → χ). This is the basic reason why some of those who wish to "nominalize" metalogic have proposed that *modal notions* being introduced, so 'M ⊢ φ' means that φ *could* be derived from M.

order Peano Arithmetic PA[123]. In short, "mathematical truth" in this sense will not require a realm of abstract mathematical objects for mathematical statements to be about.

Furthermore, if an arithmetic assertion φ is "mathematically true", we can in principle (and quite easily in this case) *discover* or *verify* that φ is "mathematically true". We simply construct a formal proof of φ (a refutation of $\neg\varphi$) from some finite subset of the axioms of PA. More generally, suppose the background axiom system is M. Suppose that φ is derivable from M. So, the assertion $M \rightarrow \varphi$ is a logical truth. Its "knowability" (in principle) is assured (at least in the first-order scenario)[124]. Since there is a complete proof procedure for first-order logical truth, all such logical truths can be routinely checked by a Turing Machine[125]. So, if mathematical truths are "really" all logical truths of the form $M \rightarrow \varphi$, then it is at least possible in principle to obtain "epistemic access" to such truths: formal deduction is enough, without recourse to empirical experience[126]. In

---

[123] I doubt whether anyone has actually performed this boring derivation. The if-thenist has to actually *write down the inscription before* he or she can legitimately assert that (1) follows from PA. To say that he or she knows that it "could be done" would be to *use* standard platonistic proof theory to support a *modal* conclusion. Not very nominalistic.

[124] The Completeness Theorem for first-order logic says that if M semantically implies φ, then there exists a derivation Γ of φ from M. It places no finite bound, or upper level of complexity, on this derivation. It is perfectly conceivable that there are consequences φ of certain axioms M such that no-one will ever discover (by formal derivation) that φ follows from M. That is, the complexity of the proof may exceed human powers of computation (although it does not exceed the Platonic idealization of such powers, as encoded in the notion of a Turing Machine).

[125] Of course, the Completeness Theorem is *proved* using mathematics (what Machover 1996 calls the "ambient set theory"). The coherent nominalist should really produce a new proof not using any mathematics (as Goodman & Quine 1947 valiantly attempted), or refrain from asserting this mathematical fact. Of course, most self-professed nominalists adopt the usual "double-think", asserting the Completeness Theorem and then denying the means of its proof.

[126] Of course, some meta-mathematical results, like Gödel 1931, Gödel 1940 and Cohen 1963, have the form: φ is not derivable from M. Of course, the if-thenist might say that Cohen still proved a *logically-true conditional*, namely, if ZFC is true, then ZFC does not imply CH. The problem is that this cannot be logically true. If it were, then ZFC would imply "ZFC does not imply CH". Thus, ZFC would imply its own consistency, and this is ruled out by Gödel's Second Incompleteness Theorem. Now every mathematician accepts that CH is not derivable from ZFC, but the simple fact is that it requires the assumption of a *stronger set theory* to construct the model of ZFC where CH fails.

There is a view that what Gödel "really" proved in 1931 is this: *if* PA is consistent, then Con is not a theorem of PA. Nonsense! Gödel proved that Con is not a theorem of PA. (No one doubts that PA is consistent). The reason this view is so pernicious is illustrated by an analogy. My wife says, at dinner, "Where's the margarine?". I reply, "*If* an external world exists, and *if* there is no wicked Cartesian demon, or a Putnamian scientist, and *if* the laws of Nature are still the same, and ..., *then* it's in the fridge". Not only would this be profoundly pretentious sophistry, it simply wouldn't reflect what I actually believe to be the truth, which is that the margarine is in the fridge. (We make no pretences at *certainty* in our assertions).

this sense, mathematical knowledge reduces to "*a priori*" logical knowledge, although of course, we do not suppose (as the logicists did) that the concepts invoked in the axiom M can be defined in purely logical terms and that M itself is an *a priori* (analytic) logical truth.

In the second-order scenario, matters are worse. For any sound formalization of deductive reasoning for a second-order notation, there will be logical truths (relative to the model theory) which are not provable in the formalism. A simple example is $PA^2 \rightarrow Con_{PA}$, which is true in all models: any model of PA^2 is isomorphic to $\mathfrak{N}$ and Con_{PA} holds in any such model, and so Con_{PA} is a semantic (logical) consequence of PA^2. But, of course, Con_{PA} is not *deducible* from PA^2, by Gödel's Second Incompleteness Theorem. It follows that, on the if-thenist analysis of "mathematical truth", there are still mathematical truths (logically true conditionals) which cannot be derived. (In order to recognize or prove the metalogical fact that $PA^2 \models Con_{PA}$, one needs to assume (as true) a set theory strong enough to characterize the intended model of PA^2 and to prove things about isomorphisms and consequences. This is a repudiation of if-thenism, unless one applies if-thenism to the metatheory! That is, one believes something like the conditional $ZF^2 \rightarrow$ "$PA^2 \models Con_{PA}$").

Actually, it is hard to find a canonical textual defence of if-thenism[127], although it is sometimes defended by others in conversation. A recent very brief description of "if-thenism" appears in Hellman 1989:

> Consider [naïve non-modal if-thenism]. Suppose it represents sentences φ of arithmetic by means of a material conditional, say, of the form:
>
> $PA^2 \rightarrow \varphi$
>
> or some refinement thereof.
>
> (Hellman 1989, p. 26).

Hellman then proceeds to present what he calls the "canonical objection" to if-thenism, an objection which is actually quite mistaken:

> Suppose also that, in fact, there happen to be no actual ω-sequences, i.e., that the antecedent of these conditionals is false. ... Then, automatically, the

[127] However, see Putnam 1967a, 'The Thesis That Mathematics is Logic'.

> translate of every sentence φ of the original language is counted as *true*, and the scheme must be rejected as wildly inaccurate. (Well, at least it gets half the answers right—not the worst imaginable performance! Compare the case of the broken watch).
>
> (Hellman 1989, p. 26).

For example, suppose (as the nominalistically-inclined if-thenist actually maintains) that there are no such things as numbers. Then the axiom PA^2 for second-order Peano Arithmetic is false. But then, obviously, for *any* sentence φ of arithmetic, the reconstrual $PA^2 \rightarrow \varphi$ will be true, simply because the antecedent is false. Does this criticism work?

I think that this criticism is mistaken. For the adequacy constraint implicitly required by the naïve if-thenist "translation" is not that $M \rightarrow \varphi$ be *true* (simpliciter). The constraint is that if φ is a theorem in M, then $M \rightarrow \varphi$ should be *logically true.* We know, merely from the soundness of $\vdash$, that this constraint is satisfied. For suppose that $M \vdash \varphi$. Then, by the deduction theorem, we infer that $\vdash M \rightarrow \varphi$. And then by the soundness of $\vdash$, we infer that $\models M \rightarrow \varphi$, and thus that $M \rightarrow \varphi$ is a logical truth.

(The converse argument only works in the first-order case. By the definitions of logical truth and semantical implication $\models$, we know that if $\models M \rightarrow \varphi$, then $M \models \varphi$; and we then infer, by the completeness of first-order derivability $\vdash$, that $M \vdash \varphi$.)

In any case, according to the if-thenist, the *truth value* of the mathematical axiom M is actually irrelevant. All that matters is that if $M \rightarrow \varphi$ has the appropriate *formal property* (logical truth, validity) if φ is derivable from M. In particular, the naïve if-thenist with nominalistic intentions is going to claim that M is in fact *not true* (or even, is perhaps meaningless: for M may be regarded as uninterpreted), but, even so, mathematical practice is faithfully reconstructed. For mathematical practice, according to the naïve if-thenist, is exhausted by *theorem-proving* and mathematical knowledge consists in the metalogical knowledge that φ is derivable or not derivable from M.[128]

---

[128] Field 1984a makes it clear that the logical knowledge that φ is *not* derivable from M presents a problem for if-thenism. It is equivalent to knowledge of *consistency* and this is not accounted for within naïve if-thenism. Normally, one proves consistency by constructing a model (e.g., the finite ordinals for PA). (Because there is an algorithmic proof procedure for *inconsistency*, there is no problem. But, by Gödel's and Church's theorems, there is no such proof procedure for consistency). The if-thenist might "ascend" a step

A problem does arise for incomplete axiom systems (or the incompleteness of the second-order derivability relation). But this is separate. The naïve if-thenist can simply rest with the claim that the "mathematical truth value" of an allegedly undecided sentence (e.g., Choice, the Axiom of Determinacy (which contradicts Choice), Cantor's Continuum Hypothesis, Gödel's Axiom of Constructibility, the existence of large cardinals) is simply undetermined by current mathematical practice. This is a very unsatisfactory position, but it might be maintained. For to determine "truth values" of (current) undecidables, new axioms must be added to decide these undecidables. Why accept these axioms and not others?

In any case, the *falsity* of M is not an objection to the adequacy of the naïve if-thenist reconstrual. Actually, the reconstrual is deeply inadequate, but one can hardly fault the nominalistically-inclined if-thenist for supposing something which he or she obviously holds (namely that the axiom M is, when construed literally, not true).

6.5 Problems for If-Thenism

There are several reasons for thinking that if-thenism is unacceptable. A preliminary argument is that mathematicians, in asserting a mathematical theorem φ, do not in fact *intend* to assert merely a conditional of the form $M \rightarrow \varphi$. The mathematician does not mean, in asserting the existence of a homomorphism from SU(2) to SO(3), that the proposition expressing this is merely deducible from ZFC. He or she means that *there is such a homomorphism.*

However, it might be possible to teach this mathematician some of the profound discoveries of the philosophy of mathematics (in particular, the two Benacerraf problems, the "identification problem" and "epistemic access"). His or her natural reaction, without knowing the further Gödelian, Quinian and Putnamian arguments for realism, might be to descend to an if-thenist or formalist position.

---

further by saying that this sort of knowledge is actually knowledge of further "logical truths" of the form: if S is true, then φ is not a theorem of M (where S is some suitably strong axiomatic set theory).

However, this view (and others like it) suffer from tremendous problems. First, this non-standard conception of "mathematical truth" does not conform to Convention T. It is a *non-standard* conception of truth. If the nominalist is going to propose some axioms governing *M-Tr*, they certainly must not entail the T-sentence,

(1) '$\exists x(\text{number}(x) \wedge \text{prime}(x))$' is *M-Tr* if and only if there are prime numbers

Indeed, the nominalist will assert something inconsistent with this, namely:

(2) '$\exists x(\text{number}(x) \wedge \text{prime}(x))$' is *M-Tr* and there are no prime numbers

Now suppose the nominalist proposal is that whenever a mathematician asserts,

(3) φ is true

we are to "reconstrue" this assertion as,

(4) φ is "mathematically true"

where "mathematically true" means "derivable in axiom system M'.

This is, to put it bluntly, ignorant. Gödel's theorem *implies* that proof and truth are quite different concepts. There is no formal axiomatic system which proves all the truths of elementary arithmetic.

Let us introduce the "if-thenist" reconstrual mapping, Γ_M: $\varphi \mapsto \Gamma_M(\varphi) = M \rightarrow \varphi$. We shall say that Γ_M is *literal* with respect to an intended interpretation $\mathfrak{I}$ just in case, for any sentence φ in L, if $\mathfrak{I} \models \varphi$ then $M \vdash \varphi$. It is possible to prove that no such literal "if-then reconstrual" exists for arithmetic.

Let M be an extension of Robinson Arithmetic Q, so M is strong enough to represent all recursive functions, including the diagonal function. Then Gödel's First Incompleteness Theorem will apply to M, and this theorem entails the non-existence of a literal translation function with respect to $\mathfrak{N}$, the intended interpretation of arithmetic.

Theorem 1: Literal If-thenist Reconstruals Γ_M Do Not Exist.

Let M be a finitely axiomatized extension of Robinson Arithmetic, Q in L, the language of arithmetic. Let $\mathfrak{N}$ be the intended interpretation of L. Then the "if-thenist" reconstrual function w.r.t M. does not have "literalness".

Proof: First, let $Prov_M$ be a provability predicate for M and let G_M be a Gödel sentence with respect to $Prov_M$ for the axiom M. That is, by the Diagonal Lemma, $M \vdash G_M \leftrightarrow \neg Prov_M(\ulcorner G_M \urcorner)$. So, M implies that "$G_M$ is true if and only if it is not provable in M". We know, of course, that $\mathfrak{N} \models G_M$ and, by the First Incompleteness Theorem, not-($M \vdash G_M$). Thus, not-($\vdash M \rightarrow G_M$). Thus, not-($\vdash \Gamma_M(G_M)$). Thus, there is a φ such that $\mathfrak{N} \models \varphi$ but not-$\vdash \Gamma_M(\varphi)$. Thus, Γ_M does not have "literalness". Thus, a "literal" if-thenist translation function cannot exist for any sound finitely axiomatized extension of Q. ■

Of course, the above theorem may be viewed as irrelevant. We are implicitly assuming that M is *true* (that is, true in the intended interpretation $\mathfrak{N}$ of L). We then recognize that G_M is similarly true, but not provable from the axiom M. But the if-thenist has a reply: he or she does not have to admit that φ is true just in case $\Gamma_M(\varphi)$ is a logical theorem. In short, this kind of formalistic if-thenist will simply reject (as irrelevant) the "literalness" condition on the translation function Γ_M.

However, it is not hard to see that any *identification* of truth with derivability within an axiom system is incoherent, as the following theorems show.

Definition 1: Deductivist "Mathematical Truth"

Let T be axiomatic mathematical theory. Then, for any mathematical sentence φ, φ is *mathematically true* w.r.t. T iff $T \vdash \varphi$.

If T is strong enough, this can be formalized within T itself. When T is an extension of PA, suppose that $Prov_T$ is a provability predicate for T. Define within T a predicate $M\text{-}Tr_T$, to be read "mathematically true in T":

$$\forall x(M\text{-}Tr_T(x) \leftrightarrow Prov_T(x))$$

Now we can prove two "inequivalence theorems":

Theorem 2: Only an Inconsistent Theory Can Equate "Deductivist Mathematical Truth" with Truth Simpliciter

Let T be an axiomatic extension in L of PA. Let $M\text{-}Tr_T$ be the deductivist "mathematical truth in T" predicate (i.e., a provability predicate for T). Let DT be the disquotation truth theory for L. Now suppose that

$$T \cup DT \vdash \forall x(Tr(x) \leftrightarrow M\text{-}Tr_T(x)),$$

Then T is inconsistent.

Proof: Let φ be any sentence of L. From our assumptions we infer that $T \cup DT \vdash Tr(\ulcorner\varphi\urcorner) \leftrightarrow Prov_T(\ulcorner\varphi\urcorner)$. By the "disquotation principle" for DT, we infer $T \cup DT \vdash \varphi \leftrightarrow Prov_T(\ulcorner\varphi\urcorner)$. Now, by the Conservativeness Theorem for DT, we infer that $T \vdash \varphi \leftrightarrow Prov_T(\ulcorner\varphi\urcorner)$. Thus, $T \vdash Prov_T(\ulcorner\varphi\urcorner) \rightarrow \varphi$. Hence, by Lob's Theorem, $T \vdash \varphi$. Thus, T is inconsistent. ■

Theorem 3: Deductivist "Mathematical Truth" is not Truth Simpliciter

Let T be an axiomatic extension in L of PA. Let $M\text{-}Tr_T$ be $Prov_T$ (i.e., a provability predicate for T). Let TS be the full Tarskian truth theory for L. Then,

$$T \cup TS \vdash \exists x(Tr(x) \wedge \neg M\text{-}Tr_T(x)),$$

That is, $M\text{-}Tr_T \neq Tr$.

Proof: By the Diagonal Lemma, one may construct a fixed-point sentence G_T such that $T \vdash G_T \leftrightarrow \neg Prov_T(\ulcorner G_T\urcorner)$. From the Non-Conservativeness Theorem for TS we know that $T \cup TS \vdash G_T$. So, $T \cup TS \vdash Tr(\ulcorner G_T\urcorner)$. But $T \cup TS \vdash G_T$ and $T \cup TS \vdash G_T \leftrightarrow \neg Prov_T(\ulcorner G_T\urcorner)$. Thus, $T \cup TS \vdash \neg Prov_T(\ulcorner G_T\urcorner)$. Thus, $T \cup TS \vdash \neg M\text{-}Tr_T(\ulcorner G_T\urcorner)$. Hence, $T \cup TS \vdash Tr(\ulcorner G_T\urcorner) \wedge \neg M\text{-}Tr_T(\ulcorner G_T\urcorner)$. Hence, $T \cup TS \vdash \exists x(Tr(x) \wedge \neg M\text{-}Tr_T(x))$. ■

The deductivist conception of "mathematical truth" is simply different from truth simpliciter. Of course, this is to be expected. The nominalist will believe that 'There are infinitely many primes' is "mathematically true" (that is, derivable within the usual axioms of arithmetic) even though he or she will believe that it is *not true* simpliciter. But what these arguments show is that the nominalist cannot blithely say 'true' (as applied to a mathematical assertion) just *means* 'provable'. Truth and provability are different concepts.

Nevertheless, on the alternative assumption that mathematical truth is genuine truth (that is, truth in the intended model of arithmetic), there can be no reconstrual scheme mapping every arithmetical *truth* to a valid theorem of first-order logic.

The above objection operates within pure mathematics. It is essentially equivalent to the claim that "mathematical truth cannot consist in provability":

> Elementary number theory is the modest part of mathematics that is concerned with the addition and multiplication of whole numbers. Whatever sound and usable rules of proof one may devise, some truths of elementary number theory will remain unprovable; this is the gist of Gödel's theorem. . . .
>
> We used to think that mathematical truth consisted in provability. Now we see that this view is untenable for mathematics as a whole, and even for mathematics in any considerable part; for elementary number theory is indeed a modest part, and it already exceeds any acceptable proof procedure.
>
> (Quine 1978b (1981a), p. 144).

> And perhaps the most significant consequence of Theorem 6 [Gödel's First Incompleteness Theorem] is what it says about the notions of *truth* (in the standard interpretation of the language of arithmetic) and *theoremhood*, or provability (in any particular formal theory): that *they are in no sense the same*.
>
> (Boolos & Jeffrey 1989, p. 180).

It seems to me that if an if-thenist were to claim that "mathematical truth consists in provability", then he or she is simply abusing the word 'true' (assigning a deviant meaning to a word which has a well-defined meaning, as explained by Tarski 1936, and which is formalized within the "Theory of Satisfaction", TS). The debate is terminological, but deeply misleading. *All of the philosophical issues in the philosophy of mathematics centre on whether mathematical assertions (or theories) are true* and it is misleading (given the quite clear Tarskian sense given to the notion of truth and the well-known Gödelian logical results demonstrating the difference of these concepts) to propose an "identification" of truth with provability.

To conclude, the if-thenist can say that he or she is *not interested* in whether arithmetic or set theory is true (i.e., not interested in the philosophy of mathematics). But he or she cannot, without contradicting Tarski, Gödel, et al., say that 'true' *means* 'provable'.

6.6 The Refutation of If-Thenism

Setting aside some further (actually important) reservations about this approach within the philosophy of mathematics (primarily, the choice of "uninterpreted" axioms in the first place; are the axioms true?), we shall now see why the if-thenist reconstrual does not work for mathematical principles integrated within scientific theory. Whatever its merits as a philosophical interpretation for mere theorem proving, if-thenism is utterly incapable of working for any mathematical theory which occurs as a component of an accepted mathematicized scientific theory. Since arithmetic, analysis and a certain amount of set theory are certainly such components of many scientific theories, it is simply incoherent to maintain if-thenism for these mathematical theories. In short, the proposal is inconsistent with the judgement as true of the mathematical statements of science.

Consider again our stand-by mathematicized scientific theory $\mathsf{Tim_S}$, whose sole mixed axiom is:

(1) the (impure) structure (Tim, Bef) is isomorphic to $(\mathbf{R}, <)$

This theory $\mathsf{Tim_S}$ is a theory which experts on space-time theory believe to be *true* (or at least, an exceedingly accurate approximation to the truth). As textbooks on space-time put it, "time is isomorphic to the continuum". Moreover, $\mathsf{Tim_S}$ is not meant to be *necessarily* true, or simply a theorem of *pure* mathematics. Indeed, $\mathsf{Tim_S}$ has contingent non-mathematical consequences. For example,

(2) between any two instants, there is another

and thus *cannot* be a consequence of any applicable mathematical axiom system (like mathematical analysis plus set theory with individuals).

Now let us drop the 's' suffix (indicating the presence of set theory within Tim). To simplify things, suppose we reconstrue the use of mathematical analysis within Tim in terms of some set-theoretical surrogates (so quantification over real numbers is replaced by quantification over sets of some kind) and let ZF^2 be the formalization of second-order Zermelo-Fraenkel set theory with ur-elements (that is, ZF^2 is now a single axiom). Let us say that ZF^2 is *quasi-pure*, for it quantifies over and talks about concreta.

The central problem is this. Tim is obviously *not derivable* from ZF^2. Indeed, Tim, although explicitly mathematicized, is a contingent law about the *physical* world. But it is possible to show that the only non-mathematical facts derivable from ZF^2 are logical truths. But Tim has numerous non-mathematical consequences, so it could not be derivable from ZF^2 (unless ZF^2 is inconsistent, of course).

Moreover, Tim is part of accepted physical science. Arid sceptical puzzles aside, it is accepted as true (it is part of the result of applying Fine's "core position"). We have seen that its acceptance cannot consist in its being *derivable* from (impure, applicable) set theory with individuals. Indeed, we naïvely think that Tim is part of an (at-worst approximately) *true description* of the nature of non-mathematical time.

Furthermore, we can quickly see that the if-thenist reconstrual of Tim is not an adequate replacement for Tim. For Tim implies (2), but the if-thenist reconstrual $ZF^2 \rightarrow$ Tim does *not* imply (2). Why is this? The negation of (the single axiom) ZF^2 certainly does not imply (2). So, there is a model of $\neg ZF^2$ in which (2) is false. Hence, there is a model of $\neg ZF^2 \vee$ Tim in which (2) is false[129]. The if-thenist reconstrual $ZF^2 \rightarrow$ Tim does not reproduce the mathematics-free theorems of Tim.

A similar argument works for *any* mathematicized law of accepted physics. Suppose that PL is Poisson's Law, governing the electrostatic potential (the laplacean ∇^2 on the potential Φ is proportional to the charge density ρ, at each space-time point e). That is,

(3) for any space-time point e, $\nabla^2\Phi(e) = \lambda\rho(e)$

[129] This is just trivial logic. If A has a model then so does $A \vee B$ (i.e., because $A \vdash A \vee B$).

The simple fact is that most working physicists accept PL as a true (or highly truthlike) description of (part of) the electromagnetic field. But again, a conditional $ZF^2 \rightarrow PL$ is clearly not a suitable "reconstrual" or replacement for PL.

In general, where T is a mathematicized theory of nature and M some (quasi-pure) mathematical axiom, the if-thenist reconstrual $M \rightarrow T$ is not an *adequate replacement* for T. In short, there are *true* mathematical theories (namely, those of science) which are not "mathematically true" in the if-thenist sense (namely, logical consequences of some uninterpreted system of mathematical axioms).

For a further example, consider Schrödinger's Law:

SL: for any system S at time t, $H\Psi(S, t) = (ih/2\pi)\partial\Psi(S, t)/\partial t$.

Schrödinger's Law has the following properties:

i. according to current science, SL is true (or at least *approximately* true).

ii. SL is a mathematicized statement (quantifying over state vectors, etc.)

iii. SL is also a contingent physical statement (it is not *necessary* that there is such an abstract state vector Ψ associated with any physical system).

But SL is not "mathematically true" on the deductivist conception of "mathematical truth". We know from a theorem given by Field in his 1980 (a sort of Conservativeness Theorem for applicable mathematics) that SL is not derivable from any system of axioms for applicable mathematics. But if SL is true, as many physicists seem to think, then (as we argued in Chapters 2 and 3) there must exist such mathematicalia as state vectors (e.g., elements of the Hilbert space, the quantum mechanical state space for physical systems).

In short, this anti-realist conception of "mathematical truth" is irrelevant and inadequate to the central use of mathematics in our scientific world view. It plays precisely no role in accounting for the *truth* of mathematicized laws like SL and countless other mathematicized laws in science. For, according to standard accepted current science,

i. SL *is true* (or at least approximately true), but

iv. SL is not "mathematically true" (derivable in any standard uninterpreted mathematical axiom system, say applicable set theory).

It seems to me that any account of "mathematical truth" that allows certain substantial mathematical statements to be true but not "mathematically true" is barely worth considering. Indeed, mathematical truth is not even coextensive with "derivability in some axiom system", unless those axioms are already true.

Consider set-theoretical axioms like pairing, infinity, power set, union, separation and so on. Scientific theories like General Relativity would naturally incorporate these axioms if the use of analysis were reinterpreted along standard lines. It logically follows that if you think that GR is true, you must think that the *axiom of infinity and so on are true.*

If-thenism is either *irrelevant* (if it makes no negative claim about the truth value of these axioms) or *false* (if it claims, for example, that the axiom of infinity is not true).

6.7 Non-Standard Mathematical Truth II: Hermeneuticism

A different, but still non-standard, conception of mathematical truth arises within a position I shall refer to (following Burgess 1983) as "hermeneutic nominalism". According to hermeneutic nominalism, the literal or disquotational truth conditions of mathematical assertions are misleading. Mathematical assertions are not meant as literal assertions about mathematicalia. In short, mathematical assertions are "true", but only in some mystical non-literal sense, accessible only to soothe-saying philosophical sophisticates, but not to simple-minded mathematicians.

A proponent of such a view is Chihara 1973, 1990. Although Chihara argues that "mathematics is a body of truths", he refuses to accept that there are such things as real numbers, sets, functions and so on. Chihara seems to agree with the human race that,

(1) 'there is a real number whose cube is 7' is true,

but also refuses to assert also that,

(2) there are real numbers.

The problem is that (1), plus the disquotation axiom that,

(3) 'there is a real number whose cube is 7' is true if and only if there is a real number whose cube is 7.

just logically *implies* (3), which Chihara refuses to accept.

It is incoherent for a nominalist to reject such disquotational equivalences[130]. For the concept of truth—indeed, the correct use of the truth predicate—is constrained *a priori* to satisfy such T-sentences (this is why 'true' cannot mean 'provable'). If I am right, disquotational T-sentences are a good example of analytic sentences.

The resolution, I suggest, is partly verbal: Chihara simply doesn't (indeed, I urge, cannot) mean *truth* when he writes 'mathematics is a body of truths'. What he actually means by an assertion,

(4) φ is true

is (something like),

(5) φ has a "correct" but "non-literal reconstrual" φ^* which is true.

Thus, this kind of hermeneutic nominalism involves the basic idea of *reconstruing* mathematical statements. This position has several versions, as I discuss later in Chapters 7 and 8. In particular, aside from one's sheer incredulity that such non-literal reconstruals are what mathematicians "really mean" when they assert mathematical theorems, most suffer from the problem of accounting for the application of mathematics. The most promising reconstruals (Hellman 1989, Chihara 1990) use *modality*: but it is hard to believe that what ordinary assertions about numbers and sets "really mean" is to be given by a modal reconstrual involving merely possible structures (Hellman) or linguistic tokens (Chihara).

---

[130] I have hinted (no more) in Chapter 5 that the incoherence of denying any sentence of the form $Tr(\langle\varphi\rangle) \leftrightarrow \varphi$ resides in its "analyticity", and that its analyticity resides in its being a "partial definition" (not: part of an *implicit* definitions) of truth. And this, I hint, is connected to the *conservativeness* of DT, that is, all these T-sentences.

Thus, according to hermeneuticism, each mathematical statement φ is to be non-literally "reconstrued" as some other statement φ^*. For example, the nominalist might argue that we should "reconstrue" the mathematical sentence,

(6) there are numbers greater than 10

as the "nominalistically acceptable" sentence about concrete linguistic entities:

(6)* there is *a numeral token* longer than ten strokes

or its weakened *modal* version:

(6)** there *could have been a numeral token* longer than ten strokes

Again there are two alleged advantages of this proposal:

i. the reconstrual φ^* will *lack ontological commitment to abstracta.*

ii. the truth of the reconstrual φ^* can be easily *known* or *verified.*

I want to suggest that the hermeneuticist can put the suggested "reconstruals" to an important use in defining a non-classical conception of "mathematical truth". For he or she may suggest that we accept every *non-disquotational* axiom:

$$M\text{-}Tr(<\varphi>) \leftrightarrow \varphi^*,$$

where φ^* is the correct nominalistic "reconstrual" of φ. Indeed, he or she can adopt the each of these axioms as a *partial definition* of hermeneuticist "mathematical truth", M-Tr, just as the standard Tarskian approach takes each of the T-sentences, the disquotational instances, $Tr(<\varphi>) \leftrightarrow \varphi$, as partial definitions of literal truth, Tr.

So, the hermeneutic nominalist could follow the disquotationalist and introduce, by similar means, a theory of *hermeneutic mathematical truth*, call it HT. The axioms of HT are non-disquotational "partial definitions", such as:

(7) 'there are numbers greater than 10' is "mathematically true" iff *there could have been a numeral token* longer ten strokes

In fact, equipped with a systematic correlation of mathematical assertions and their hermeneutic reconstruals φ^*, given by a translation function Γ such that $\varphi^* = \Gamma(\varphi)$, the axioms of HT are $M\text{-}Tr(\langle\varphi\rangle) \leftrightarrow \Gamma(\varphi)$, for each mathematical assertion φ.

Indeed, the nominalist might think that 'true' just means 'mathematically true', as partially defined by HT. Strictly speaking, however, this procedure is not even consistent, if he or she also,

i. accepts the disquotation theory, wherein 'true' is partially defined by DT;

ii. accepts that certain existential mathematical assertions are true;

iii. claims that that numbers or sets do not exist.

Suppose that this nominalist proposes that Tr and $M\text{-}Tr$ *express the same concept* (at least in application to mathematical sentences). That is, we should accept $\forall x(Tr(x) \leftrightarrow M\text{-}Tr(x))$, where x ranges only over mathematical assertions. It is easily seen that HT and DT, plus this definition, jointly entail $\varphi \leftrightarrow \Gamma(\varphi)$, for each such mathematical assertion φ. The problem is simply that the nominalist *accepts* the reconstrual $\Gamma(\varphi)$, but *rejects* the mathematical assertion φ.

Perhaps, he or she can always reply that by 'true' (as predicated of a mathematical sentence), he or she always *means* 'mathematically true' as *partially defined* by HT axioms alone. The above argument shows clearly that this manoeuvre (plus the proposer's background theory) simply assigns a *deviant meaning* to 'true'. That is, a meaning extensionally different from concept expressed by 'true' as partially defined by DT.

Finally, let us quickly look at the possible relation of φ to its "reconstrual" φ^* within the interpreted language of the hermeneutic nominalist. In the scenario under discussion, there will be many mathematical sentences φ such that,

i. the hermeneuticist *accepts* φ^* (as literally true),

ii. the hermeneuticist *rejects* φ (as literally false).

It follows from this that ,

iii. the consistent hermeneuticist cannot *accept* the biconditional $\varphi \leftrightarrow \varphi^*$.

And it follows from this that,

iv. φ and φ^* cannot "mean the same thing" (from the nominalist's point of view).

For let φ be an existential mathematical assertion with explicit ontological commitment to abstracta. Let φ^* be its hermeneutic reconstrual. Assume now that φ and φ^* mean the same in the interpreted language of the hermeneutic nominalist. Then, minimally, in the combined language of these sentences, the biconditional $\varphi \leftrightarrow \varphi^*$ must be analytically true. Now suppose the nominalist accepts φ^*. He or she would then be compelled ("rationally") to accept φ. But this, on his or her nominalist principles, he or she rejects. Therefore, φ and φ^* cannot mean the same.

Indeed, the hermeneutic nominalist cannot even (rationally) believe that $\varphi \leftrightarrow \varphi^*$ is *true*! From this, it follows that the hermeneutic reconstrual does not preserve truth! Oddly enough, a similar argument shows that a platonist *can* (if he or she wishes) argue that $\varphi \leftrightarrow \varphi^*$ is *true* (but presumably will add that they obviously do not mean the same).

CHAPTER 7

Nominalism: Reconstrual & Reconstruction

Numquam ponenda est plurales sine necessitate.

William of Ockham, *Super Quattuor Libros Sententiarum.*[131]

It would be satisfying to contrive a systematic account of the world while staying strictly within an ontology of physical objects, and indeed physical objects big enough to be perceived. ... We have to conclude that multiplication of entities can make a substantive contribution to theory. ... Pad the universe with classes or other supplements if that will get you a simpler, smoother overall theory. Otherwise, don't. Simplicity is the thing, and ontological economy is one aspect of it, to be averaged in with the others.

W.V. Quine 1966-74 (1976), 'On Multiplying Entities', p. 259.

To say what someone is talking about is to say no more than how we propose to translate his terms into ours. We are free to vary the decision with a proxy function. The translation adopted arrests the free-floating reference of the alien terms only relatively to the free-floating reference of our own terms, by linking the terms.

W.V. Quine 1981b (1981a), 'Things and Their Place in Theories', p. 20.

The Hungarian phrase 'Can you direct me to the railway station?' is here translated as 'Can you please fondle my buttocks?'.

Monty Python, 1971.

7.1 Nominalism: Introduction

According to the argument of Chapters 1 to 3, an ontology of (presumably abstract) mathematicalia is built into much of modern science. The examples given covered quantitative laws, space-time theory, syntax and metalogic. Trivially, if one wishes to *reject* such an ontology, one cannot *also* adopt Fine's allegedly neutral "core position", a realist *acceptance as true* of the "certified theories of science" and a *realist view of truth* (e.g., Tarski's) wherein a theory is true just in case the entities, processes, events and

[131] Quoted from Kneale & Kneale 1962 (1988), p. 243.

properties it quantifies over exist and are related as the theory says they are. In short, the renunciation of mathematicalia is simply inconsistent with science.

Nominalism is the renunciation or repudiation of abstract entities, including mathematicalia. The puzzle is obvious. Nominalism contradicts science. How can nominalism and modern science be reconciled if, as Quine and Putnam argue, they are incompatible? The remaining task of this thesis is to discuss and assess some recent attempts by nominalistically-inclined philosophers of mathematics to square this circle.

Nominalism emerged historically in the thought of certain mediaeval thinkers, notably Garland and Roscelin[132], who opposed the reality of Platonic universals. Instead, they urged that general terms like 'angel' and 'man' are not *names* or *designations* of universals. Rather, such term are nothing more than predicates, applying individually to each and every concrete instance, be it a particular angel or man. In the notorious words of Roscelin (as reported by his pupil, Abelard), universals are *flatus vocis*, the "breath of voice". The most celebrated of the later, fourteenth century, nominalists (the so-called *nominales*) was William of Ockham:

> For the truth of the proposition 'This is an angel' ... it is sufficient and necessary that the subject and the predicate suppone for [refer to] the same ... And so it is not meant that this has angelity, or that angelity is in this, or anything of the kind, but simply that this is truly an angel, not indeed that it is that predicate but that it is that for which the predicate suppones.
>
> William of Ockham, *Summa Totius Logicae*, ii.2[133]

The proposal is that talk of abstract universals be eliminated in favour of talk of *linguistic entities*: predicates. The truth of '*a* is an angel' requires only that the concrete referent of '*a*' be a thing of which the *predicate* 'angel' is true. It does not require that there be an abstract universal (*angelity*) which inheres in the referent of '*a*'.

What is the motivation for such a repudiation of abstract universals? By common consent, the motivation resides in Ockham's famous "Razor", sometimes called "The

[132] See Kneale & Kneale 1962 (1988) for a discussion, in the context of logical theory, of this tradition of nominalism.

[133] From Kneale & Kneale 1962 (1988), p. 270.

Principle of Parsimony", expressed (probably apocryphally) by the slogan "*entia non multiplicanda sine necessitate*".

Scholarly debates rage as to the precise meaning and intent of Ockham's anti-Platonist slogans. Some recent commentators have de-emphasized the *ontological* component (the parsimonious renunciation of abstract entities) and emphasized instead the idea of *conceptual simplicity* (the parsimonious renunciation of redundant *concepts* and *assumptions*). An early advocate, though hardly a scholar, of this interpretation of Ockham was the *Tractarian* Wittgenstein,

> Occam's razor is, of course, not an arbitrary rule nor one justified by its success. It simply says that unnecessary elements in a symbolism mean nothing. Signs which serve *one* purpose are logically equivalent, signs which serve *no* purpose are logically meaningless.
>
> (Wittgenstein 1922 (1981), 5.47321, p. 129).

A notable advocate of parsimony, both ontological and conceptual, as an organizing (pragmatic) principle of reason within our evolving "conceptual scheme" is Quine:

> It is not to be wondered that theory makers seek simplicity. When two theories are equally defensible on other counts, certainly the simpler of the two is to be preferred on the score of both beauty and convenience. ... [the maxim of simplicity of nature] seems to be implicitly assumed in every extrapolation and interpolation, every drawing of a smooth curve through plotted points. And the maxim of uniformity of nature is of a piece with it, uniformity being a species of simplicity.
>
> (Quine 1963b (1976), p. 255).

Quine points out that there may even be a tension between these two edges of Ockham's Razor: conceptual simplicity and ontological parsimony may *conflict*:

> We move into a conceptual scheme of electrons, neutrons, and other hypothetical particles that can never be directly observed; a conceptual scheme of kinky four-dimensional space-time, and of mathematical abstractions—sets, relations, functions, integers, ratios, irrational numbers, imaginary numbers, infinite numbers. None of these extras are observable. ... What then is all this extra apparatus of ours? Is it sheer mythmaking, unwarranted by observational evidence?
>
> Paradoxically, the purpose of all this extra apparatus is *simplification*. We are out to *systematize* and integrate the testimony of our senses by devising laws that relate the observable phenomena systematically to other phenomena; and the *most systematic network* of relations for this purpose turns out to be a network that links all these phenomena up with a lot of additional, hypothetical entities that are *only assumed for integrating the system.*
>
> (Quine 1973 (1976), p. 62. Emphasis added)

Nominalism re-emerged in the early twentieth century, advocated by able logicians, like Stanislaw Lesniewski at Warsaw, and Nelson Goodman and W.V. Quine at Harvard. Let us refer to this particular form of nominalism as *Logical Nominalism*, a scientific research programme aimed at showing how to redevelop or reconstruct mathematics or science so as to avoid reference to abstract entities. The eliminative programme had its central goal of ridding science, and even mathematics itself, of reference to abstractions, like numbers and sets. Lesniewski's *mereology* and Goodman's *calculus of individuals* were proposed as nominalistic *replacements* for standard set theory.

If such a constructive programme of eliminative nominalism were workable, the *indispensability component* of the Quine-Putnam argument would fail. Although science and mathematics would still, when literally construed, count as a platonic system, the nominalistic replacement would provide the correct fluff-free description of the facts. Any fluffy Platonic talk of abstract mathematicalia could (in principle) be dropped, or regarded as a mere *façon de parler*, a convenient fiction.

7.2 Non-Nominalizability: Preliminaries

Let us say that an assertion which carries *prima facie* commitment (usually via quantification or by naïve semantics) to abstracta is *platonistic*. Let us say when such a statement may be "re-expressed" without such platonistic commitment, that it is *nominalizable*.

An example of a nominalizable platonistic statement is the assertion,

(1) the number of eyes = *2*

whose *prima facie* logical structure requires an entity, presumably abstract, for the singular terms 'the number of *F*s' and '*2*' to be names. The "nominalization" of (1) is the statement,

(2) $\exists x \exists y (E(x) \wedge E(y) \wedge x \neq y \wedge \forall z (E(z) \rightarrow z = x \vee z = y))$

whose explicit logical structure require the existence solely of two concrete eyes.

So, the natural question for the programme of eliminative nominalism is this:

A. Are some (perhaps scientifically important) platonistic statements *non-nominalizable*?

This question is complicated by a related problem. The nominalist might introduce novel or non-standard "logical apparatus" which, it will be urged, carries no platonic commitments. So, the related question is this:

B. Suppose a nominalization strategy is proposed, involving some novel "logical apparatus". Is this apparatus genuinely free of abstract commitments and thus nominalistically acceptable?

For Lesniewski, the new logical apparatus included *mereology* and *substitutional quantification*. Goodman and Quine likewise adopted mereology as part of their strategy for "constructive nominalism" (although the inadequacy of mereology as a replacement for set-theoretical mathematics was soon recognized by Quine). More recent nominalistic or quasi-nominalistic strategies have considered introducing *higher-order quantification* (e.g., Field, Boolos) and *modal operators* (Putnam, Chihara, Field, Hellman).

We shall pursue both questions in a piecemeal fashion.

Examples, now classic, of non-nominalizable statements were found and given detailed attention by the post-nominalist Quine, from 1947 onwards. Statements of *fixed definite cardinality*, like (1) above, seem nominalizable. But what about cardinality comparisons? For example,

(3) there are more Fs than Gs

This assertion can be expressed using mathematical notions thus:

(4) there is an injection from $\{x: G(x)\}$ to $\{x: F(x)\}$ and there is no injection from $\{x: F(x)\}$ to $\{x: G(x)\}$

As such, it is committed by the obvious quantification to various mathematicalia (to sets and injections). Is it possible to maintain that (3) is "really" a nominalistic statement? To begin, note that (3) implies each of the nominalistically expressible statements,

(3′) if there is at most 1 *F*, then there are no *G*s;

(3″) if there are at most 2 *F*s, then there is at most 1 *G*;

(3‴) if there are at most 3 *F*s, then there are at most 2 *G*s;

… … …

Since each of this infinite set can be nominalized using a numerical quantifier, one might wonder then whether we could re-express (3) using these quantifiers (one quantifier for each numeral n: $\exists_n x$) and a *substitutional quantifier*, thus,

(5) $\Pi n(\exists_{\leq n+1} xF(x) \rightarrow \exists_{\leq n} xF(x))$

The problem with this proposal is the acceptability of the logical apparatus. Is substitutional quantification with respect to an infinity of numeral types nominalistically acceptable?

Issues deepen. Even with this apparatus, we cannot re-express an assertion like,

(6) There are more space-time points than points in an ω-region

for a quite simple reason. Standard space-time physics implies that there are $\mathfrak{c}$-many space-time points. According to *physics*, the totality of space-time points forms an ur-element set which can be structured as a 4-dimensional topological manifold. But there is no nominalistically definable numerically definite quantifier $\exists_{\mathfrak{c}} x$ corresponding to 'there are $\mathfrak{c}$-many *F*s'. (One cannot re-express $\exists_{\mathfrak{c}} x\varphi$ using a finite formula, using just identity and no mathematical notions).

The assertion about injections (4) can be re-expressed in *second-order logic* thus,

(7) $\exists f(\mathrm{Inj}(f) \wedge \forall x(G(x) \rightarrow F(f(x))) \wedge \neg\exists f(\mathrm{Inj}(f) \wedge \forall x(F(x) \rightarrow G(f(x)))$

where $\mathrm{Inj}(f)$ is the formula $\forall x \forall y(x \neq y \rightarrow f(x) \neq f(y))$. But, again, it is doubtful that this use of second-order logic is nominalistically acceptable.

Chapter 7. Nominalism: Reconstrual And Reconstruction

Another example discussed by Quine is the non-nominalistically-definable relation expressed by 'ancestor',

(8) *x* is an ancestor of *y*

For example, Eve is an ancestor of Cherie. That is, Eve is a parent of a parent of ... a parent of Cherie. We may *recursively* define 'ancestor' thus,

(a) any parent of *x* is an ancestor of *x*

(b) any parent of an ancestor of *x* is an ancestor of *x*

(c) if *y* is an ancestor of *x*, then either *y* is a parent of *x* or *y* is a parent of an ancestor of *x*

This may then be converted to an *explicit definition* ("directified") using set theory. This explicit definition of 'ancestor', due essentially to Frege 1879, can be formulated as,

(9) *y* is a member of every class that contains *x* and contains any parent of *z* if it contains *z*

(Frege's use of this is to define "natural number": a natural number is something such that *0* is an ancestor under the successor relation). As Quine emphasizes, this assertion quantifies platonistically over classes. If one wants to reduce 'Eve is an ancestor of Cherie' to a claim about parents, one must use set theory (or second-order logic):

(10) Eve is a member of every class that contains Cherie and contains any parent of *z* if it contains *z*

This analysis of 'ancestor' has the consequence that an apparently nominalistic and true assertion like,

(11) Napoleon is not an ancestor of Boolos

implies the existence of a *class*[134]. Is this odd? Part of the use of mathematics is to provide a "finite encapsulation" of the *infinite*, as we sometimes indicate in our informal language by ellipsis, '...', or by expressions like 'and so on'. Thus, (11) means,

---

[134] E.g., see the first paragraph of Boolos 1985:

(12) (Napoleon is not a parent of Boolos) and (Napoleon is not a parent of a parent of Boolos) and ... (so on!)

It is quite unsurprising that the encapsulation of this infinitary statement, involving the enigmatic '...' should introduce mathematical entities like classes.

More recent examples of non-nominalizability have come to light. For example, in *Methods of Logic*, Quine mentions the so-called "Geach-Kaplan sentence",

(13) Some critics admire nobody but one another

This can be expressed using set theory straightforwardly thus,

(14) There is a *set* Z of critics such that, for any x, if $x \in Z$, then x admires y only if $y \neq x$ and $y \in Z$.

Or, more formally, (taking a domain of critics only), as,

(15) $\exists Z(\exists x(x \in Z) \wedge \forall x \forall y(x \in Z \wedge A(x, y) \rightarrow (y \neq z \wedge y \in Z)))$

Now this can be rewritten using second-order logic, as,

(16) $\exists Z(\exists x Z(x) \wedge \forall x \forall y(Z(x) \wedge A(x, y) \rightarrow (y \neq x \wedge Z(y)))$

The important thing about (13) is its *non-nominalizability*. Kaplan showed how to prove that the second-order statement (16) is not equivalent to any first-order formula containing just $A(x, y)$[135].

Further examples of difficulties in nominalizing platonistic statements have been provided by Putnam 1971 and Field 1980. Following Quine's early lead, both authors

Frege's definition of "x is an ancestor of y" is: x is in every class that contains y's parents and also contains the parents of any member. A philosopher whom I shall call N. once asked me. "Do you mean to say that because I believe that Napoleon was not one of my ancestors, I am committed to such philosophically dubious entities as classes?" Although it is certain that Frege's definition, whose logical utility, fruitfulness and interest have been established beyond doubt, cannot be dismissed for such an utterly crazy reason, it is not at all easy to see what a good answer to N.'s question might be.

[135] Kaplan's proof that (16) is not equivalent to a formula containing just A is based on demonstrating that the second-order formula (16) discriminates between certain elementarily-equivalent interpretations $\mathfrak{I}_1$ and $\mathfrak{I}_2$. This is impossible for any first-order formula (or set of such) containing just A. See Boolos 1984, pp. 432-433.

discuss those mathematicized statements used in formulating *scientific laws*. In particular, statements using *mixed predicates* and quantification over mathematicalia.

Consider a platonistically definable relation amongst concreta:

(17) (x is r-mass-related to y) $\leftrightarrow$ ($m_{kg}(x) = r \times m_{kg}(y)$)

Here's a "law" using this notion relating concreta,

(18) if x is π-mass-related to y, then $f(x)$ is λ-mass-related to $g(y)$

which is expressible platonistically as,

(19) if $m_{kg}(x) = \pi \times m_{kg}(y)$ then $m_{kg}(f(x)) = \lambda \times m_{kg}(g(y))$

Now, Field explained in his 1980 how *some* platonistically definable relations amongst concreta may be redefined nominalistically using certain "nominalistic" primitives:

'x is massless',

'x is less-massive than y',

'the mass-difference of x and y is congruent to the mass difference of z and w'.

For example, given these, we can re-express the relation,

(20) x is twice as massive as y

by,

(21) (y is less-massive than x) $\wedge$ $\exists w$(w is massless $\wedge$ cong-mass(wy, yx))

However, it is not so easy to re-express,

(22) x is 3.14 times as massive as y

But it is possible. We can in fact re-express (22) as follows,

(23) there is something z such that z is 314 times as massive as y and z is 100 times as massive as x

That is, $m_{kg}(z) = 314m_{kg}(y)$ and $m_{kg}(z) = 100m_{kg}(x)$. One then has to define 'z is 314 times as massive as y' and 'z is 100 times as massive as x' using the afore-mentioned nominalistic primitives. Notice that the proposed nominalistic definition of this rather complicated mass relation between concreta requires the *postulation of sufficiently many other objects*, in the appropriate primitive mass relations. If these other objects did not exist, then x simply could not be 3.14 times as massive as y.

Even with such notions alone, we can re-express,

(24) x is π times as massive as y

Such a statement requires quite advanced nominalization techniques. For example, we might introduce a notion,

the line segment r measures the mass of x,

and then, using geometrical techniques (in particular, invoking the fact that such line segments can be thought of a real number surrogates[136]), re-express (23) roughly as,

(25) there is a line segment r_1 which measures the mass of x and there is a line segment r_2 which measures the mass of y, and the ratio r_1: r_2 is equal to π.

where 'π' is now defined nominalistically as 'the ratio $L : D$, where L is a circumference and D a diameter for any circle C', and where a circle is a locus of points all equidistant from a fixed point. (More exactly, 'the ratio $x : y$' is an "incomplete symbol" which is governed by a Eudoxian principle: '$x : y = z : w \leftrightarrow x \times w = z \times y$', where it is assumed that multiplication ($\times$) is somehow made sense of for concrete line segments x, y, z and w).

The above examples are attempts to obtain what I below identify as *expressive conservativeness*. Suppose that we attempt the following reconstruction. We begin with our standard two-sorted mathematicized theory T. The aim is to eliminate the reference to secondary entities (mathematicalia). We expand T in L to a new theory $\mathsf{T}^\S$ in an expanded language $\mathsf{L}^\S$ by adding new nominalistic primitives (e.g., 'x is massless', 'x is less massive than y') and new axioms governing these primitives.

136 There are several "mechanical" ways of converting masses to distances. Old-fashioned analogue weighing scales convert mass to a distance along a marked scale. Likewise a spring balance.

First, we require *deductive conservativeness.* The reconstruction $T^{\S}$ in $L^{\S}$ of T in L must imply all the original non-mathematical (primary) consequences of T. That is, if T $\vdash \varphi$, where φ is a mathematics-free assertion in L, then we require that $T^{\S} \vdash \varphi$.

Second, $T^{\S}$ must be able to *re-express* using its primary predicates (in $L^{\S}$) all the platonistically definable relations, expressible in the language L of T, amongst primary entities. That is, for any mixed mathematicized formula $F(x_1, \ldots, x_n)$ in L containing only primary variables, we require that there be a mathematics-free formula $F^{\S}(x_1, \ldots, x_n)$ in $L^{\S}$ such that $T^{\S} \vdash \forall x_1 \ldots \forall x_n(F(x_1, \ldots, x_n) \leftrightarrow F^{\S}(x_1, \ldots, x_n))$. For example, we might be able to prove (in $T^{\S}$) that $m_{kg}(x) = 0$ if and only if x is massless (so we can eliminate '$m_{kg}(x) = 0$'.

If we now take the primary restriction of $T^{\S}$ in $L^{\S}$, we obtain a nominalistic theory $(T^{\S})^{\circ}$ in $(L^{\S})^{\circ}$ such that, if T $\vdash \varphi$, then $(T^{\S})^{\circ} \vdash \varphi$, for any mathematics-free assertion φ; and, furthermore, if T $\vdash \varphi(F, G, \ldots)$, where $F, G, \ldots$ are mixed predicates in L, then $(T^{\S})^{\circ} \vdash \varphi(F^{\S}, G^{\S}, \ldots)$.

For example, our nominalistic reconstruction of "mass theory" must imply,

(26) for any objects x, y, ((x is π times as massive as y) $\leftrightarrow F^{\S}(x, y)$)

This will guarantee that, if T $\vdash \exists x \exists y(m(x) = \pi \times m(y))$, then $(T^{\S})^{\circ} \vdash \exists x \exists y(F^{\S}(x, y))$.

I think it is clear that the demand of expressive conservativeness is acceptable. If it is dropped, then the mathematicized theory expresses *relations* amongst concrete (primary) entities which the nominalistic reconstruction loses. For example, it is natural to demand that the relations *amongst concreta* defined by Newton's laws for mass, acceleration and force should be somehow encoded within the nominalistic reconstruction. This (abstraction-laden) information is central the explanation of the dynamics of moving bodies. If the nominalistic reconstruction is just a *list* of the mathematics-free predictions of the platonic theory T, these mathematics-free predictions will look completely *ad hoc*. So, expressive conservativeness is important.

Moreover, a nominalistic redefinition will be required for *every* platonistically definable relation amongst primary entities! This is no easy task. The extent to which mathematics can be thus eliminated from our scientific statements about the world is the topic of the remaining Chapters.

7.3 Three Construals of Nominalism

Typically, an advocate of nominalism will begin by expressing his or her outrage at the introduction of abstractions in science. Platonistic mathematicized theories, whether within pure unapplied mathematics or whether fully integrated within science, are an affront to our *a priori philosophical intuitions*; they are an affront to our *epistemology* (abstractions are "unknowable"). At any cost, such platonistic theories must be dropped.

In an important position paper (Burgess 1983), John Burgess explained why he "is not a nominalist". The method was that of exhaustion. Burgess argues that nominalism about mathematicalia may be construed in three quite distinct ways:

i. Instrumentalist Nominalism

Mathematicized theories that require such mathematicalia are "convenient fictions". The world simply behaves "as if" there are such things, even though there are no such things.

ii. Hermeneutic Nominalism

Mathematicized theories which appear to imply the existence of such mathematicalia do not "really" imply any such thing. When properly "interpreted", such theories are committed only to mundane concreta.

iii. Revolutionary Nominalism

Science is to be revolutionized, dropping all mathematicization. Presently accepted mathematicized theories of Nature, like relativity and quantum theory, are to be replaced by mathematics-free reconstructions.

Burgess explains why none of these positions is satisfactory or even appealing. He thereby concludes that nominalism, in any of its forms, is unacceptable.

We shall briefly discuss the *instrumentalist* variety of nominalism, which Burgess compares to the fictionalism of Hans Vaihinger, the author of *The Philosophy of "As If"* (1913), and to the sceptical instrumentalist philosophy of van Fraassen 1980. Burgess refers approvingly to Putnam's arguments against instrumentalism and scepticism (especially Putnam 1971). I will not go into details, for the obvious reason that extreme scepticism is a trivial way of avoiding *any* argument. The extreme sceptic just thinks that *nothing* is rationally compelling[137]. But this is sheer nihilistic *a priorism.* With the greatest of respect to van Fraassen 1980, Laudan 1981, and Cartwright 1983, I can see no reason to accept their irrational *a prioristic* conclusions about rationality. The plain fact is that real rational flesh-and-blood working scientists do find Relativity, Quantum Theory, Transformational Grammar, Evolutionary Theory and so on, rationally compelling and, arid sceptical puzzles aside, they quite correctly ignore the rationally *uncompelling* (and often technically deficient) arguments of some philosophers. If Schrödinger's Law is not true ("it lies"), as Cartwright has amazingly "discovered", isn't it just a little *odd* that no-one has ever found a counter-example, despite countless experiments, predictions and explanations? Of course, it *could* be false (which no one denies), but such extremist (and quite pretentious) "arguments" simply do not touch the matter of whether it *is* false. If these arguments were rationally compelling, then the whole physics community (unless they are irrational) would say: "Gosh! Van Fraassen and Cartwright have shown how the whole of modern high-level quantum physics and space-time theory is an *elaborate myth*, spreading terrible *lies* about such *fictions* as the topological non-triviality of the vacuum. Let's close down CERN and Fermilab!".

The issue of extremist scepticism about science is somewhat tangential to the general theme of this dissertation. I assume that the reader is the normal person who thinks that standard accepted scientific theories like GR, QED and so on, are pretty good

[137] Not even simple mathematical truths, logical truths, or the observationally "given". As Descartes pointed out, a "malicious demon" may be supposed who forces us to think in a manner that appears valid, but in fact sometimes isn't (i.e., leads from truths to falsehoods).

descriptions of the Universe. I shall quickly come to a sceptical conclusion about extreme scepticism.

A succinct summary of the unsatisfactory nature of scepticism, and, in fact, of the untenability of "uncompromising empiricism", was given long ago by Russell, in a classic discussion of Logical Positivism:

> There is one matter of great philosophic importance in which a careful analysis of scientific inference and syntax leads ... to a conclusion which is unwelcome to me and (I believe) to almost all logical empiricists. The conclusion is that *uncompromising empiricism is untenable.* From a finite number of observations no general proposition can be inferred to be even probable unless we postulate some general principle of inference which cannot be established empirically. ... As to what is to be done in consequence there is no agreement. Some hold that truth does not consist in conformity with fact, but only in coherence with other propositions already accepted for some undefined reason. Others, like Reichenbach, favour a posit which is a mere act of will and is admitted not to be intellectually justified. Yet others make attempts—to my mind futile—to dispense with general propositions. For my part, I assume that science is broadly speaking true, and arrive at the necessary postulates by analysis. But *against the thoroughgoing sceptic I can advance no argument except that I do not believe him to be sincere.*
>
> (Russell 1950 (1956), pp. 381-382. Emphasis added).

In the last sentence, Russell presents what might be called the *Shamming Objection* to scepticism. Namely, those who affect to be sceptics, really aren't: *they are "shamming"*. On this view, irrespective of van Fraassen's detached lip service to scepticism, he is (in some sense) shamming when he affects to express disbelief or agnosticism about atoms, molecules, electrons, genes and so on. Field 1980 calls the analogue of this (in the philosophy of mathematics) "intellectual dishonesty", namely "refusing to assert in one's philosophical moments what one assumes all the time in doing science".

7.4 Reconstrual: Preliminaries

In ordinary language, a range of statements carry *prima facie* ontological commitment to abstracta. Examples[138] are:

(1) Humility is a virtue

[138] I take the first two of these examples from Quine 1960, pp. 122-123.

(2) Redness is a sign of ripeness

(3) The number of apples on the desk is 2

Anyone with nominalistic scruples might well balk at the admission of such abstracta (humility, redness and the number of apples on the desk) into their ontology, and might wish to eliminate such abstract references. One procedure would be to engage in an eliminative programme of showing such statements to be merely perverse ways of saying humdrum things about mundane concrete bodies.

Such an eliminative programme was discussed long ago in Goodman & Quine 1947. In 1947, Quine wrote:

> ... Another and more serious case in which a man frees himself from ontological commitments of his discourse is this: he shows how some particular use which he makes of quantification, involving a prima facie commitment to certain objects, *can be expanded into an idiom innocent of such commitment ...* In this event the seemingly presupposed objects may justly be said to have been explained away as convenient fictions, manners of speaking.
>
> (Quine 1947 (1980), pp. 103-104. Emphasis added).

Quine was no doubt thinking of Russell's 1905 paradigmatic theory of descriptions, wherein we expand out any context,

$G(\text{the } F)$

as,

$\exists x[G(x) \wedge F(x) \wedge \forall y(F(y) \rightarrow y = x)]$.

Quine referred to this kind of expansion, following Russell, as a *contextual definition*[139]. That is, a definition which permit the elimination of 'the' from any context. Indeed, one can, and probably should, think of Russell's explanation as a logical analysis of the *meaning* of 'the' (or as an analysis of truth conditions of sentences containing 'the'), as it is usually used, prefixed to a polyadic predicate to form a definite description.

The suggestion that a person may "free himself from ontological commitments of his discourse" by showing how a statement exhibiting a *prima facie* commitment to certain

[139] Quine traces such contextual definition back to Bentham's "paraphrasis".

objects “can be expanded into an idiom innocent of such commitment” and thus that “the seemingly presupposed objects may justly be said to have been explained away as convenient fictions, manners of speaking” lies behind the recent flurry of nominalistic reconstruals, which are to be discussed below. We expand platonistic mathematical statements “into an idiom innocent of such commitment”.

Actually, it would be more accurate to say that, for each of these reconstruals, a mathematical assertion is *non-literally* reconstrued, rather than simply “expanded out”. For it would seem that Russell, Quine and others would have argued that,

(4) $\exists x[G(x) \wedge F(x) \wedge \forall y(F(y) \rightarrow y = x)]$

is actually the *literally correct* logical form of,

(5) $G(\text{the } F)$.

Russell, reflecting later on this matter, wrote:

> Meinong maintains that there is such a thing as the round square only it does not exist, and it does not even subsist, but nevertheless there is such an object, and when you say 'The round square is a fiction', he takes it that there is an object 'the round square' and there is a predicate 'fiction'. No-one with a sense of reality would so analyse that proposition. He would see that the proposition wants analysing in such a way that you won't have to regard the round square as a constituent of the proposition.
>
> (Russell 1918-1919 (1956), p. 223).

Another paradigm of “expanding out” statements with commitment to abstracta is clearly illustrated by the use of numerically definite quantifiers. For example,

(3) The number of apples on the desk is *2*

may be “expanded out” as,

(6) $\exists x \exists y(x \neq y \wedge A(x) \wedge A(y) \wedge \forall z(A(z) \rightarrow (z = x \vee z = y)))$

The mathematicized statement (3) contains an apparent reference to an abstractum (the number *2*). The reconstrual (6) using numerically definite quantifiers is only committed to the existence of a pair of apples on the desk (I discuss such quantifiers below).

However, it might be urged that a *literal* regimentation of (3) is,

(7) $\#A = 2$

where '#' is an operator on predicates yielding a singular term which designates numbers and '*2*' is a bona fide singular term designating a number. If this is correct, then we must explain the *relationship* between the ontologically committed (7) and its mathematics-free reconstrual (6).

Similarly, there is a literal regimentation of,

(8) The cube root of seven is less than 2,

namely,

(9) $\exists! x(x^3 = 7 \wedge x < 2)$.

To regiment 'The cube root of 7 is less than 2' as something *logically inequivalent* to (9) is to adopt a *non-literal* reconstrual.

Returning to our earlier example, consider,

(1) Humility is a virtue

Someone philosophically puzzled by the apparent reference to an abstractum, namely humility, might wish to reconstrue this statement nominalistically, namely as,

(10) All humble persons are virtuous

which only quantifies over persons. A similar proposal might be made for (2), with its reference to another abstractum, redness. We simply reconstrue (2) as,

(11) All ripe fruit are red

In 1960, Quine put such eliminativism as follows:

> One might, with laudably scientific motives, resolve to sweep these abstract objects aside. One might begin by explaining 'Humility is a virtue' and 'Redness is a sign of ripeness' away as perverse ways of saying of humble concrete persons and red concrete fruit that they are virtuous and ripe.
>
> (Quine 1960, p. 122).

However, Quine[140] was sceptical that this kind of programme might *fully* succeed in its aims:

> But such a program cannot without difficulty be carried very far. What of 'Humility is rare'? We may for the sake of argument construe 'Humility is a virtue' and 'Humility is rare' as 'Humble persons are virtuous' and 'Humble persons are rare'; but the similarity is misleading. For whereas 'Humble persons are virtuous' means in turn that each humble person is virtuous, 'Humble persons are rare' does not mean that each humble person is rare; it means something rather about the *class* of humble persons, viz., how small a part it is of the class of persons. But these classes are abstract objects in turn—not to be distinguished from attributes save on a technical point. So 'Humble persons are rare' has only the appearance of concreteness; 'Humility is rare' is a more forthright rendering. Maybe this abstract reference can still be eliminated, but only in some pretty devious way.
>
> (Quine 1960, pp. 122-123).

If we bracket questions of whether such abstractum-eliminating "reconstruals" might be found for such recalcitrant cases as 'Humility is rare', a more serious question arises. This was first raised by Alston 1958, a paper sharply critical of Quine's method of nominalistic paraphrase. Roughly, Alston's question is: what exactly legitimizes the claim that any such nominalistic reconstrual φ° is an *adequate reconstrual* of some such (allegedly problematic) abstractum-committed assertion φ?

In short, noone has clarified what it means to say that,

(12) φ° is an *adequate reconstrual* of φ.

For example,

(13) 'Humble persons are virtuous' is an adequate reconstrual of 'Humility is a virtue'

Now, Quine himself was deeply opposed to (even the scientific intelligibility of) saying that,

(14) φ° means the same as φ.

---

[140] Remember, it was Quine who, with Goodman, initiated modern logical nominalism in Anglo-American philosophy. Quine was amongst the keenest to explore the possibilities of and limits of such "abstractum-eliminating paraphrases".

However, if we provide *no* constraints on reconstrual, then we could suppose that *anything* might count as a reconstrual of φ. (Then, when I utter some sentence φ in the presence of my interlocutor, he or she will be completely in the dark as to what *I* mean, even if he or she knows the conventional meaning of φ).

Let me clarify. The hermeneuticist has an "intuition", that

(3) $\#A = 2$

may be reconstrued as,

(6) $\exists x \exists y(x \neq y \wedge A(x) \wedge A(y) \wedge \forall z(A(z) \rightarrow (z = x \vee z = y)))$

I am not disputing the "intuition". But intuitions are slippery beasts, and if nothing counts as a *constraint* on the introduction of reconstruals, we have the beginning of a slippery slope towards total anarchy. If some hermeneutic anarchist comes along and says that statements like,

(15) Churchill was a great leader

(16) I dreamt last night that my mother tripped over and broke her hip

are to be construed as,

(15)° I implicitly endorse the mythologizing authoritarian apotheosis of leadership and coercion

(16)° I have a repressed desire to re-enter my mother's womb, but also know that to escape from my present obligations would lead to pain

then what are we to say? Is there an "intuition" behind such reconstruals?

Returning to the serious case, if there is an "intuition" behind the reconstrual of (3) as (6), and even perhaps behind the reconstruals of (15) and (16) as (15)° and (16)°, then it needs to be "brought to the surface", clarified and criticized.

Progress on this problem was made somewhat accidentally in the late 1960s in some important work of Putnam. For our purposes, the problem was immensely clarified when Field 1980 introduced a theory of "the abstract counterpart relation". This theory is discussed below. It may be said to have the nominalistic advantage of not introducing such things as *meanings* for φ and its reconstrual $\varphi°$ to share.

7.5 Abstract Counterparts

I now want to discuss the important theory of the *relation* between certain kinds of mathematical statements and their "intuitive" non-mathematical reconstruals. Once pointed out, this theory has the sort of obviousness and elegance that important truths often have. This theory (so far as I know) was first introduced by Putnam in 1967:

> Now consider an inference from (applied) arithmetic rather than from logic, say:
>
> There are two apples on the desk
> There are two apples on the table
> The apples on the desk and table are all the ones in this room
> No apple is both on the desk and on the table
> Two plus two equals four
> *therefore* There are four apples in this room
>
> The logicist account of such an inference is well known. The logicist definitions of 'there are two As' and 'there are four As' are such that one can prove that 'There are two As' is equivalent to a statement of pure quantification theory (with identity) namely: 'There is an x and there is a y such that x is an A and y is an A and $x \neq y$ and such that, for every z, if z is an A then either $z = x$ or $z = y$'. ... The entire inference above is equivalent, line by line (except for two plus two equal four) to an inference in pure logic, by the narrowest standard—quantification theory with identity. What of the line 'two plus two equals four'? The answer is that the above inference is still valid with that line omitted!
>
> (Putnam 1967a (1979), p. 27).

To this, he adds later:

> ... let us abbreviate the statement 'the set of planets belongs to the number nine' as *P*, and the statement 'there is an x and there is a y and ... such that x is a planet and t is a planet and ... and $x \neq y$ and ... and such that for every z, if z is a planet then $z = x$ or $z = y$ or ...', which expresses 'the number of planets is nine' is a purely first order way, as P^*. The equivalence $P \leftrightarrow P^*$ is a theorem of *Principia*, and hence holds in all models.
>
> (Putnam 1967a (1979), p. 31).

Building on Putnam's work, Field 1980 put these ideas to great use. First, in developing a "theory of abstract counterparts"; and second in trying to explain the *utility* of mathematics within science (his explanation is exactly like the explanation Putnam gives of the above inference).

Putnam's insight was to notice that the connection between statements like,

(1) The number of apples on the table = 2

(which carries a *prima facie* commitment to numbers), and nominalistic statements like,

(2) There are two apples on the table

which is committed just to apples, is really very simple. They are *provably equivalent modulo a background theory.*

Before we discuss the mathematical case more fully, let us return to Quine's correlated pair of statements:

(3) Humility is a virtue

(4) All humble persons are virtuous

To illustrate the Putnam-Field idea, consider the platonistic theory P with the axioms:

P_1 $\forall x(\text{humble}(x) \leftrightarrow \text{has}(x, \text{humility}))$

P_2 $\forall x(\text{virtuous}(x) \leftrightarrow \exists y(\text{virtue}(y) \wedge \text{has}(x, y)))$

Clearly, P is a platonistic theory which explains being humble as having the attribute of humility, and explains being virtuous as having some attribute which is a virtue. It is now quite simple to prove the conditional,

$P \vdash (3) \rightarrow (4)$

That is, the platonistic theory P licenses the inference from (3) to (4). Presumably, a platonist about ethics will believe the two axioms P_1 and P_2 (and much else, of course). Not surprisingly then, he or she will be able to deduce (4) from (3). In fact, with extra platonic theory, he or she might be able to show that (3) and (4) are equivalent, modulo the platonistic background theory.

This is the Putnam-Field idea. A humdrum statements about concreta and their "abstract counterparts" are *provably equivalent, modulo a platonistic background theory.*

Let B be some background theory. Then it may be the case, for two assertions, say φ and φ° (neither of which, in the interesting case, are theorems of B), that their equivalence is provable in B. That is,

$$B \vdash \varphi \leftrightarrow \varphi^\circ$$

If φ and φ° are theorems of B, then $\varphi \leftrightarrow \varphi^\circ$ is trivially a theorem of B. If φ is a theorem and φ° isn't, or vice versa, then of course $\varphi \leftrightarrow \varphi^\circ$ cannot be a theorem of B.

In the case of simple mathematicized statements, the proposal works like this. Consider the pair of statements:

(5) $\{x: \text{rabbit}(x)\} \subseteq \{x: \text{mammal}(x)\}$

(6) $\forall x(\text{rabbit}(x) \rightarrow \text{mammal}(x))$

Field agrees that (5) is committed to abstracta, a set of rabbits and a set of mammals. In fact, statement (5) asserts the *existence* of sets Z_1 and Z_2 such that Z_1 is the set of rabbits and Z_2 is the set of mammals and $Z_1 \subseteq Z_2$. However, he insists that (6) is free of ontological commitment to anything but concrete rabbits and mammals. Intuitively speaking, (5) is an "abstract counterpart" of the non-mathematical statement (6).

Now introduce standard set-theory with ur-elements ZFU which contains a theorem:

(7) $\exists Z \forall x(F(x) \leftrightarrow x \in Z)$

for any non-mathematical predicate F. (This theorem follows from the specification that F is an ur-element predicate, so $\forall x(F(x) \rightarrow U(x))$ is a theorem, that *there is a set* of ur-elements, and the Axiom of Separation. This was explained in Chapter 2).

Field's idea is that (5) is an abstract counterpart of (6) because, given standard applicable set theory ZFU, we can prove the equivalence, (5) $\leftrightarrow$ (6). Thus,

$$\text{ZFU} \vdash (\{x: F(x)\} \subseteq \{x: G(x)\}) \leftrightarrow \forall x(F(x) \rightarrow G(x))$$

In this way, we can show that certain set-theoretical statements (which are *not* theorems of ZFU!) are "abstract counterparts" of certain non-mathematical statements about concreta.

This account readily transfers to the application of arithmetic. Consider the pair of statements,

(8) $\#F = m$

(9) $\exists_m x F(x)$

where F is any urelement predicate true only of concreta (e.g., 'apple' or 'space-time region') and m is any specific numeral.

The statement (9) uses the *numerically definite quantifier* $\exists_m$, where m is a specific numeral (0, $s(0)$, $s(s(0))$, etc.). The theory of such quantifiers is given by the following recursive definition:

(Def$_0$) $\exists_0 x F(x) \leftrightarrow_{df} \neg\exists x F(x)$

(Def$_1$) $\exists_{\geq s(0)} x F(x) \leftrightarrow_{df} \exists x F(x)$

(Def$_2$) if $n \neq 0$, then $\exists_{\geq s(n)} x F(x) \leftrightarrow_{df} \exists x(F(x) \wedge \exists_{\geq n} y(y \neq x \wedge F(y))$

(Def$_3$) if $n \neq 0$, then $\exists_n x F(x) \leftrightarrow_{df} \exists_{\geq n} x F(x) \wedge \neg\exists_{\geq s(n)} x F(x)$

(Def$_2$ says that: there are at least $n + 1$ Fs just in case there is an x such that Fx and there are least n things y such that $y \neq x$ and Fy. Def$_3$ says that: there are exactly n Fs just in case there are at least n Fs and it is false that there are at least $n + 1$ Fs).

We can show (model-theoretically) that these definitions capture what is meant by specific attributions of finite cardinality.

Theorem 1: Correctness of the Numerical Quantifiers

For any finite whole number n and any interpretation $\Im$,

if $\Im \models \exists_n x F(x)$, then $|\Im[F]| = n$.

Proof: This is proved by induction as follows. It clearly holds for $\exists_0 xF(x)$. Assume it holds for n and that it does not hold for $n + 1$. Thus, there is an $\mathfrak{I} = (D, F)$ such that $\mathfrak{I} \models \exists_{s(n)} xF(x)$ and $|\mathfrak{I}[F]| \neq n + 1$. Thus, $|\{x: F(x)\}| \neq n + 1$. And, from the definition of $\exists_{s(n)}$, $\mathfrak{I} \models \exists x(F(x) \wedge \exists_n y(y \neq x \wedge F(x)))$. So, there is an $e \in D$ such that $e \in \mathfrak{I}[F]$ and $\mathfrak{I}^x_e[\exists_n y(y \neq x \wedge F(x))] = \mathbf{T}$. But, by assumption, if $\mathfrak{I} \models \exists_n y(G(y))]$, then $|\mathfrak{I}[G]| = n$. Thus, $|\{x: F(x) \wedge x \neq e\}| = n$. Thus, since $e \in F$, we conclude that $|\{x: F(x)\}| = n + 1$. Contradiction. ■

Returning to the counterpart relation, Field agrees that (8) is committed to an abstractum (the number m). However, he insists that (9) is free of ontological commitment to anything but concrete Fs. Of course, for (9) to be true, there must be m Fs, but there needn't exist the number m in addition to the m concrete Fs. Again, the Putnam-Field idea is that (8) is an abstract counterpart of (9) in the sense that, given standard applicable (platonistic) arithmetic T, we can prove that:

(*) for any canonical numeral $\mathbf{n}$, $\mathsf{T} \vdash (\#F = \mathbf{n}) \leftrightarrow \exists_n xF(x)$.

Let us refer to this formula $(\#F = \mathbf{n}) \leftrightarrow \exists_n xF(x)$ as Red_n. Presumably, it is possible to prove each formula Red_n in second-order Frege Arithmetic FA, using the recursive definition of the numerically definite quantifiers above. To prove this inductively, we first show that Red_0 is a theorem. Next we assume that Red_n is a theorem of FA. We want to show that $\mathsf{Red}_{s(n)}$ is also a theorem of FA. It then follows by induction that, for each natural number n, $\mathsf{FA} \vdash \mathsf{Red}_n$. That proves (*).

To summarize, Putnam and Field have explained how a non-mathematical statement $\varphi°$ may be correlated with an "abstract counterpart" φ, where the condition for $\varphi°$ to be an abstract counterpart of φ is simply that $\mathsf{M} \vdash \varphi \leftrightarrow \varphi°$, where M is the background (platonistic) mathematical theory.

7.6 Hermeneuticism in the Philosophy of Mathematics

Hermeneutics is "the science of interpretation, especially Scripture"[141]. A hermeneuticist is an exponent of the science of interpretation. Hermeneutics has been an integral part of philosophy down the ages, for philosophers have traditionally been concerned with what certain kinds of statements "really mean", are "intended" to mean, or how they are to be "interpreted".

Modern theologians, perhaps sensing the untenability (that is, falsity[142]) of a literal interpretation of Biblical stories about "God", "the Holy Spirit", "Creation", "Adam and Eve" (and other myths) seek to inform us what religious doctrines (about the wrath of God, or transubstantiation, say) are "really about". Namely, *something else.*

> ... reconstructive nominalists may be likened to those ecumenically minded thinkers who have suggested that religion can be made perfectly congenial to humanists by (re)-interpreting religious language so that 'God' refers, not to a transcendent supernatural being, but to something more innocuous, such as the good in human beings or an immanent historical process of liberation and enlightenment. But there is a great difference between offering such a reinterpretation as a substitute for more traditional creeds in which humanists have lost faith or offering it as an exegesis of what the canonical scriptures have really meant all along, despite appearances to the contrary that have misled the unsophisticated.
>
> (Burgess & Rosen 1997, p. 6).

In the context of debates about mathematical realism, the hermeneutic philosopher of mathematics promises to provide a "correct" (but non-literal) interpretation of mathematical discourse.

> On what may be called the hermeneutic conception [of nominalism], the claim is instead 'All anyone really means—all the words really mean—is ...' (here again giving the reconstrual or reinterpretation). Reconstrual or reinterpretation is taken to be an analysis of what really 'deep down' the words of current theories have meant all along, despite appearances 'on the surface'. It is taken to be a means to the end of substantiating the claim that nominalist disbelief in numbers and their ilk is in the fullest sense compatible with belief in current mathematics and science.
>
> (Burgess & Rosen 1997, pp. 6-7).

---

141 Chambers English Dictionary, from the Greek *hermeneutikos*, from the god Hermes.

142 We philosophers are inclined to rephrasing strong assertions by weaker, often subjective, assertions. E.g., to say "it seems to me that ..." instead of "it is true that ...", or "it is implausible (or untenable) that ...." instead of "it is false that ...". I acquiesce, but register my protest.

For example. Suppose a group of mathematicians are debating the existence of non-modular elliptic functions[143]. The hermeneutic philosopher of mathematics suggests that such talk is somehow *not* "really" about elliptic functions, but is somehow "really" about something else (perhaps constructible linguistic tokens, or merely possible structures, or performing concrete valid derivations, or even empirical matters of manipulating eggs).

Consider the simple existential arithmetic assertion,

(1) there is a prime number greater than 13

According to hermeneuticism, this is not literally true (there are no such things as numbers). However, there is some sense in which this is not even what it "really means". What (1) "really means" is to be given by a hermeneutic reconstrual, which will in fact turn out to be *true*, *not* committed to such things as numbers, and (preferably) more easily "knowable" or "epistemically accessible" that (1).

For three illustrative examples, consider the following:

The If-Thenist Reconstrual with respect to an axiom M (say PA^2):

$(1)^{\circ}$ M $\rightarrow$ (there is a prime number greater than 13)

The Modal Constructibility Reconstrual:

$(1)^{\circ\circ}$ there might have been a numeral token with the "primeness property" which is longer than '|||||||||||||'

The Structuralist If-Thenist Reconstrual:

$(1)^{\circ\circ\circ}$ 'there is a prime number greater than 13' holds in any ω-sequence

Obvious questions arise. Is it plausible to hold that, when mathematicians engage in mathematical discussion, they are "not really" making assertions about numbers, functions, sets, and so on? Is it plausible to maintain that what their assertions "really mean" is given by some hermeneutic reconstrual? What evidence is there for this reconstrual? Would it be evinced by experimental linguistics? (Burgess 1983 takes this

[143] The non-existence of such functions is equivalent to Fermat's Last Theorem (FLT): this is how Andrew Wiles proved FLT in 1994.

quite seriously: the hermeneuticist is making a claim about what statements "really mean" and such a claim should presumably be supported by empirical linguistic evidence).

Is some "reconstrual" already *in the mind* of the mathematician? Then it might be appropriate simply to ask the mathematician what he or she means by asserting, or seeming to assert, some theorem. (Perhaps, one might have to hypnotize the mathematician). I think it would be conceded that this would lead nowhere. For it is just *not true* (and obviously not true) that mathematicians always have some reconstrual "in mind" when they assert things like the existence of non-standard models of first-order analysis, or the existence of a homomorphism between SU(2) and SO(3).

But if a mathematician asserts a theorem, and *philosophers* are not satisfied with it (because it implies the existence of mathematicalia), then to what extent is it even legitimate to insist that the mathematician must "reconstrue" his or her theorem as some philosophically acceptable reconstrual? What exactly makes it legitimate for philosophers to tell mathematicians what they mean? Philosophy has hardly covered itself in glory in its analyses of "meaning" (e.g., the logical positivists were blatantly wrong about the meaning of theoretical statements).

Furthermore, if a mathematical theorem is really about the nominalistically acceptable entities quantified over in its "correct" reconstrual (possible numeral tokens, say, or possible structures), then *why exactly* do mathematicians bother with the indirection, with the talk of real numbers, sets, functions, modular curves, topological spaces and so on?

Despite my serious misgivings about the whole programme of hermeneuticism in the philosophy of mathematics, I think it is worthwhile trying to get the basic components of the programme clear.

7.7 Reconstrual: Adequacy Constraints

The underlying strategy of the hermeneutic programme about mathematical discourse is to provide a system of nominalistic reconstrual. Reconstrual is a form of *translation*. We have some puzzling mathematical statement φ and we attempt to find another statement φ°, the nominalistic "reconstrual" of φ. Since the statements that the hermeneuticist is reconstruing are often already formalized (or easily formalizable), the problem in many cases resolves into studying *translation mappings* from one formalized language L_M (say, a mathematicized notation for set theory, arithmetic, or analysis) to another L_N (say, a "mathematics-free" notation for space-time points, or numeral tokens, or modal logic).

A general or unified theory of translation between formalized languages does not exist, although fragments of such a treatment can be found, and indeed, the important notion of a (*relative*) *interpretation* is really just an example of a translation function, namely one that satisfies the constraint that theorems are mapped to theorems.

The idea of a translation function is simple. Suppose the signature of L^* is $(P_1^*, \ldots, P_m^*)$ and the signature of L is $(P_1, \ldots, P_k)$. Then an *elementary translation* from L^* to L is a map Γ from the formulas of L^* to those of L such that, first,

for each atomic formula $P_i^*(x_1, \ldots, x_n)$ in L^*,

$\Gamma(P_i^*(x_1, \ldots, x_n))$ is an L-formula containing exactly the same free variables.

In fact, we shall suppose that,

$$\Gamma(P_i^*(x_1, \ldots, x_n)) = \Psi_i(P_1, \ldots, P_k)(x_1, \ldots, x_n)$$

where $\Psi_i(P_1, \ldots, P_k)(x_1, \ldots, x_n)$ is an L-formula built from the P_i. We shall call the formulas Ψ_i the *defining formulas* of Γ. We shall see below that they determine the associated "structure map".

And, second, where φ^* and χ^* are L^*-formulas,

$$\Gamma(\neg\varphi^*) = \neg\Gamma(\varphi^*)$$

$$\Gamma(\varphi^* \wedge \chi^*) = \Gamma(\varphi^*) \wedge \Gamma(\chi^*)$$

$\Gamma(\exists x_n \varphi^*) = \exists x_n \Gamma(\varphi^*)$[144]

It is possible to verify (model-theoretically) that if $\Delta^* \vdash \varphi^*$, then $\Gamma(\Delta^*) \vdash \Gamma(\varphi^*)$, so that implication is preserved. However, it is possible that Δ^* be consistent while $\Gamma(\Delta^*)$ is inconsistent. For example, let L^* be the propositional notation (p^*, q^*). Define a map Γ from L^* to a propositional notation (p, q) thus:

$\Gamma(p^*) = p$

$\Gamma(q^*) = p \wedge q$

Clearly, the consistent L^*-formula $\neg p^* \wedge q^*$ is mapped to the inconsistent $\neg p \wedge (p \wedge q)$. So, consistency-preservation is non-trivial constraint.

What about truth-value preservation? To answer this, we need to look at the fundamental theorem of translation, what I call the "Co-extensionality Theorem". To explain it, we need to define the induced "structure map". Let $A = (D, R_1, \ldots, R_n)$ be an L-structure. Then, Γ induces a map (we'll call it Γ also) from A to some L^*-structure, $\Gamma(A) = (D^*, R_1^*, \ldots, R_m^*)$. This structure map is defined as follows:

$\mathrm{dom}(\Gamma(A)) = \mathrm{dom}(A)$

$R_i^* = \Psi_i(A)$ [i.e., $\Gamma(A)[P_i^*] = \Psi_i(A)$]

Each defining formula Ψ_i first-order defines a relation $\Psi_i(A)$ on the structure A. I.e., $(a_1, \ldots a_n) \in \Psi(A)$ iff $A \models \Psi[a_1, \ldots, a_n]$.[145] The idea is that this relation $\Psi_i(A)$ is then identified as the *i*th relation in the L^*-structure $\Gamma(A)$, the extension of the L^*-predicate P_i^*.

Then we can prove:

---

[144] What model theorists call an *interpretation* is slightly more general (see Hodges 1997, Ch. 4). A "reduction formula" is introduced in image notation (L) and all translations of quantifications are then *relativized* to this formula. For example, consider the reduction of arithmetic in L_A to set theory in $\mathsf{L}_\in$. Any statement of arithmetic $\forall x\varphi$ is translated as the set-theoretical assertion $\forall x(x \in \omega \rightarrow \Gamma(\varphi))$. The reduction formula is $x \in \omega$. (Choosing Γ correctly, the result is that every axiom of PA, say, is mapped by Γ to a theorem of ZFC. We say: PA is *relatively interpretable* within ZFC).

[145] I use the usual model-theoretic notation: $A \models \Psi[a_1, \ldots a_n]$ just in case any A satisfies $\Psi(x_1, \ldots, x_n)$ when the sequence $(a_1, \ldots, a_n)$ from dom(A) is assigned to x_i. See Hodges 1997.

Theorem: The "Co-extensionality Theorem"

for any L*-formula $\varphi^*(x_1, \ldots, x_n)$ and any L-structure A,

$$A \models \Gamma(\varphi^*)[a_1, \ldots, a_n] \quad \Leftrightarrow \quad \Gamma(A) \models \varphi^*[a_1, \ldots, a_n].$$

Proof: Let φ^* be an atomic formula, say $P_i^*(x_1, \ldots, x_n)$. Then, $\Gamma(\varphi^*) = \Psi_i(x_1, \ldots, x_n)$. Then, assume that $A \models \Gamma(\varphi^*)[a_1, \ldots, a_n]$. Thus, $A \models \Psi_i[a_1, \ldots, a_n]$. Thus, $(a_1, \ldots, a_n) \in \Psi_i(A)$. Thus, $(a_1, \ldots, a_n) \in \Gamma(A)[P_i^*]$. Thus, $\Gamma(A) \models \varphi^*[a_1, \ldots, a_n]$. Conversely, suppose that $\Gamma(A) \models \varphi^*[a_1, \ldots, a_n]$. Then, similarly, $(a_1, \ldots, a_n) \in \Psi_i(A)$ and thus $A \models \Gamma(\varphi^*)[a_1, \ldots, a_n]$. By induction on complexity, a similar argument shows that the relation holds if φ^* is molecular. ■[146]

Specializing to closed formulas, we immediately deduce from the above the very useful "Equivalence Theorem":

$$A \models \Gamma(\varphi^*) \quad \Leftrightarrow \quad \Gamma(A) \models \varphi^*$$

For our purposes, to examine hermeneutic translations, what matters is translational adequacy. So, here is a simple list of some adequacy constraints:

a. Preservation of relations of implication and consistency;

b. Preservation of truth values;

c. Preservation of meanings;

d. Preservation of modal properties of necessity and possibility:

e. Preservation of non-mathematical consequences.

We have seen that implication is automatically preserved by Γ, but consistency isn't. What about truth? Suppose that (L*, A^*) and (L, A) are *interpreted* first order languages and Γ: L* → L an elementary translation. Then we can prove:

Theorem: Truth Preservation

[146] Although it took me two years to arrive here, this is in fact a *standard result* in model theory, known as the "Reduction Theorem". See Hodges 1997, Chapter 4.

$A^* \equiv \Gamma(A)$ if and only if, for any $\varphi^* \in {}_0\mathsf{L}^*$, $A^* \models \varphi^*$ iff $A \models \Gamma(\varphi^*)$.

Proof: Suppose $A^* \equiv \Gamma(A)$. Then, for any $\varphi^* \in \mathsf{L}^*$, $A^* \models \varphi^*$ iff $\Gamma(A) \models \varphi^*$. By the Equivalence Theorem, we deduce that $A^* \models \varphi^*$ iff $A \models \Gamma(\varphi^*)$. Conversely, suppose that for any $\varphi \in \mathsf{L}^*$, $A^* \models \varphi^*$ iff $A \models \Gamma(\varphi^*)$. Then, quickly, $A^* \models \varphi^*$ iff $\Gamma(A) \models \varphi^*$, so $A^* \equiv \Gamma(A)$. ■

This shows that the assumption of elementary equivalence of A^* and $\Gamma(A)$ is equivalent to the constraint that the translation Γ from (L^*, A^*) to (L, A) preserves all truth values. Thus, if A^* and $\Gamma(A)$ are *not* elementarily equivalent, truth preservation will fail. In fact, it is quite trivial to find a pair of interpreted languages and a translation that does not preserve truth. E.g., take (p^*, **T**) and (p, **T**) and the translation map $\Gamma(p^*) = \neg p$.

An odd thing about any reconstrual mapping introduced by some hermeneutic strategy is that it *cannot* preserve truth values. We stressed in Chapter 6 that the hermeneutic nominalist cannot think that mathematical assertions are literally (disquotationally) true; but his or her idea is to map these literal *falsehoods* onto *truths* about the concrete domain. For example, Chihara thinks that there are no such things as prime numbers, and so must think that the mathematical assertion,

> For any prime number, there is prime that is larger

is literally *false*. But, his strategy involves translating this (allegedly false) mathematical assertion to a (rather bizarre) *true* assertion, namely,

> For any numeral token with the "*P*-property" that it is possible to construct, it is possible to construct a longer token with the "*P*-property".

(where he must define 'token x has the *P*-property' so that the resulting, mostly unactualized (!), set of such tokens work as "surrogates" for the prime numbers).

There are number of neat results about the structure map Γ. For example, it preserves elementary equivalence and isomorphism. So, if $A_1 \equiv A_2$, then $\Gamma(A_1) \equiv \Gamma(A_2)$[147]; and if A_1

[147] This follows quickly from the "Equivalence Theorem" above.

$\approx A_2$, then $\Gamma(A_1) \approx \Gamma(A_2)$[148]. There are also some interesting connections between the properties of injectiveness/surjectiveness for the translation map and the corresponding properties for the associated structure map. One can prove that if the structure map is surjective, then the translation map is *consistent* (and this result strengthens to "iff" in the case of propositional languages). Furthermore, one can show that if the translation map satisfies appropriate surjectiveness and injectiveness condition, then it is *invertible*; and then, if $A^* = \Gamma(A)$, that these structures are *definitionally equivalent* (as are the theories of the structures).

Returning to adequacy constraints in reconstrual, we need to quickly mention "meaning preservation" and preservation of modal properties. It is clear that extensional model-theoretic semantics has nothing to say about either of these topics (unless modality is understood metalogically: see next Chapter). Occasionally a philosopher will suggest that the meaning of a sentence is connected to its class of models (or that the proposition expressed by a sentence is its equivalence class under narrow logical equivalence). This is nonsense, for 'Derrida is a genius' and 'Derrida is a charlatan' have the same models but express different propositions. Similarly, 'Hace calor hoy' and 'It's hot today' are plainly not logically equivalent, but happen to express the same proposition (they are *synonymous*, and synonymy is not the same as logical equivalence).

The final adequacy constraint for a hermeneutic nominalist to honour is *preservation of mathematics-free consequences*. Take our stand-by theory, Tim. This theory implies that time is dense and continuous (in technical terms, order-complete). If a hermeneuticist wishes to translate this into his or her favourite language, the result Tim* must still imply these things. We saw in Chapter 6 that the so-called "if-thenist" cannot cope with this. Just as he or she translates each assertion φ in the language of a mathematical theory M as the conditional M $\rightarrow \varphi$ (and then checks to see if the result is valid), he or she will perhaps wish to translate Tim as $\mathsf{ZF}^2 \rightarrow$ Tim. But this conditional is *not* valid (obviously we cannot prove contingent things about *time* from ZF^2 alone), and it does not imply the

148 To prove this, one needs to show that if a relation R^* on a domain D^* is the image of another R on a domain D under a bijection ρ, then various relations $\exists R^*$ and $\exists R$, $D^* - R^*$ and $D - R$, and so on (generated by logical operations) are also in bijective correspondence.

mathematics-free consequences implied by Tim. We shall see in the next Chapter that Hellman's modal proposal also falls foul of this requirement (the solution to this technical hiccup involves something rather *metaphysical*, called an "actuality operator").

7.8 Reconstruction Techniques à la Burgess

In this section Burgess's discussion of various nominalistic reconstruction techniques is outlined (see Burgess 1984 and Burgess & Rosen 1997).

Imagine a two-sorted theory T in a two-sorted notation L. This theory T talks about primary entities (i.e., in our case, concreta) and secondary entities (in our case, mathematicalia). Assume that T implies the existence of secondary entities. What we want to provide is a nominalistic reconstruction T* in L* which satisfies the following constraints:

(1) T* in L* expresses and derives all the information about primary entities expressible and derivable in T in L.

(2) T* in L* no longer talks about nor implies the existence of secondary entities.

It is worth emphasizing some aspects of constraint (1) which were not made clear in Chapter 3. Constraint (1) is a *conservativeness* constraint. In Chapter 3, we proved the "trivial indispensability theorem", which says that an EMT T in a mathematicized notation L might not be *deductively* conservative over its primary restriction T° in L° (i.e., the result of deleting all mixed and secondary axioms). An example is our stand by, the analytically formulated theory of time Tim.

But there are in fact two important notions of conservativeness:

i. Deductive Conservativeness

T in L is a *deductively conservative extension* of T* in L* if and only if, for any closed formula φ^* in ${}_0L^*$, if $T \vdash \varphi^*$ then $T^* \vdash \varphi$

ii. Expressive Conservativeness

> T in L is an *expressively conservative extension* of T* in L* if and only if for any formula $\varphi(x_1, ..., x_n)$ in L containing only primary free variables, there is a formula $\varphi^*(x_1, ..., x_n)$ in L* containing the same variables such that T $\vdash \forall x_1 ... \forall x_n(\varphi \leftrightarrow \varphi^*)$.

We have explained and utilized the notion of deductive conservativeness before, in Chapter 3 (to prove the trivial non-conservativeness of some theories over their primary restrictions: e.g., Tim is not a deductively conservative extension of Tim°) and Chapter 5 (to prove the conservativeness of the deflationary theories of truth and the non-conservativeness of Tarski's theory of satisfaction).

The notion of expressive conservativeness is explained in Burgess & Rosen 1997:

> If for every formula *F* in the language L of T having only primary free variables, there is a formula *F*° of the language L° of T° with the same free variables such that it is deducible from T that *F* and *F*° hold of exactly the same primary entities, then in jargon T is called an *expressively conservative* extension of T°. In this case, the class (if there is just one free variable) or relation (if there are several free variables) determined by a formula *F* of L is also determined by the formula *F*° of L°, and in this sense any assertion about classifications of or relationships among primary entities that is expressible in L is (according to T itself) already expressible in L°.
>
> (Burgess & Rosen 1997, pp. 83-84).

Expressive conservativeness is clearly a stronger constraint than mere deductive conservativeness. Indeed, it is a natural constraint on any adequate reconstruction of a mathematicized theory like Tim. Roughly, a mathematicized theory might imply facts about primary (concrete) entities which are not in any way represented in the primary consequences. Such facts are expressed using *mixed* predicates. For example, T in L might imply a mixed consequence $\forall x \exists X R(x, X)$, for some mixed formula $R(x, X)$. This consequence says that *every primary entity x* has a certain (mathematically definable) property (i.e., the monadic property of *primary entities* expressed by $\exists R$). This consequence will automatically be omitted from the primary restriction T°, for it quantifies over secondary entities (mathematicalia).

So constraint (ii) demands that mixed consequences like $\forall x \exists X R(x, X)$ *are* "preserved" in the reconstruction T* in L*, by a *reconstrual* within L* of the mixed formula $\exists X R(x, X)$. One cannot simply *delete* this concept $\exists R$ which concerns primary entities, for then one simply deletes a piece of information about primary entities, namely $\forall x \exists X R(x, X)$.

For example, suppose the mixed formula $\exists X R(x, X)$ classifying primary entities is in the language L of T. The requirement of expressive conservativeness is that there be a primary formula $R^*(x)$ of L* such that T $\vdash \forall x(\exists X R(x, X) \leftrightarrow R^*(x))$. As Burgess & Rosen 1997 point out, if there is no such formula $R^*(x)$, then the factual classification of all those primary entities that satisfy $\exists X R(x, X)$ is simply not expressed within L*.

Burgess & Rosen 1997 sketch three technical methods of achieving such a reconstructive goal:

a. Tarskian elimination

b. Skolemite elimination

c. Craigian elimination

We shall quickly describe these techniques now.

a. Tarskian Elimination

The term 'Tarskian elimination' derives from a technique introduced in a 1953 monograph (quite unrelated to nominalism) by Tarski, Mostowski and Robinson. Tarskian elimination is the most natural form of elimination of secondary entities which conforms to the constraints (i) and (ii) above.

We say that a theory T in a two-sorted notation L has the *elimination property* just in case it has a merely redundant extension $T^{\ddagger}$ in an extended notation $L^{\ddagger}$ that is fully conservative over its primary restriction.[149]

[149] $T^{\ddagger}$ in $L^{\ddagger}$ is a *merely redundant extension* of T in L just in case it is a deductively and expressively conservative extension of T in L. See Burgess & Rosen 1997, pp. 83-85.

Tarskian elimination depends upon T in L satisfying a certain strong representational constraint. Briefly, the appropriate metatheorem says that,

if T in L has the *representation property*, then it has the elimination property

T in L has the representation property just in case there is a mixed "representation formula" $R(x_1, ..., x_n, X)$ in L such that,

A. $\mathrm{T} \vdash \forall X \exists x_1 ... \exists x_n R(x_1, ..., x_n, X)$

saying that every secondary entity X is represented by some n-tuple $x_1, ..., x_n$ of primary entities, and

B. $\mathrm{T} \vdash \forall x_1 ... \forall x_n \forall X_1 \forall X_2 [(R(x_1, ..., x_n, X_1) \wedge R(x_1, ..., x_n, X_2)) \rightarrow X_1 = X_2]$

saying that there is at most one secondary entity X represented by any n-tuple $x_1, ..., x_n$ of primary entities.

Burgess & Rosen 1997 point out that "by modifying the definition of R, arbitrarily designating some one secondary entity (such as *0* in the case where the secondary entities are real numbers) and arbitrarily stipulating that a k-tuple that did not represent anything else under the original definition of R is to be counted as representing this special secondary entity, one may also assume",

C. $\mathrm{T} \vdash \forall x_1 ... \forall x_n \exists X R(x_1, ..., x_n, X)$

What is going on here? Roughly, the "representation theorems" say that each mathematical entity has a "concrete counterpart". E.g., each real number r (e.g., π) might be represented by some concrete geometric space-time line segment L_r. The idea is to do away with the abstract reals in favour of their representatives, the line segments L_r. (The problem that arises is when set theory enters the picture. Unless our physical theory contains a *spectacular physical assumption* that every abstract transfinite set has a "concrete exemplification", the Tarskian method does not work to eliminate a set theoretic ontology).

On the assumptions (A)-(C), the technical metatheorem is then not difficult to prove. Namely, that there is a merely redundant extension $\mathrm{T}^\ddagger$ in some extended notation $\mathrm{L}^\ddagger$ that it is fully conservative over its primary restriction $T^{\ddagger\circ}$. The proof is given on pp. 87-90 of Burgess & Rosen 1997. They also make clear that this elimination method actually provides an implication-preserving *reconstrual* Γ: $L \mapsto L^{\S}$ such that, for any primary formula φ in L°, $\Gamma(\varphi) = \varphi$. In particular, it follows that if φ is a closed primary formula, then if $T \vdash \varphi$, then $\Gamma(T) \vdash \varphi$.

b. Skolemite Elimination

If we drop the demand that there exist a representation formula for T (i.e., a formula in L, $R(x_1, \ldots, x_n, X)$ such that existence and uniqueness principles follow from T), an easier method of reduction is available. One expands the original two-sorted language L to L^+ by adding a new unexplained primitive representation predicate ®(x, X), and one adds as axioms governing this predicate,

i. $\forall X \exists ! x$®(x, X)

ii. $\forall x \exists ! X$®(x, X)

to obtain a new theory T^+. The method of eliminating quantification with respect to X then proceeds as before.

The bizarreness of such a procedure is easy to describe. Suppose T contains set theory, which posits a vast transfinite hierarchy of sets. To eliminate quantification over these sets, one must introduce the stupendous physical assumption that *every* transfinite rank is represented by some unique concrete physical object. This assumption, that everything implied to exist in standard set theory has a physical exemplification, is a pretty hard one to swallow.

The resulting theory T^+ in L^+ can be proved deductively conservative over T using the Löwenheim-Skolem Theorem[150].

[150] See Burgess & Rosen 1997, pp. 91-92.

c. Craigian Elimination

Let us say, just for a moment, that a theory T is "nice" when it is recursively axiomatizable. It can be specified by an effectively decidable class of axioms (which we identify with T)[151]. The theory proper is just the *deductive closure* of T. If the class of axioms of T is decidable, then, under a Gödel coding system, the Gödel numbers of its class of axioms is *recursive*. On this assumption, the class of theorems of T is *recursively enumerable*[152].

It is possible to show that, for any r.e. set of numbers A, there is a Turing Machine M_A which prints out all and only the elements of A. Then, Craig's Reaxiomatization Theorem (Craig 1953) is this:

> Any deductively closed r.e. set of formulas is the deductive closure of a recursive set of axioms[153].

This means that any deductively closed r.e. set of formulas can be recursively axiomatized (no matter how inelegantly). That is, if Δ is r.e. and closed, then there is a decidable (recursive) set of formulas Δ^* such that $\Delta = \mathsf{Cn}(\Delta^*)$.

Now consider a recursively axiomatized "nice" theory T in a two-sorted notation L. Consider the deductive closure of T, Cn(T). This set is r.e. Next consider definable

---

[151] Incidentally, by Gödel's Incompleteness Theorems, full first-order arithmetic is not a "nice" theory! The class of truths in $\mathfrak{N}$ is not recursively axiomatizable. "Niceness" is a kind of *accessibility constraint*, that we impose because we think of the human mind as a "finite recognition device". But Gödel's Incompleteness Theorem, as Putnam once remarked, implies that the whole truth can never be fully "accessed".

[152] Any set A of numbers is r.e. if and only if there is a recursive relation R such that, for any n, $n \in A$ iff $\exists m R(m, n)$. E.g., the set of (g.n.s of) theorems of Robinson Arithmetic Q is not recursive, but is r.e. Indeed, the set of valid theorems of first-order logic is not recursive, but is r.e. (Church's Theorem).

[153] A proof of Craig's Theorem is to be found in Boolos & Jeffrey 1989, solution to Exercise 15.4, pp. 180-181. Another proof sketch is given in Burgess & Rosen 1997, p. 93. The proof is this. Assume that X is a set of formulas. Suppose that X^*, the class of g.n.s of X's members, is r.e. So, there is a recursive relation R such that $X^* = \{n: n \text{ is the g.n. of a formula in X}\} = \{n: \exists m R(n, m)\}$. The proof involves constructing a decidable set Y of formulas whose closure is X. For any formula φ and number n, let $\varphi^n = (\varphi \wedge \varphi) \wedge \varphi \wedge \ldots \wedge \varphi$, the conjunction of $n + 2$ occurences of φ. Let $Y = \{\varphi^n: R(\#\varphi, n)\}$. If $\varphi \in X$, then for some n, $R(\#\varphi, n)$, so $\varphi^n \in Y$. And if $\varphi^n \in Y$, then $\varphi \in X$. But φ and φ^n are equivalent, so X and Y must imply all the same fomulas, and thus X must be the deductive closure of Y. Finally, Y is decidable. Because R is recursive, it is a decidable matter whether or not $R(\#\varphi, n)$, for given φ and n. If $R(\#\varphi, n)$, then $\varphi^n \in Y$. If not-$R(\#\varphi, n)$, then $\varphi^n \notin Y$. This set Y is thus a decidable axiomatization of X.

restrictions of Cn(T). In particular, the class Cn°(T) in L°, the class of formulas containing only *primary notation* in L. That is, the primary restriction of Cn(T). If Cn(T) is r.e., then Cn°(T) is also r.e.[154] Since Cn°(T) in L° is r.e., then it is recursively axiomatizable, by Craig's Theorem. Call the resulting decidable set of axioms Craig°(T). This is a "nice" theory in L° which has exactly the same consequences in L° as does T in L. In short, Cn(Craig°(T)) = Cn°(T). Clearly, Craig°(T) does not imply the existence of secondary entities. Indeed, it talks only of primary entities. However, it implies all the original primary consequences of T. If T is a mathematicized theory in a two-sorted mathematicized notation L, then Craig°(T) is a potential nominalistic replacement for T.

The question now is: is Craig°(T) a "good replacement" for T?

A common objection to Craigian reaxiomatizations is that they omit much of the mixed "internal information" about primary entities encoded in T. That is, T is not an *expressively conservative* extension of Craig°(T). For example, if T implies a mixed consequence $\forall x \exists X R(x, X)$, then this consequence, which says that every primary entity x has a certain mathematically definable property (i.e., the monadic property $\exists R$), is simply *omitted* from the Craigian reaxiomatization, for it mentions secondary entities. More exactly, although T is, by construction, a *deductively* conservative extension of Craig°(T), it need not be an *expressively* conservative extension of Craig°(T).

Burgess & Rosen 1997 describe a quick and novel route around this objection. First extend T in L to a new theory T^+ in L^+ by adding, for each formula $P(x_1, \ldots, x_n)$ in L with no free secondary variables a new primitive $F_P(x_1, \ldots, x_n)$ along with an axiom B_P, $\forall x_1 \ldots \forall x_n[F_P(x_1, \ldots, x_n) \leftrightarrow P(x_1, \ldots, x_n)]$. One obtains a new theory T^+ in L^+, which, by construction, is *expressively conservative* over T. But note: L^+ now has *infinitely many primitives*, F_P. Now suppose $T^+ \vdash \varphi$, where φ is a primary assertion (no secondary variables). Simply add each such nominalistic consequence φ as an axiom of a new

[154] When T is recursively axiomatized, there is a Turing Machine M_T that prints out all (and only) the theorems of T. This machine need only be modified so as to print out all the theorems containing just primary notation, and this condition is effectively checkable (a sub-machine is easy to construct such that for any input formula φ, the sub-machine decides whether or not φ is a primary formula).

theory $T^{\ddagger}$. Finally, discard from the extended language L^{+} all secondary and mixed assertions, to obtain a primary language $(L^{+})^{\circ}$. The restriction of $T^{\ddagger}$ to this language is the final theory. Call this theory T°. Again, as Burgess & Rosen make clear, it is possible to apply Craigian reaxiomatization to obtain a decidable set of axioms T^{∞} whose deductive closure is just T°. The advantage of this method is that $T^{\ddagger}$ yields an *expressively conservative extension* of T.

However, there is a consensus (even among nominalists) that there is something very fishy about the latter two techniques of non-Tarskian reduction: Skolemite and Craigian eliminations. Skolemite elimination simply introduces a "completely unexplained representation primitive and completely unjustified existence and uniqueness axioms" (Burgess & Rosen 1997, p. 92). The Craigian reaxiomatization is "little more than a formal counterpart of the instrumentalist 'theory' consisting of the bare assertion that concreta behave *as if* abstracta existed and standard scientific theories are true" (Burgess & Rosen 1997, p. 94).[155]

Burgess & Rosen conclude that,

> Ultimately dissatisfaction with non-Tarskian reduction [Skolemite, Craigian] depends upon the grounds for dissatisfaction with instrumentalism, on the grounds for dissatisfaction with a merely negative, destructive nominalism, on the motivation for engaging in a positive, reconstructive nominalistic project in the first place.
>
> (Burgess & Rosen 1997, p. 94).

7.9 Two Strategies: Modal and Geometrical

In the next two Chapters, we shall concentrate on two specific nominalistic strategies for reconstruction of mathematicized theories:

[155] See Putnam 1979, Chapter 14, 'Craig's Theorem', for a further discussion.

a. Modal Strategies

This group of strategies is based on *modal logic*, having its roots in work by Putnam 1967 and Chihara 1973. This strategy provides quite a powerful way of reconstruing mathematics within modal logic (indeed, *suspiciously* powerful!).

b. Geometrical Strategy

This strategy is geometrical and based on the idea of aiming for a more recognizably nominalistic (non-modal) physics. It has its roots in Field 1980 (and in some respects, in Hilbert 1899). Although this strategy provides a way of reconstruing *mathematical analysis* (real numbers) within geometry, it cannot get rid of those nasty big sets.

We examine these strategies in the next two Chapters.

CHAPTER 8

Modalism in Mathematics

> ... there is not, from the mathematical point of view, any difference between the assertion that there exists a set of integers satisfying an arithmetical condition and the assertion that it is possible to select integers so as to satisfy the condition. Sets, if you will forgive me for parodying John Stuart Mill, are permanent possibilities of selection.
>
> Hilary Putnam 1967b (1996), 'Mathematics Without Foundations', p. 174.

> The basic idea of the approach to be taken in this work is to develop a mathematical system in which the existential theorems of traditional mathematics have been replaced by constructibility theorems: where, in traditional mathematics, it is asserted that such and such exists, in this system it will be asserted that such and such can be constructed.
>
> Charles Chihara 1990, *Constructibility and Mathematical Existence*, p. 25.

8.1 Modalism: A Counter-Revolutionary Movement?

When demonstrative mathematics emerged in Ancient Greece, much mathematical talk was deeply "constructive", "dynamic" or "modal"[156]. Mathematical axioms and postulates were standardly framed as assertions about what "it is possible to construct", and so on. The pattern is set by Euclid's *The Elements*[157], as described in Boyer 1968,

> In most manuscripts of *The Elements* we find the following ten assumptions:
>
> *Postulates*. Let the following be postulated,
>
> 1. To draw a straight line from any point to any point
> 2. To produce a finite straight line continuously in a straight line
> 3. To describe a circle with any centre and radius
> 4. That all right angles are equal
> 5. That, if a straight line falling on two straight lines makes the interior angles on the same side less than two right angles, the two straight lines, if produced indefinitely, meet on that side on which the angles are less than two right angles.

[156] These descriptions are taken from Shapiro 1997, Chapter 6, 'Construction, Modality, Logic'.

[157] For a standard translation of Euclid see *The Thirteen Books of Euclid's Elements*, translated and edited by T.L. Heath (Heath 1956).

Common Notions:

1. Things which are equal to the same thing are also equal to one another
2. If equals be added to equals, the wholes are equal
3. If equals be subtracted from equals, the remainders are equal
4. Things which coincide with one another are equal to one another
5. The whole is greater than the part

(Boyer 1968, pp. 116-117).

Some of these postulates concern what "can be constructed" or "produced", in a rather literal sense, using a straight-edge and so called Euclidean "collapsible" compasses, as Boyer goes on to describe:

> Postulate 3 is interpreted in the very limited literal sense, sometimes described as the use of Euclidean (collapsible) compasses, whose legs maintain a constant opening so long as the point stands on the paper, but fall back upon each other when they are lifted. That is, the postulate is not interpreted to permit the use of a pair of dividers to lay off a distance equal to one line segment upon a noncontinuous longer line segment, starting from an end point. It is proved in the first three propositions of Book I that the latter construction is always possible, even under the strict interpretation of Postulate 3. The first proposition justifies the construction of an equilateral triangle ABC on a given line segment AB by constructing through B a circle with a centre at A and another circle through A with centre at B, and letting C be the point of intersection of the two circles
>
> (Boyer 1968, p. 117).

Plato protested as this talk of "constructing", "extending", "producing", "adding" and so on. Plato urged that mathematicalia (like line segments, triangles and circles) are "static" abstract entities in the real Parmenidean world of "Being", and not ephemeral entities in the imperfect Heraclitean world of "Becoming":

> Then if geometry compels us to view Being, it concerns us; if Becoming only, it does not concern us? . . . Yet anybody who has the least acquaintance with geometry will not deny that such a conception of the science is in flat contradiction to the ordinary language of geometricians. . . . They have in view practice only, and are always speaking, in a narrow and ridiculous manner, of squaring and extending and applying and the like—they confuse the necessities of geometry with the those of daily life; whereas knowledge is the real object of the whole science . . . the knowledge at which geometry aims is knowledge of the eternal, and not of aught perishing and transient.
>
> (Plato, *Republic*, Book VII, 527. In Plochman 1973, p. 442).

Whatever the intrinsic merits of Plato's Platonic protests, the *history* and practice of mathematics suggests that the Platonic vision has won out. A Platonic revolution occurs in mathematics every time a modal locution of the form "it is possible to find ..." or "it is possible to construct ..." is replaced by a static Platonic assertion of the form "there exists ...". We have what might be termed the *Platonic demodalization of mathematics.*

Consider the modal notion of provability. The locution "φ is provable in M", means, intuitively, "*it is possible to construct a proof* of φ in M". But the standard modern analysis of provability, a recent development *within* mathematics[158], indicates the demodalization nicely. Nowadays, "φ is provable in M" is taken to mean (in proof theory) "*there exists a derivation* Γ of φ from the axioms of M".[159]

For another example, consider the important locution with mathematical analysis,

> the real-valued function f is continuous at x.

This may be explained in the modal constructive vernacular as follows,

> $|f(x') - f(x)|$ *can be made* as small as you like by *taking* $|x' - x|$ ever smaller

The exact demodalized Platonic definition was not discovered until the nineteenth century, in the works of the great analysts Cauchy and Weierstrass, and results in the standard ε-δ definition:

> for every $\varepsilon > 0$, there is a $\delta > 0$ such that, for any x', if $|x' - x| < \delta$ then $|f(x') - f(x)| < \varepsilon$

Finally, perhaps the most famous example. Hilbert's axiomatization of Euclidean Geometry appeared in 1899. Within his presentation, he explicitly and deliberately replaced Euclidean modal or constructive locutions, like 'for any two points p and q, *one may construct* a line L such that ... ' by demodalized locutions like 'for any points p and q, *there exists* a line L such that ...'. Paul Bernays, one of Hilbert's later collaborators, described this 35 years later,

> If we compare Hilbert's axiom system to Euclid's, ignoring the fact that the Greek geometer fails to include certain postulates, we notice that Euclid speaks of figures to be *constructed*, whereas, for Hilbert, systems of points, straight

[158] The major figures in the development of the mathematical study of mathematical reasoning, mathematical proof theory, are Hilbert, Gentzen and Gödel.

[159] A standard text on modern (Gentzen-style) proof theory is Takeuti 1975.

> lines and planes exist from the outset. Euclid postulates: One can join two points by a straight line; Hilbert states the axiom: Given any two points, there exists a straight line on which both are situated. "Exists" refers here to existence in the system of straight lines.
>
> This example shows already that the tendency of which we are speaking consists in viewing the object as cut off from all links with the reflecting subject.
>
> Since this tendency asserted itself especially in the writings of Plato, allow me to call it "platonism".
>
> (Bernays 1935 (1983), pp. 258-259).

Surely the above are examples of fundamental *progress* within mathematics, introducing an improved and clearer understanding of our mathematical concepts. Examples like the notions of continuity and provability above suggest that sometimes progress *consists in* demodalization.

> The value of platonistically inspired mathematical conceptions is that they furnish models of abstract imagination. These stand out by their simplicity and logical strength. They form representations which extrapolate from certain regions of experience and intuition.
>
> (Bernays 1935 (1983), p. 259)

Naturally, modal or constructive talk continues in the vernacular of the working or teaching mathematician:

> Here are two ways of thinking about continuity of f at a.
>
> (1) In terms of approximations: we can ensure that $f(x)$ approximates $f(a)$ within any prescribed degree of accuracy by *choosing* x to approximate a sufficiently accurately.
>
> (2) Geometrically: given a horizontal band of any positive width 2ε centred on height $f(a)$, we *can choose* a vertical band of some suitable width 2δ centred on $x = a$ such that the part of the graph of f in this vertical band is also in the horizontal band.
>
> (Sutherland 1975, p. 14. Emphasis added).

Indeed, I am no Platonic puritan, having said many things like "*one may construct* a fixed point sentence ...". Of course, I view such colloquial assertions exactly as Plato would have done. As Burgess & Rosen 1997 put it,

> Historical relics of the former practice persist even today in colloquial mathematical language, as when one speaks of an integer or equation or function or space or formula that is divisible or solvable or differentiable or metrizable or provable, as being one that can be divided or solved or differentiated or metrized or proved—whereas the formal definition is that it is one for which there exists a divisor or solution or derivative or metric or proof.

(Burgess & Rosen 1997, p. 125).

If this is correct, the working mathematician will agree that the underlying reality behind constructive talk of divisibility, solvability, differentiability, metrizability or provability, the reality that he or she studies is a "static" realm of Platonic, abstract and demodalized mathematicalia: divisors, solutions, derivatives, metrics or proofs, and further of choice functions, isomorphisms, limits and so on.

Given this Platonic demodalization of mathematics, it is then rather surprising that some recent philosophical interpreters of mathematics have suggested *turning back the clock*. We describe some of these proposals below. But what is to be gained? Should we similarly return to Ptolemaic Astronomy, or to co-ordinate-dependent formulations of General Relativity? To ether theories of light? The Platonic tide is powerful. In a sense, Platonic demodalization is a component of the basic methodology of mathematicization. The suggestion that mathematics be *re*-modalized looks regressive and reactionary.

8.2 The Analysis of Modality

Almost all recent philosophical discussion of modality occurs within the context of various formalized treatments of modal logic. Ironically, modern modal logic is, in a certain sense, developed *non-modally*. (The semantics for modal logic is developed using Platonic demodalized mathematics). We have in the possible-worlds semantics of Kripke[160] (and others, like Kanger and Hintikka) a sort of Platonic demodalization of modality itself!

Within the recent philosophical literature on modality, there are roughly two ways of understanding modal notions: the *metalogical* construal and the *metaphysical* construal.[161]

160 Kripke's original paper is Kripke 1959. A more philosophical approach is Kripke 1963.

161 Of course, other construals of modality are possible: e.g., a proposition may be "physically impossible", in the sense of being logically incompatible with a "necessary" law of Nature.

i. The Metalogical Conception of Modality

The metalogical analysis of modality invokes the metalogical notions, like *satisfiability* or, perhaps, proof-theoretic consistency. This construal of modality is related to the *epistemic* notion of modality,

(1) it may be the case, *for all we know*, that ...

However, the only attributions of modality that make sense on this analysis are *de dicto* attributions of possibility and necessity to (the contents of) *closed sentences*. The standard notation for such the above "modalization" is,

(2) $\Diamond(\ldots)$

The analysis of this operator proceeds in the semantic metatheory, in assigning truth-conditions to modal formulas. For example, on this analysis, if φ is a closed formula (sentence), then,

(3) the formula $\Diamond\varphi$ is true just in case φ is *satisfiable.*

The reason proof-theoretic consistency is not the right analysis is that there are proof-theoretically consistent closed *second-order* formulas which are not satisfiable, and thus do not seem to correspond to any logical possibility. For example, let φ be $\mathsf{PA}^2 \wedge \neg\mathsf{Con}_{\mathsf{PA}}$. This formula $\mathsf{PA}^2 \wedge \neg\mathsf{Con}_{\mathsf{PA}}$ is consistent but unsatisfiable, because $\mathsf{Con}_{\mathsf{PA}}$ holds in any model of PA^2. It seems then that this formula is necessarily false, so $\Diamond\varphi$ is not true, whereas on the proof-theoretical analysis, $\Diamond\varphi$ would be true[162].

It is difficult to make sense of a metalogical analysis of an open formula like $\Diamond F(x)$. Model-theoretic satisfiability is a property of *closed* formulas. How might we deal with open formulas? A simple suggestion is that $\Diamond F(x)$ is *true of* an object *e* (in an interpretation $\mathfrak{I}$) just in case there is a term t designating (in $\mathfrak{I}$) *e* such that $\Diamond F(t)$ holds in $\mathfrak{I}$. So, $\Diamond F(x)$ is true in $\mathfrak{I}$ of an object *e* just in case there is a term such that $\mathfrak{I}[t] = e$ and $\mathfrak{I}$

162 However, there is growing literature (within modern mathematical logic) on provability interpretations of modal logic. Roughly, $\Box\varphi$ means 'it is provable that φ', where provability is usually taken to be provability in some background mathematical theory, such as PA or ZFC. For example, a modal formula like $\Box\varphi \rightarrow \varphi$ corresponds to a reflection sentence for PA, i.e., a formula $Prov_{\mathsf{PA}}(\ulcorner\varphi\urcorner) \rightarrow \varphi$. An excellent text on this interpretation of modality is Boolos 1993, *The Logic of Provability*.

$\models F(t)$. The problem with this is that objects do not have "canonical designators". There may be no term designating e, or there may be two terms t_1 and t_2 both designating e such that $\Diamond F(t_1)$ is true and $\Diamond F(t_2)$ is not true.[163]

A famous example (modified from examples given by Quine in Quine 1953c) is given by the open sentence '$x \neq 9$'. Clearly, '$\Diamond(9 \neq 9)$' is false but '$\Diamond$(the number of planets $\neq$ 9)' is true. And, of course, '9' and 'the number of planets' are terms which designate the same object, (i.e., the number 9). In Quine's terminology, modal contexts like '$\Diamond(x \neq 9)$' , may be *referentially opaque*. The usual resolution involves introducing possible worlds and saying that while '9' refers to 9 in all possible worlds, 'the number of planets' may refer to different numbers in different possible worlds: that is, the descriptive phrase 'the number of planets' is not, in Kripke's terminology, a "rigid designator"[164].

ii. The Metaphysical Conception of Modality (c.f., Tense)

The alternative analysis of modality is *metaphysical*. After decades of sometimes acrimonious dispute between the enemies of metaphysical modality (notably, Quine[165]) and its advocates[166], over the very intelligibility (or, perhaps, scientific respectability) of modality, it emerged that metaphysical modality is the prime notion at play when one wants to *quantify into* modal contexts, as, for example, in formulas like $\exists x \Box F(x)$. Such a notion of modality is sometimes called *de re* modality, for it involves the attribution of

[163] For further discussion see Burgess & Rosen 1997, pp. 139-140.

[164] See, e.g., Kripke 1980, for the classic discussion of how Venus, Nixon and Gödel might have been. See Forbes 1985 for further discussion.

[165] Since modal logic is a form of intensional logic, Quine's position might be called *extensionalism*. The reason modal logic is intensional is that the context $\Box(\ldots)$ is non-extensional. That is, one may have a pair of sentences φ_1 and φ_2 such that $\varphi_1 \leftrightarrow \varphi_2$ is true, but $\Box\varphi_1 \leftrightarrow \Box\varphi_1$ is not true. This is a sign of non-extensionality. Indeed, Quine's extensionalism is rather extreme: modal locutions are regarded as simply unintelligible.

[166] Famously, Ruth Barcan who first developed a formal system of quantified modal logic, and then, later, Saul Kripke, who along with others, developed a model-theoretic semantics and adequacy theorems for modal systems. Paul Benacerraf has a nice anecdote on these matters:

> ... was it in 1962 that Hilary Putnam, then at MIT, called me to tell of this young man who was making sense of what, to our everlasting shame, we used to call "muddle logic"?
>
> (Benacerraf 1996, p. 17)

necessary (or essential) properties to *particular things*. This is "essentialism". The properties of an object (irrespective of how described) are classified into the *essential* (holding in all possible worlds) and *accidental*.

For example, a formula like $\exists x \Box F(x)$ says that there is something x, such that the property expressed by F *holds necessarily or essentially of x*. An example cited by Putnam involves the statue and the lump of clay from which it is formed. According to the metaphysical analysis, the property of *being a statue* is a necessary or essential property of the statue, but is not a necessary property of the lump of clay. The lump of clay might not have had this property. In short, the lump of clay and the statue are *two different things*! They are discriminated by a modal property, expressed by "x might not have been a statue". The lump of clay is a statue at the actual world w^*, but there are accessible possible worlds w where the lump of clay exists where it isn't a statue. In contrast, there is no accessible possible world at which the statue exists at which it is not a statue. It is an essential property of the statue *itself* that it be a statue.[167]

Early appreciation of this metaphysical point lies at the heart of Quine's lifelong campaign to expose the scientific un-intelligibility of quantified modal logic. *De re* modality, with its concomitant essentialism, is to be contrasted with the less menacing so-called *de dicto* modality, associated with the metalogical analysis discussed above, which solely attributes necessary truth to closed propositions.

The standard model-theoretical analysis of *de re* metaphysical modal quantification theory introduces (in the metatheory) quantification over possible worlds. Although possible worlds themselves are not mathematicalia, the "Kripke models" built from such possible worlds are:

> The development of modal logic was greatly advanced with the introduction, by Saul Kripke and others, of mathematical models (now called Kripke models) of Leibniz's fantasy of the actual world as one "possible world" amongst others. In Kripkean semantics, sentences are true or false *at* various possible worlds, but,

[167] On the provability interpretation of modality, things are less problematic. For example, consider a modal extension of the language of PA by adding, for any formula φ, the formula $\Box\varphi$. Then stipulate that the modal formula $\Box F(x_1, \ldots, x_n)$ is true under an assignment which assigns the numbers n_i to variables x_i just in case the formula $F(n_1, \ldots, n_k)$ is provable in PA. E.g., $\exists x \Box F(x)$ is true if there is a number n such that PA $\vdash F(n)$. See Boolos 1993, p. xxxiv.

> typically, not all worlds are possible relative to, or "accessible from", others. A Kripke model is a triple (W, R, V), consisting of a domain W, the set of (possible) worlds, a binary relation R on W, the accessibility relation, and a relation V between worlds and sentence letters specifying which sentence letters are true at which worlds. The truth value of a truth-functional compound at a world is computed in the familiar manner from the truth values of its components at w. And a sentence $\Box A$ is true at w iff A is true at all worlds x such that wRx. (Thus the box acts like a universal quantifier over possible worlds). A sentence is valid in a model (W, R, V) iff it is true at all worlds in W.
>
> (Boolos 1993, p. xix).

On the standard philosophical conception, possible worlds are supposed to be *causally isolated* from the actual world. We lonely inhabitants of the actual world cannot influence events in other possible worlds. More precisely, there is no transmission of energy-momentum from one possible world to another[168].

A Kripke model for modal logic (W, R, V) is a highly static Platonic entity, built of a domain W of static possible worlds and their Platonic inhabitants, a static accessibility relation R and a static assignment function V. The same may be said of Kripke models for intuitionist logic, and of models for tense logic: temporal instants, and entities that *will* exist (in the future) and *did* exist (in the past), are treated as static "untensed" entities (such as Tony Blair-at-11.59pm-Dec-31-1999). In fact, a recent text on the metaphysics of modality provides a full *translation scheme* taking modal assertions into first-order logic, with the resulting quantified variables ranging over possible worlds and sets of entities existing at such worlds[169]. For example, a modal schema $\Box p \rightarrow p$ is translated as $\forall w p(w) \rightarrow p(w^*)$.

Finally, it is worth mentioning two significantly different ways of analysing so simple a *de re* modal assertion as,

(1) Quine might have been a chef

168 Science is strange. Modern quantum physics can be interpreted as a theory about a plurality of possible worlds. Indeed, several recent physicists have contemplated the construction of quantum-computation devices, which work by operating massively "in parallel", multitasking in countless parallel worlds. Perhaps this is nonsense. But if it isn't, it suggests that there is *some kind of interaction* between physical possibilities (e.g., they can interfere; they can be combined in superpositions). It is unclear how this relates to modality as studied by philosophers and logicians. It may be the case that both tense logic and modal logic are really parts of *physics*!

169 Forbes 1985, *The Metaphysics of Modality*, Appendix.

According to the view favoured by Kripke, (1) says, *of Quine himself*, that he might have been a chef. In other words, the modal property, "might have been a chef" applies to the actual Quine (who actually is not a chef). Modality is just of way of saying *new* things about ordinary entities. It is this introduction of new ideology or notions that permits us to modally discriminate objects which are identical in the actual world, such as the statue and the lump of clay. Kripke thus tends to think of possible worlds, rather Platonically, as "abstract possibilities".

In contrast, consider the highly metaphysical view of David Lewis, who is the arch champion of the reality of possible worlds. According to Lewis, (1) says of some object x in some accessible possible world w, first that x is a "*counterpart*" at w of the actual Quine, and second that x is a chef. This possible object, Quine-at-w is simply different from Quine-at-w^*. For the actual Quine, Quine-at-w^*, is not a chef and this "unactualized Quine", Quine-at-w, is a chef. Lewis's theory is called *Counterpart Theory*[170] and the semantics for such an analysis of modality is called *Counterpart-Theoretic Semantics*[171].

If Counterpart Theory sounds strange, there is an analogy with some analyses of tense and temporality. For example, a perfectly intelligible (and true) tensed assertion is,

(2) Quine was once taller than he is now

According to one analysis of temporality—an analysis which dovetails nicely with modern relativistic space-time physics—the Quine is a *four-dimensional entity* with *temporal parts*, or time-slices. A *time*, or a *temporal instant*, is a *spatial hypersurface* of the physical Universe. A *time-slice* of such an entity x at a time t (that is, the time-slice x-at-t) is just the mereological intersection of x with the spatial hypersurface whose time label is t. In particular, the "at" function is injective: so, if $t_1 \neq t_2$, then x-at-$t_1 \neq x$-at-t_2. Thus, the time-slice Quine-at-(3pm-June-3rd-1936) is different from the time-slice Quine-at-(3pm-June-3rd-1998). The former had no wrinkles, for example, and was taller. Indeed, the time that is "now" is exactly such a temporal instant (unlike normal singular

---

[170] See Lewis 1979.

[171] For a discussion of Counterpart Theory and many other important topics in the philosophy of modal logic, see Forbes 1985.

terms, "now" is an *indexical* singular term that designates different times at different times. At time t, 'now' designates t). Quine himself is the mereological aggregate of all these time-slices. Thus, (2) is analysed as,

(3) For some time t, before now, Quine-at-t is taller than Quine-at-now.

Lewis's idea then is that Quine is, in fact, a vast "transmundal" entity (c.f., a four-dimensional aggregate over time-slices, at different times) with disjoint "modal slices" in particular unactualized possible worlds. Quine himself is the transmundal aggregate of all these modal slices. Thus, on the counterpart analysis, (1) is analysed as,

(4) For some possible world w, accessible from the actual world w^*, Quine-at-w is a chef

8.3 Modern Modalism: Putnam 1967

The first recent philosopher of mathematics to seriously propose such a (counter-revolutionary) remodalization of mathematics was Hilary Putnam, whose 1967b proposes a new picture of mathematics, the "modal picture", which is contrasted with the "objects picture" (advocated by platonists). The sting in the tail is that Putnam goes on to claim that these pictures are "equivalent pictures"[172].

Putnam gave as an example Fermat's Last Theorem (FLT)[173]. FLT is the statement,

(1) There do not exist non-zero natural numbers n, x, y, z such that $n > 2$ and $x^n + y^n = z^n$

The negation of FLT is then, prima facie, an *existence* claim,

(2) There exist non-zero natural numbers n, x, y, z such that $n > 2$ and $x^n + y^n = z^n$

[172] Putnam was motivated by an analogy: the allegedly "equivalent" pictures in the interpretation of Quantum Mechanics: e.g., the Schrödinger picture versus the Heisenberg picture; the wave picture versus the particle picture.

[173] Or, perhaps, more correctly, the Fermat-Wiles Theorem, since Andrew Wiles of Princeton managed to prove it in 1993-1994.

Call this statement ¬FLT. Putnam argued that if ¬FLT is true, then it will be derivable from some finitely axiomatizable portion of elementary arithmetic. Indeed, had ¬FLT been true, it would have been derivable from a simple extension (the single axiom for) Robinson Arithmetic **Q**. The extension adds the recursive definition of exponentiation: i.e., the axioms $\forall x(exp(x, 0) = s(0))$ and $\forall x \forall y(exp(x, s(y)) = exp(x, y) \times x)$. All recursive functions are representable in (this extension of) **Q**. Addition and exponentiation are recursive. So, if $(n_1)^n + (n_2)^n = (n_3)^n$, then **Q** $\vdash exp(n_1, n) + exp(n_2, n) = exp(n_3, n)$. From this it follows that, if ¬FLT is true, then **Q** ⊢ ¬FLT. Actually, Putnam does not specify that it is this extension of **Q** that will serve this purpose. This is why I have tried to spell it out. Next, Putnam points out (in effect) that **Q** ⊢ ¬FLT just in case the conditional **Q** → ¬FLT is valid. Furthermore, **Q** → ¬FLT is valid just in case the modal statement □[**Q** → ¬FLT] is true. Next, Putnam asks us to imagine that the component non-logical symbols '*0*', '*s*', '+' and '×' and '*exp*' are replaced by purely schematic function symbols. Call the resulting uninterpreted schemes **Q*** and **FLT***. But the interpreted sentence □[**Q** → ¬FLT] is true just in case the modal formal schema □[**Q*** → ¬FLT*] is a valid theorem of modal logic. Let us write '$\vdash_{ML} \varphi$' to mean 'φ is a theorem of modal logic'. Thus, we have:

(3) if **Q** ⊢ ¬FLT, then $\vdash_{ML}$ □[**Q*** → ¬FLT*]

Thus, we have converted (or translated) a certain existential mathematical statement (that is, one which says that there are certain entities) into an uninterpreted modal schema. In the language of my earlier discussion of reconstrual theory, the (implicit) modal reconstrual function Γ_{mod} converts theoremhood to modal validity (theoremhood).

Now let us examine what Putnam says about the significance of the relationship between the mathematical statement and the associated schema. He says:

> Now the mathematical content of the assertion □[Q* → ¬FLT*] is *certainly the same* as the assertion that there exist numbers *x, y, z, n* ($2 < n$, $x, y, z \neq 0$) such that $x^n + y^n = z^n$. Even if the expressions are not synonymous, *the mathematical equivalence is so obvious* that they might as well be synonymous as far as the mathematician is concerned. Yet the pictures in the mind called up by these two ways of formulating what one might as well consider to be the same mathematical assertion are different. When one speaks of 'the existence of numbers' one get the picture of mathematics describing eternal objects; while (2)

> simply says that Q* entails ¬FLT*, no matter how one may interpret the predicate letters, and this scarcely seems to be about 'objects' at all.
>
> (Putnam 1967b (1996), p. 173).

In fact, this looks exactly like a version of if-thenism, which we argued against in Chapter 6. The truth value of FLT is being reduced to its theoremhood (derivability) in Q. Allegedly, when we say 'FLT is false' we just mean '¬FLT is a theorem of Q'. So, how does Putnam's position differ from if-thenism? Briefly, in addition to saying that there are modalized if-thenist truths of the form,

(4) necessarily, if (X, f) is an ω-sequence, then ...,

Putnam says that we must add "categorical" assertions like,

(5) ω-sequences are possible.

Thus,

> 'Numbers exist'; but all this comes to, for mathematics anyway, is that (1) ω-sequences are possible (mathematically speaking); and (2) there are necessary truths of the form 'if α is an ω-sequence, then ...' (whether any concrete examples of an ω-sequence exists or not).
>
> (Putnam 1967b (1996), p. 174).

Actually, Putnam does not clarify sufficiently what he means by a 'concrete example of an ω-sequence' (although he explains later what he means by a concrete example of a model of set theory).

In any case, what Putnam's modalism adds to naïve if-thenism is a *categorical component*. The modalist (if he or she is a nominalist) does not believe to be true a mathematical theorem φ as it stands, even if φ has been correctly derived from PA or ZFC. This would imply the existence of mathematicalia. Instead, he or she believes that,

i. the inference is correct, and

ii. the axioms from which φ was derived are consistent (possible)

Note that the modal categorical assertion is very weak. It asserts merely the *consistency* (or *possibility*) of M. The realist, in contrast, says that M is *true*.

In this respect, Putnam's modalism closely resembles formalism. Mathematical knowledge consists in knowledge—for various examples of M and φ—that φ is a logical consequence of some consistent axiom M. That is, in appropriate cases, the sort of mathematical knowledge that formalism says mathematicians have is of the form:

i. φ is derivable from M

ii. M is consistent

Putnam's point (I think) is that we may put all this *modally*. If φ is derivable from M, then $\Box(M \rightarrow \varphi)$ is true. If M is consistent then $\Diamond M$ is true. So, the modalist believes,

i°. $\Box(M \rightarrow \varphi)$

ii°. $\Diamond M$.

It might seem then that the modalist has discovered a deflationary way of analysing "mathematical practice". That is, identifying the sorts of propositions that mathematicians *really believe* (or: really that they *ought* to believe, if they were philosophically sensitive enough).

I mentioned the sting in the tail. Putnam claims that objects-platonism and modalism are "equivalent pictures" of mathematics:

> In my view the chief characteristic of mathematical propositions is the very wide variety of equivalent formulations that they possess. ...; what I mean is rather that in mathematics the number of ways of expressing what is in some sense the same fact (if the proposition is true) while apparently not talking of the same objects is striking.
>
> ... Reichenbach coined the happy term 'equivalent descriptions' for this situation [in QM].
>
> ... the two theories are, so to speak, on the same explanatory level. Any fact that can be explained by one can equally well be explained by the other. And in view of the systematic equivalence of statements in one theory with statements in the other theory, there is no longer any point to regarding the formulation of a given fact in terms of the notion of one theory as more fundamental (or even as significantly different from) the formulation of the fact in terms of the notions of the other theory.
>
> (Putnam 1967b (1996), pp. 170-171).

> My purpose is not to start a new school in the foundations of mathematics (say, 'modalism'). Even if in some contexts the modal logic picture is more helpful than the mathematical objects picture, in other contexts the reverse is true. ...

> Looking at things from the standpoint of many different 'equivalent descriptions', considering what is suggested by all the pictures is both a healthy antidote to foundationalism and of real heuristic value in the study of first-order scientific questions.
>
> (Putnam 1967b (1996), p. 181).

However, it is strikingly *unclear* that objects-platonism and modalism are in fact, as Putnam claims, "equivalent pictures" of the same mathematical facts. I will argue later that some simple modal translations (using just the modal $\Box$ operator) of mathematicized theories of Nature are not equivalent to their standard objects-platonism formulation.

I shall suspend criticism of Putnam's modalism and return to it later. Next we turn to another version of hermeneutic modalism in recent philosophy of mathematics.

8.4 Modal Constructibility Theory: Chihara 1990

Charles Chihara has adopted a quasi-nominalist strategy striking similar to the programme of the mediaeval nominalists of old. Talk of abstract mathematicalia (c.f., universals) is to be replaced by talk of *linguistic entities.* In Chapters 2 and 7 we mentioned the 1947 programme of Quine and Goodman, the attempt to develop a genuinely nominalistic *syntax*. Quine noted later that more could have been achieved by adding a component of modality, but commented, "... we would not for a moment have considered enlisting the aid of the modalities. *The cure would in our view have been worse than the disease.* (Quine 1986, p. 397. Emphasis added). Unlike extensionalists (like Quine), and other philosophers (like Putnam) who regard the introduction of a robust notion of modality as *anti*-nominalistic, Chihara does indeed propose enlisting the aid of the modalities. And, he also proposes invoking (again much like the mediaeval nominalists) a basic *semantical notion* (viz., satisfaction) with which to replace talk of the basic mathematical notion of membership.

Chihara's calls his modal system *Constructibility Theory* (CT). It is really a kind of notational variant of (Russellian) *Theory of Types*, replacing the membership symbol $\in$ by a satisfaction predicate, and replacing standard quantifiers by modal "constructibility quantifiers", *Cx*. The purpose of this "constructibility quantifier" is to replace, or to reinterpret, the usual existential quantifier $\exists x$ used in canonical formulations of arithmetic, analysis and set theory. So, the modal strategy involves reinterpreting standard existential assertions about mathematicalia as assertions of constructibility:

> The basic idea of the approach to be taken in this work is to develop a mathematical system in which the existential theorems of traditional mathematics have been replaced by constructibility theorems: where, in traditional mathematics, it is asserted that such and such exists, in this system it will be asserted that such and such can be constructed.
>
> (Chihara 1990, p. 25).

Chihara emphasizes that his constructibility quantifier should have no intuitionist or constructivist connotations, for that would severely limit the amount of classical mathematics reconstructible in his system:

> Now it is clear that I will need a more powerful notion of constructibility than that of the Intuitionists if I am to obtain anything like classical mathematics
>
> (Chihara 1990, p. 25).

> Furthermore, the discussion should make it clear that my constructibility quantifiers are very different from the quantifiers of Intuitionistic mathematics, according to which $\exists xFx$ is assertable just in case there is available a "presentation" of an object *b* and a proof of *F*(*b*). Thus, the assertion (*Cx*)*Fx* should not be taken as implying that one has (or even that one could in principle devise) an effective procedure for producing an *x* such that *Fx*.
>
> (Chihara 1990, p. 38).

Another reason for distinguishing between the notions of constructiblity in intuitionism and his system, is that Chihara certainly does not want to say (as some intuitionists do) that it is *mathematical objects*, like numbers, functions and sets, that are constructed.

Instead, Chihara's idea is to reconstrue mathematical existence assertions as asserting the constructibility of concrete linguistic items. In particular, the constructibility of *tokens* of open-sentences:

> What kinds of things are to be said to be constructible by the mathematical theorems in my system? In Intuitionistic mathematics, it is proofs or mental constructions that are asserted to be constructed or constructible. In my system, mathematics will be concerned with the constructiblity of *open-sentences*—indeed, *tokens* (as opposed to types) of open-sentences will be said to be constructible.
>
> (Chihara 1990, p. 40. Emphasis added).

Two further passages clarify the overall hermeneutic strategy, including the implicit reconstrual mapping:

> When the formal system is interpreted classically (so as to reflect the intended Fregean ontology), the predicate variables takes as values Fregean concepts and the existential quantifier is explained in the standard referential way. So, for each type of concept, we have a standard existential quantification over the totality of concepts of that type. All of these existential quantifiers will be replaced in my system by *constructiblity quantifiers* involving variables for *open-sentences* of the corresponding type. Essentially, then, a phrase of the form '*There is* a *concept F* such that' in the classical system gets replaced in my system with the corresponding phrase '*It is possible to construct* an *open-sentence F* such that'.
>
> (Chihara 1990, pp. 44-45. Emphasis added).

> Thus, corresponding to the assertion in the classical system that *there is* a monadic concept of level 1 which all objects fall under, there is an assertion in my system that *it is possible to construct a monadic open-sentence* of level 1 which all objects *satisfy*.
>
> (Chihara 1990, p. 45. Emphasis added).

In summary,

i. Mathematical statements about sets are to be systematically reconstrued as statements about *possible tokens of open sentences*;

ii. The basic set-theoretical relation $\in$ is then reinterpreted as *the satisfaction relation* between objects and such open sentence tokens.

The very simplest example would be the following theorem of type theory,

(1) there is a level-1 set z such that no level-0 object is an element of z

which is reconstrued as a theorem of the Constructibility Theory,

(2) it is possible to construct a level-1 open sentence token x such that no level-0 object satisfies x

Formalizing somewhat, we have, corresponding to (1), the set-theoretical theorem,

(1)° $\exists z^1 \forall x^0 (x^0 \notin z^1)$

which is reconstrued, corresponding to (2), as the modal constructibility theorem,

(2)° $C\alpha^1 \forall x^0 \neg Sat(\alpha^1, x^0)$

In fact, as Chihara notes, in some cases it is possible to reformulate the modal constructibility quantifier Cx using the more transparent modal quantifier $\Diamond\exists x$, meaning "it is possible that there is an x such that . . .". Thus, (2)° becomes,

(3) $\Diamond\exists\alpha^1\forall x^0\neg Sat(\alpha^1, x^0))$

However, Chihara says that this reconstrual can in other cases be misleading to the intended meaning of Cx. Chihara cites an objection due to Dale Gottlieb concerning this rendering of the constructibility quantifier:

> Let us classify any sentence consisting of the words 'There are no more than', followed by an Arabic numeral, which in turn is followed by the word 'star', as an *S*-sentence. For example, 'There are no more than 1,000,000 stars' is an *S*-sentence. Now consider the following sentence
>
> [1] Cx(x is an *S*-sentence and x is true).
>
> As I explicated such sentences, one can regard [1] as saying, informally,
>
> [2] It is possible to construct an *S*-sentence that is true.
>
> Gottlieb suggests that we can regard these sentences as asserting
>
> [3] $\Diamond\exists x$(x is an *S*-sentence and x is true)
>
> Now as [1] is intended, it is true just in case there are finitely many stars. But [3] is true even if, in fact, there are infinitely many stars. For no matter how many stars may in fact exist, it is possible that only finitely many stars exist. Strangely, what Gottlieb concludes from this is not that it is improper to replace '(Cx)' with '$\Diamond(\exists x)$', but rather that the constructibility interpretation of the existential quantifier that I gave does not "assign truth values in an acceptable manner". It should be clear to the reader, from what has been given above, that Gottlieb's interpretation of my constructibility quantifier, using the possibility operator, is a distortion.
>
> (Chihara 1990, pp. 37-38).

It has to be checked that every axiom of standard set theory is reinterpreted by the modal constructibility translation function Γ_{mc} as a theorem of **CT**. In fact, the axioms of **CT** are more or less a *notational variant* of the axioms of Russellian Type Theory. Indeed, as Chihara develops **CT** in Chapter 4 of his 1990, he introduces first the many-sorted (ω-sorted) notation (he calls it **Lt**) and then suitable axioms for formulas in **Lt**. These include The Axiom of Extensionality, Axioms of Identity, The Axiom of Abstraction (Comprehension) and a "Hypothesis of Infinity" (guaranteeing the existence of infinitely many level-0 ur-elements). We shall simply suppose that standard Type Theory translates under the implicit interpretation map Γ_{mc} properly into **CT**. For example, Shapiro 1997 points out that it is possible to translate the type-theoretic version of Cantor's Continuum Hypothesis into **CT**,

> In Chihara's system, there is a sentence equivalent to the following:
>
> > For every level-3 open sentence α, if α can be satisfied by uncountably many surrogate natural-number open sentences, then α can be satisfied by continuum many such open sentences.
>
> Such a sentence is obtained by translating a type-theoretic version of the continuum hypothesis into Chihara's language. The sentence is fully objective, and, of course, it is independent of the axioms of the system.
>
> (Shapiro 1997, p. 231).

Many questions now arise. Are **CT** theorems, like (2), meant to *have the same meaning* as their standard set-theoretical progenitors, like (1)? If so, then assertions about sets may be understood as assertions about the constructibility of open sentence tokens. But, conversely and symmetrically, assertions about the constructibility of certain open sentence tokens must be construed as assertions about sets. It looks as though this could provide no succour to the nominalist. The synonymy relation is symmetric.

The hermeneutic nominalist adopts the following hermeneutic principle:

> The literal, disquotational, truth conditions of mathematical assertions are misleading: their correct mathematical truth conditions are given by their nominalistic reconstruals.

As we noted in Chapter 6, a systematic hermeneutic translation mapping can be put to use in partially defining non-literal (non-disquotational) "mathematical truth" conditions, a theory **HT** of *hermeneutic truth*, analogous to the (incomplete) theory **DT** of disquotational truth. Let φ be some mathematical assertion about numbers or sets. Let $\Gamma_{mc}(\varphi)$ be its modal constructibility hermeneutical reconstrual. Then the hermeneutic nominalist may adopt:

(3) φ is *mathematically true* iff $\Gamma_{mc}(\varphi)$ is literally true.

For example,

(4) 'There is an empty set' is *mathematically true* iff 'it is possible to construct an open sentence token x such that nothing satisfies x' is literally true.

But literal truth includes disquotational truth, so,

(5) 'There is an empty set' is *mathematically true* iff it is possible to construct an open sentence token x such that nothing satisfies x

Statements such as (5) then form the axioms of HT, the theory of hermeneutic truth. Of course, we also have the *literal* (disquotational) truth condition:

(6) 'There is an empty set' is true iff there is an empty set

But a nominalist is someone who asserts 'There are no sets' and thus rejects 'There is an empty set' (indeed, any existential mathematical assertion φ). A coherent nominalist thus must therefore reject the truth claim ''There is an empty set' is true' (that is, 'φ is true'). However, the hermeneutic nominalist can accept 'φ is *mathematically* true' (that is, he or she accepts this just in case he or she accepts the reconstrual $\Gamma(\varphi)$).

Having briefly sketched Chihara's position, we shall subject it to criticism below.

8.5 Modal Structuralism: Hellman 1989

Modal structuralism was introduced by Geoffrey Hellman in 1989 and is an attempt to combine the modalism of Putnam 1967b with the structuralism in the philosophy of mathematics that goes back (at least) to famous discussions by Dedekind 1888 and Poincaré 1905. Hellman's position involves an explicit hermeneutic claim concerning how mathematical statements are to be reconstrued, and is advocated as a means of avoiding the "objects-platonism" discussed by Putnam[174].

---

[174] The main problems of so-called "objects-platonism" are the two conditions I shall call "Benacerrafitis". The problem which structuralism aims to resolve is the one articulated by Benacerraf 1965, the identification problem. If mathematicalia, like numbers and sets, are definite objects, then there ought to be *definite identities* between them. But numbers can be identified with sets in countless acceptable ways, and none seems priveleged. According to structuralism, it is just the structures that count. As long as we have the "natural number structure" or the "real number structure", then it simply doesn't matter what we identify the

One of the most pleasing aspects of structuralism is its very attractive (but still Platonist, I should say) account of application of mathematics (briefly introduced in Chapter 2 above):

i. The Structuralist Thesis

Theories in pure mathematics semantically characterize (and assert the existence of) certain types of abstract structure (e.g., progressions or ω-sequences, rings, modules, ordered fields, etc.);

ii. The Exemplification Thesis

Abstract structures are sometimes *exemplified* by physical systems in the physical world (e.g., time exemplifies the structure of the ordered continuum);

iii. The Application Thesis

Mathematics is applicable *because* it characterizes precisely those structures which we find exemplified in the physical world.

There are two components to Hellman's position:

A. Structuralism

Introduced to avoid reference to some particular intended model, with a particular domain of particular abstracta (numbers or sets);

B. Modalism

positions in the structure with. Arithmetic and analysis are branches of mathematics which try to figure out the facts about these structures. How such structures are exemplified, or identified with distinct but isomorphic structures, is unimportant. I lack the space to discuss the merits of structuralism. Suffice it to say, that while it seems to defuse the Benacerraf problem, it leaves the very notion of structure unexplained. The problem is that structures are treated with great success by model theorists as *sets*. So, it looks as if structuralism is just a way of advocating set-theoretical foundationalism, while refusing to worry about *which* set-theoretical structure is the "real" natural numbers. (In practice, set theorist just take ω, the set of finite ordinals). The other problem (if it is a problem) is that structures are abstract entities, and so the position is a version of Platonism. Shapiro 1997, for example, accepts all this: the structure-theoretic way of putting things is "on a par" with the set-theoretic way. (Incidentally, the further final step of saying that set theory is also the attempt to characterize a certain structure, the full cumulative hierarchy, is deeply unconvincing: this position presupposes what it wishes to eliminate).

> Introduced to avoid commitment to the actual existence of such structures as are characterized by the "structuralized" (second-order) axioms.

On the structuralist theme, (A), Hellman writes,

> Perhaps the central puzzle raised by traditional "objects-platonism" ... is the difficulty in seeing how it is that the posited abstract objects "play any role"—"make any difference"—in our knowledge and in our language. Prima facie, at least, it is difficult to understand, for example, how it is that we can justify construing mathematical reference as reference to particular abstracta, as opposed to other forming a structurally isomorphic system—or how, for that matter, such reference ever gets established in the first place.
>
> (Hellman 1989, pp. 3-4).
>
> It is a widely, if not universally accepted view, that, in the theory of arithmetic, what matters is structural relations among the items of an arbitrary progression, not the individual identity of those items. As one commonly says: "Any ω-sequence will do".
>
> (Hellman 1989, p. 11).

As for the modalism, Hellman writes,

> On the view we should like to articulate, *mathematics is the free exploration of structural possibilities, pursued by (more or less) rigorous deductive means.*
>
> (Hellman 1989, p. 6)
>
> In what follows, we seek to overcome these obstacles [lack of an explicitly worked-out structuralist interpretation; the difficulty of seeing in structuralism any genuine alternative to set theoretical objects-platonism] by combining insights of structuralism with a related, more recent strand of thought in philosophy of mathematics, roughly summed up as the view of mathematics "as modal logic". The locus classicus of this approach is Putnam's 'Mathematics Without Foundations', in which it is suggested that many of the problems plaguing objects-platonism (and, in particular, the identification of mathematics with set theory) could be overcome by reinterpreting mathematics, as standardly presented, in a modal language, in which a notion of mathematical or logical possibility is taken as primitive.
>
> (Hellman 1989, pp. 7-8).

Let us begin considering the structuralism. Hellman invokes a *second-order* formalization of the axioms of arithmetic. An advantage of a second-order formulation is that non-finitely axiomatized first-order theories (using schemes) become finitely axiomatizable. That is, PA^2 is a single axiom. The first-order axiom scheme of induction is replaced by a second-order axiom. PA^2 is the axiom:

$$[\forall x(s(x) \neq 0) \wedge \forall x \forall y(s(x) = s(y) \rightarrow x = y) \wedge \forall X(X(0) \wedge \forall x(X(x) \rightarrow X(s(x)))] \rightarrow \forall x X(x))$$

(N.B., + and × are explicitly definable in second-order PA^2. One converts the recursive axioms of PA governing + and × to direct definitions using second-order quantification over functions. Furthermore, second order logic contains a comprehension axiom scheme, $\exists X \forall x(X(x) \leftrightarrow A(x))$, where the second-order variable X is not free in the formula A).

Second-order arithmetic PA^2 is categorical. Any model of PA^2 is isomorphic to the countable intended interpretation $\mathfrak{N}$ of the language of arithmetic. So, the second advantage of going second-order is the possibility of categoricity. One may characterize a structure "up to isomorphism"[175].

Broadly speaking, structuralism is a hermeneutic strategy which attempts to provide a non-literal re-interpretation of mathematical assertions. The structuralist has to specify a translation mapping. A mathematical assertion φ is to be reconstrued (to a first approximation) as a statement which quantifies over structures, and says something like,

(*) for any structure Ω satisfying certain conditions, φ holds in Ω

There are two prima facie problems with this:

i. Collapse to If-Thenism

If the conditions are simply satisfying some axiom system M, then the reconstrual simply says:

(*) φ holds in every structure in which M holds

and this is just equivalent to:

(**) φ is a *logical consequence* of M

[175] A powerful defence of second-order logic for the foundational analysis of mathematics is Shapiro 1991. The categoricity of certain second-order axiomatizations of arithmetic and analysis is a major theme and adduced as a significant component in the role of formal systems as "implicit definitions" of structures. This Hilbertian theme is continued in Shapiro 1997 (except that Shapiro is no formalist. He thinks that *there are* what he calls "*ante rem*" structures, characterized by second-order arithmetic and analysis).

and this smacks strongly of if-thenism (although, for higher-order logic, the model-theoretical notion of logical consequence exceeds the proof-theoretical notion of derivability).

ii. The Mapping is Metalinguistic

The "first approximation" structuralist reconstrual (*) sketched above translates a mathematical statement φ about mathematicalia into a *metalinguistic* statement about φ itself and about φ "*holding in*" or "*being satisfied by*" some structure.

Both defects can be remedied. The first defect and its remedy are quickly summarized by Burgess & Rosen 1997:

> ... a proof of a theorem about the natural numbers:
>
> (o) For the natural order on the natural numbers, it is the case that ...
>
> typically will prove more than is stated, namely:
>
> (i) For any progressive order, it is the case that ...
>
> ... However, if there are no structures of the given type, it will be equally and vacuously true that:
>
> (i') For any progressive order, it is not the case that ...
>
> So structuralism is, at a second approximation, the view that theorems of mathematics like (o) should be construed as assertions like:
>
> (ii) For all progressive orders (*and there are some*), it is the case that . . .
>
> (Burgess & Rosen 1997, pp. 146-147).

So the remedy for the first defect is this: the structuralist must *add* to any hypothetical conditional of the form:

(Hyp_M) for any structure Ω satisfying condition M, φ holds in Ω

a *categorical assertion*:

(Cat_M) *there exists* a structure Ω in which M holds

For example, the following assumptions are required by the structuralist,

there exists a structure in which the axioms of PA hold

there exists a structure in which the axioms of ZFC hold

... and so on.

The remedy for the second-defect is more involved. An appropriate structure Ω for the language of arithmetic, say, must be (D, z, f, g, h), composed of a non-empty domain of entities D, a special object z in D, a one-place function f on D, and two further two-place functions g and h on D. Indeed, the structure $\mathfrak{N} = (\mathbf{N}, 0, s, +, \times)$ is the *intended* structure for the language of arithmetic and first-order arithmetic itself is the (non r.e.) class of first-order sentences that hold in $\mathfrak{N}$. Another way to put this is to say that the *interpreted* language of arithmetic is $(\mathsf{L}, \mathfrak{N})$.

The structuralist wants to say that if Ω is a structure of the "right sort", then φ is a mathematical truth just in case it holds in Ω. For example, the theorems of arithmetic are simply those statements which always hold in any *progression* (or *ω-sequence*)[176]. Note that if a structure Ω^* is *isomorphic* to Ω, then it follows that the theories of these structures (the respective sets of sentences that hold in each) are identical (that is, Ω^* and Ω are elementarily equivalent). Finally, the structuralist has to say that there are such things as progressions or ω-sequences.

In Chapter 1 of his book, Hellman develops the *modal structuralist translation scheme* for Peano Arithmetic and Analysis. As a starter, he writes,

> Intuitively, we should like to construe a (pure) number-theoretic statement as elliptical for a statement as to what would be the case in any structure of the appropriate type. In this case, the structures are, of course, "progressions" or "ω-sequences", so what we seek to make precise is a translation pattern that sends a statement of arithmetic S to a conditional such as,
>
> If X were any ω-sequence, S would hold in X. (1.1)
>
> (Hellman 1989, p. 16).

Notice that the first approximation (1.1) talks of a statement φ's "holding" in an ω-sequence. Hellman's way of developing a structuralist translation of an arithmetical

[176] In set-theoretical terminology, an *ω-sequence* or a *progression* is a structure (X, e, s) where:

(i) there is no $x \in X$ such that $s(x) = e$;

(ii) s is an injection from X to $X - \{e\}$; and

(iii) for any subset Y of X, if $e \in Y$ and, for any x in X, if $x \in Y$ then $s(x) \in Y$, then $Y = X$.

In the terminology of Drake & Singh 1996 (p. 81), ω-sequences are the same as "Peano Systems".

statement $\varphi(z, s)$ without ascending to metalinguistic talk *about* $\varphi(z, s)$ (and thus introducing the semantic relation of "holding in" or "satisfaction") is to introduce the apparatus of second-order logic in a way reminiscent of the "Ramsification" of a finitely axiomatized scientific theory:

i. replace the constant z by an individual variable x and replace the successor function symbol s by a second-order 1-place function variable, f. We thus obtain $\varphi(x, f)$.

ii. relativize quantifications of any individual variables within φ to a second-order variable X (e.g., any subformula $\forall x\psi$ of φ is replaced by $\forall x(X(x) \rightarrow \psi)$; any subformula $\exists x\psi$ is replaced by $\exists x(X(x) \wedge \psi)$. Call the result $\varphi^X(x, f)$

From now on let $PA(z, s)$ (without the ^{2} superscript) be *second-order* Peano Arithmetic. Then $PA^X(x, f)$ is the intermediate reconstrual of this axiom. Finally, for any $\varphi \in {}_0L$, we define the structuralist reconstrual mapping $\Gamma^{PA}{}_{stru}$ by,

$\Gamma^{PA}{}_{stru}(\varphi)$: $\forall X\forall x\forall f[PA^X(x, f) \rightarrow \varphi^X(x, f)]$

And we define the categorical assertion Cat_{PA} thus:

Cat_{PA}: $\exists X\exists x\exists f[PA^X(x, f)]$

Part of the importance of these reconstruals is explained thus:

i. for any $\varphi \in {}_0L$, $PA \models \varphi$ iff $\Gamma^{PA}{}_{stru}(\varphi)$ is necessarily true.

ii. Cat_{PA} is true iff PA holds in some structure

Now, from the structuralist perspective, the important thing about $\Gamma^{PA}{}_{stru}(\varphi)$ and Cat_{PA} is that they are both "logically pure", containing no *non-logical symbols*. In a sense,

i°. $\Gamma^{PA}{}_{stru}(\varphi)$ asserts that φ holds in every (full second-order) model of PA.

ii°. Cat_{PA} asserts that *there exists* a structure satisfying PA.

In particular, the Categorical Assertion simply asserts the existence of some PA satisfying structure, some ω-sequence (progression), without specifying a distinguished one (the genuine, natural numbers "themselves").

In consequence, the translation of an arithmetical statement, like,

(1) for any numbers x, y, $x + y = y + x$

becomes, roughly (using set theoretical language)

$(1)_{str}$ for any ω-sequence $(X, f, +, \times)$, if $e_1, e_2 \in X$, then $+(e_1, e_2) = +(e_2, e_1)$.

The next step is the *modalization* (described in pages 16-24 of Hellman 1989). This is triviality itself. For any $\varphi \in {}_0\mathrm{L}$, we define the modal structuralist reconstrual mapping Γ^{PA}_{msi} by,

$$\Gamma^{PA}_{msi}(\varphi): \quad \Box\forall X\forall x\forall f[PA^X(x, f) \rightarrow \varphi^X(x, f)]$$

And we define the modal categorical assertion Mod-Cat_{PA} thus:

$$\text{Mod-Cat}_{PA}: \quad \Diamond\exists X\exists x\exists f[PA^X(x, f)]$$

By analogy with the above, we have,

i°°. $\Gamma^{PA}_{msi}(\varphi)$ asserts that φ *would hold* in any (full second-order) model of PA;

ii°°. Mod-Cat_{PA} asserts that *there might have been* a structure satisfying PA.

In particular, the Modal Categorical Assertion now asserts the *possible existence* of some PA satisfying structure, some possible ω-sequence (or progression). Hellman seems to think that asserting merely the *possible existence* of a natural number structure will improve philosophical matters. As we shall see later, this is quite dubious.

This, effectively, is all there is modal structuralism. Generalizing, each mathematical theory is to be characterized by a certain second-order axiom M in a formalized language L. Ignoring some technicalities about relativizing first-order quantifiers, one performs the "Ramsification" and replaces the axiom M by the Modal Categorical Axiom

$\Diamond\exists X_1\exists X_2 \ldots M(X_1, X_2, \ldots)$. Second, any closed formula φ of L is translated as the quantified conditional, $\Box\forall X_1\forall X_2 \ldots [M(X_1, X_2, \ldots) \rightarrow \varphi(X_1, X_2, \ldots)]$.

We subject Hellman's modal structuralism[177] to criticism below.

8.6 Criticism I: Adequacy of the Modal Reconstrual

Having described the recent modalisms of Putnam, Chihara and Hellman, I now want to discuss some criticisms of such approaches. In our refutation of if-thenism in Chapter 6 above we noted a major difficulty in reconstruing (as the if-thenist should like) *mathematicized science*: those theories of Nature which contain pure and mixed mathematical axioms in their description of the Universe. Our example was Tim, whose main (contingent) hypothesis is that the set of temporal instants under the "before" relation is isomorphic to the set of reals under <. We pointed out that,

i. Tim is not *derivable* from pure mathematics alone (say, ZF^2, with ur-elements);

ii. $ZF^2 \rightarrow$ Tim lacks the empirical consequences of Tim.

Unless the modal logic used is extended in a certain way, some of the above strategies may suffer this same fate.

Consider first Putnam's modalism. Putnam does not explain how mathematicized assertions are to be reconstrued, so it is not clear how he would have proposed to reconstrue a theory like Tim modally. Let us try some possibilities (no pun intended). If it is reconstrued as the conjunction of,

i. $\Box(ZF^2 \rightarrow \text{Tim})$

ii. $\Diamond ZF^2$

then the same problems arise as before. First, the modal schema (i) is straightforwardly false, because not:-($ZF^2 \models$ Tim). The contingent theory of time Tim is not a logical

[177] Modal Structuralism is significantly reformulated (in particular, using mereology and Boolos' logic of plural quantifiers) in Chapter II.C of Burgess & Rosen 1997. This reformulation, which uses *actuality operator*, avoids an objection I sketch below.

consequence of ZF^2. Second, the sole addition beyond if-thenism is the modal claim (ii). But adding (i) and (ii) does not permit the derivation of non-mathematical empirical consequences of Tim. Suppose (simplifying enormously) that Tim $\vdash \varphi_O$, an observation statement. Then, at best, $[\Box(\mathsf{ZF}^2 \rightarrow \mathsf{Tim})] \wedge \Diamond\mathsf{ZF}^2] \vdash \Diamond\varphi_O$. And this is no good. For φ_O will be a *contingent* testable statement about the world (like "if the apparatus is set up *thus*, the counter will read *so-and-so*"), Consequently, the modalization $\Diamond\varphi_O$, "it is possible that φ_O", is just a modal *logical truth*, and nothing like a testable empirical consequence. In short, the reconstrual does not imply φ_O.

The same holds of Hellman's modalism, as I shall show in rather more detail. Let Γ_{msi} be the modal-structuralist reconstrual mapping and let T be a mathematicized scientific theory. Then, we shall see that,

$\Gamma_{msi}(T)$ simply does not imply the non-mathematical consequences of T.

That is, there are non-mathematical assertions φ such that $T \vdash \varphi$ but not:-($\Gamma_{msi}(T) \vdash \varphi$)[178]. Since any nominalistically-inspired reconstrual of a scientific theory is constrained by the requirement that it preserve non-mathematical consequences, it follows that $\Gamma_{msi}(T)$ is not an acceptable or adequate reconstrual of T.

On this topic of representing *applied* mathematical discourse within the modal structuralist interpretation, Hellman writes:

> To illustrate, let us consider a simple statement of numerical comparison, say "There are more spiders than apes (and a definite number of each)". Using the second-order formalism of Chapters 1 and 2, with our language expanded to include the relevant non-mathematical predicates (in this case, just 'spider' $S(x)$ and 'ape' $A(x)$), we can represent the statement by,
>
> $\Box\forall X\forall f(\omega(X, f) \rightarrow \exists\phi\exists\psi\exists n\exists m($"$\phi$ is a 1-1 correspondence between the class of all x such that $S(x)$ and the f-predecessors of n in X" & "ψ is a 1-1 correspondence between the class of all x such that $A(x)$ and the f-predecessors of m in X" & $m <_f n$)] (3.1)
>
> (Hellman 1989, pp. 98-99).

[178] Γ_{msi} will, of course, preserve implication: that is, if $\varphi_1 \vdash \varphi_2$ then $\Gamma_{msi}(\varphi_1) \vdash \Gamma_{msi}(\varphi)$. But what is required is different. The statement $\Gamma_{msi}(T)$, if it is to be a proper replacement for T, must still imply the *same non-mathematical assertions* as T, not their modal structuralist reconstruals. This problem could be remedied if an implication-preserving translation Γ_{msi} could be found such that, for any non-mathematical assertion φ, $\Gamma_{msi}(\varphi) = \varphi$.

It is easy to see what's wrong with this. We have our original statement,

(1) There are more spiders than apes (and a definite number of each)

which under a natural mathematical formalization becomes something like (or, at least mathematically equivalent to):

(2) There are bijections ϕ and ψ, and natural numbers n, m such that $\phi(\{x: S(x)\}) = \{x \in \omega: x < n\}$ and $\psi(\{x: A(x)\}) = \{x \in \omega: x < m\}$ and $n < m$

or, more formally,

(3) $\exists\phi\exists\psi\exists n\exists m(\phi \text{ is a bijection} \wedge \psi \text{ is bijection} \wedge \phi(\{x: S(x)\}) = \{x \in \omega: x < n\} \wedge \psi(\{x: A(x)\}) = \{x \in \omega: x < m\} \wedge n < m)$

Of course, (1) can be expressed more simply. I have chosen to represent (1) thus to remain faithful to Hellman's version[179]. Let us call this mathematicized statement A. To illustrate the objection, consider the simple fact that A (plus the usual axioms of set theory) logically implies the non-mathematical conditional:

(4) if there is exactly one spider, then there are no apes

or, if you prefer,

(4′) if $\exists x(S(x) \wedge \forall y(S(y) \rightarrow y = x)$ then $\neg\exists x A(x)$

Call this statement B. This statement is non-mathematical, quantifying only over spiders and apes. So,

$\mathsf{ZFC} \cup \{\mathsf{A}\} \vdash \mathsf{B}$.

That is, *B* is part of the non-mathematical content of the theory *A*. Now Hellman's modal structuralist reconstrual of A is this,

(5) $\Box\forall X\forall f[\omega(X, f) \rightarrow \exists\phi\exists\psi\exists n\exists m($"$\phi$ is a 1-1 correspondence between the class of all x such that $S(x)$ and the f-predecessors of n in X" & "ψ is a 1-1

[179] One can of course represent any such cardinality comparison statement 'there are more *F*s than *G*s' without reference to numbers. That is, just say that there is an *injection* from *G*s to *F*s and no injection from *F*s to *G*s.

correspondence between the class of all x such that $A(x)$ and the f-predecessors of n in X" & $m <_f n$)]

This, roughly, is a *modalization* of a *quantified conditional.* Call it A_{msi}. It is not hard to see that A_{msi} does *not* imply B.

In fact, it is sufficient to show that a slightly different statement A_{mod}:

(6) $\Box(PA^2 \to A)$

does not imply B, on the sole assumption that A implies B (which can be represented as the assumption that $\Box(A \to B)$ is true). The reason we can simplify is that the "Ramsification" within A_{msi} makes no difference. We need to show that the following set Σ of modal formulas is consistent:

Σ: $\Box(PA^2 \to A), \Box(A \to B), \neg B$

A simple propositional model of Σ is obtained by taking $PA^2 = p \wedge q$, $A = p \wedge q$, $B = p$ and $p = T$ and $q = F$. Clearly, $PA^2 \vdash A$, so $\Box(PA^2 \to A)$ is true; also, $A \vdash B$, so $\Box(A \to B)$ is true; and finally $\neg B$ is true. To summarize, $A \vdash B$ but not-($A_{msi} \vdash B$).

It might be thought that adding the modalized "categorical axiom", Mod-Cat_{PA}, would help:

(7) $\Diamond\exists X \exists z \exists f(PA(X, z, f))$

But this is utterly futile. For then we see that {A_{msi}, Mod-Cat_{PA}} implies only a *modalization* $\Diamond B$ and, again, {A_{msi}, Mod-Cat_{PA}} does not imply B itself. To see that {A_{msi}, Mod-Cat_{PA}} does *not* imply B, we simply need to show that the following set Σ' of formulas is consistent:

Σ': $\Box(PA \to A), \Box(A \to B), \Diamond PA, \neg B$

A propositional model of Σ' is obtained by taking $PA = p \wedge q$, $A = p \wedge q$, $B = q$ and $p = T$ and $q = F$. Clearly, $PA \vdash A$, so $\Box(PA \to A)$ is true; $A \vdash B$, so $\Box(A \to B)$ is true; also, not-($\vdash \neg(p \wedge q)$), so $\Diamond PA$ is true; and finally $\neg B$ is true.

In short, for any chosen arithmeticized mixed statement A, the modal structuralist reconstrual A_{msi} is simply incapable of delivering the non-mathematical consequences of A. From this it follows that neither A_{msi} alone, nor A_{msi} + Mod-Cat$_{PA}$, are adequate replacements for A.

The resolution of these problems was carefully explained in Burgess & Rosen 1997. It is also discussed in Field 1988. Technically, as Field explains, two things are required:

i. One must, in modalizing *mixed* mathematicized science, invoke an "*actuality operator*", @, whose purpose is to undo the modal effect of modal operators whose scope includes quantifications over non-mathematicalia.

Using this operator, a modal reconstrual of T can be obtained that implies all the primary consequences of T, as it should. Using @, define the actuality predicate $Act(x)$ by $\forall x(Act(x) \leftrightarrow @\exists y(y = x))$. Then, assuming that T is finitely axiomatized, the reconstrual T_{mod} is, roughly, $\Diamond(T \wedge \forall x(\neg Math(x) \rightarrow Act(x)))$. It is possible to show[180] that the main adequacy condition, that mathematics-free content be preserved, is satisfied. That is, for any primary assertion φ, if $T \vdash \varphi$, then $T_{mod} \vdash \varphi$.

But there is a second objection. Even with @, there will still be substantial non-mathematical content of T which is not preserved, as Field 1988 explains. Suppose that T implies a formula $\forall x F(x)$, where $F(x)$ is a *mixed* formula containing quantification over mathematicalia. This formula will not be implied by T_{mod} above. Instead, T_{mod} now implies the uninteresting consequence $\forall x \Diamond F(x)$.

For example, suppose that T implies that there are objects x and y such that $m(x)$ is π times $m(y)$. That is, $T \vdash \exists x \exists y(m(x) = \pi \times m(y))$. Then, $T_{mod} \vdash \exists x \exists y \Diamond(m(x) = \pi \times m(y))$, which is an uninteresting and trivial modal claim (namely, that there are objects whose mass *might have been* thus related, e.g., any two objects you care to mention!). Furthermore, T_{mod} does not imply the original consequence.

180 See Field 1988 (1989), pp. 258-259 (the proof is given in footnote 30).

If the modalization is to work properly, one must first find a *nominalistic formula* $F^{\circ}(x, y)$ in possibly an extension L^{+} of L such that, where T^{+} is an extension of T, $T^{+} \vdash \forall x \forall y(m(x) = \pi \times m(y) \leftrightarrow F^{\circ}(x, y))$. That is, this nominalistic formula $F^{\circ}(x, y)$ *re-expresses* nominalistically the content of the platonistic formula $m(x) = \pi \times m(y)$.

It seems clear that,

> ii. one must, *before applying the modalization procedure*, first show how to associate with each mathematicized statement a non-mathematical assertion which says the "same thing" about concreta. Given our mathematicized theory T in a two-sorted notation L, then, one must extend L to a new notation L^{+}, and add new axioms to form a new theory T^{+} such that, for each mixed formula $F(x_1, \ldots, x_n)$ of L with just free primary variables, there is a mathematics-free primary formula $F^{\circ}(x_1, \ldots, x_n)$ in L^{+} in the same variables such that $T^{+} \vdash \forall x_1 \ldots \forall x_n(F(x_1, \ldots, x_n) \leftrightarrow F^{\circ}(x_1, \ldots, x_n))$. In other words, one must show how to obtain an expressively conservative reduction of all mixed mathematical apparatus.

A simple but unsatisfying method to obtain this goal is simply to *add*, for every (!!) such mixed formula $F(x_1, \ldots, x_n)$ (with just primary variables free), a new *primitive primary formula* $F^{\circ}(x_1, \ldots, x_n)$ and an new axiom $\forall x_1 \ldots \forall x_n(F(x_1, \ldots, x_n) \leftrightarrow F^{\circ}(x_1, \ldots, x_n))$. This is just the Craigian reconstruction given by Burgess & Rosen 1997 and described in Chapter 7 above. It is not satisfactory. One must add *infinitely many new primitives* (e.g., one just adds a primitive meaning 'the mass of *x* is π times the mass of *y*', so on for all mathematically definable relations amongst primary entities).

The alternative is to look for a small set of nominalistic primitives and axioms governing them such that adding them to T in L yields a theory T^{+} in L^{+} which is an expressively and deductively conservative extension of T. For example, one might add, as nominalistic primitives, '*x* is massless', 'the mass of *x* is between that of *y* and *z*', 'the difference in mass of *x* and *y* is the same as the difference in mass of *z* and *w*'. This is Field's proposal, as set out in Field 1980.

Having done all this (*if* it can be done), one obtains a (non-modal) mathematicized theory T^+ in L^+ which is a merely redundant extension of (i.e., is deductively and expressively conservative over) the original T in L, and which is deductively and expressively conservative over its primary restriction. At this stage, one may drop all mixed and secondary axioms of T^+, pertaining to mathematicalia. The result $(T^+)^\circ$ in $(L^+)^\circ$ is a nominalistic reconstruction of T.

Field 1988 stresses that if such a full nominalization can be achieved, there is then simply no point in *modalizing* the remaining mathematical components of T^+ in L^+. For one may then simply *drop* all the mathematical axioms of T^+ (and regard these axioms as falsehoods, "convenient fictions").

To conclude, if Field is right, the programme of modalization is *irrelevant* to the nominalization of *applied* mathematicized theories of Nature.[181]

8.7 Criticism II: Modal Primitivism

The main modal strategies we have looked at are:

i. Simple Modalism (Putnam),

ii. Modal Constructibility Theory (Chihara),

iii. Modal Structuralism (Hellman).

But how are we to understand the modality as invoked within these strategies?

I suspect that Putnam and Hellman would prefer some kind of metalogical analysis, wherein $\Diamond$ means "it is *logically consistent* that ...". Even this is unclear. To explain $\Diamond\varphi$ metalogically one asserts that $\Diamond\varphi$ is true just in case *there is a structure* in which φ holds. so, Platonic mathematicalia reappear in the metatheory. But we have seen that a simple metalogical *de dicto* conception is inadequate for the modalization of mathematicized theories of Nature. Neither $\Box(ZF^2 \rightarrow Tim)$, nor $\Diamond Tim$, nor $\Diamond ZF^2$ (nor a combination) will

[181] N.B. Field does however think that (meta-)logic is to be nominalized via modalization. See next Chapter.

do as a nominalistic replacement for Tim. At the very least, one must introduce an the actuality operator @ or predicate *Act*(*x*). This apparatus is equivalent to invoking a distinction between entities that are actual and those are merely potential or possible. The values of Chihara's constructibility quantifier *Cx* are such non-actual possibilia. Thus, we are driven to a metaphysical *de re* conception of modality. So, let us concentrate on Chihara's invocation of *de re* modality, via his constructibility quantifier *Cx*.

There is strong evidence that Chihara actually intends his use of modality to be understood *metaphysically*. Indeed, Chihara suggests a *possible worlds* reading of his constructibility quantifiers:

> The constructibility quantifier (*Cx*) is to be understood as saying, roughly,
>
> It is possible to construct an *x* such that
>
> but in many contexts, it could be rendered
>
> It is possible that there be an *x* such that
>
> depending upon what sort of *x* we are talking about. In the beginning, at least, the second rendition may appear to be somewhat more natural. Thus, in terms of the familiar Kripkean possible worlds semantics,
>
> –(*Cx*)(*x* is a tangram in the shape of a circle)
>
> can be understood as saying (but only as a first approximation)
>
> There is no possible world in which a tangram in the shape of a circle exists
>
> (Chihara 1990, pp. 25-26).

Note only does Chihara introduce such informal renderings, he introduces in Chapter 2 ("The Constructibility Quantifiers") and Chapter 4 ("The Deductive System") of his 1990 a wholesale Kripkean possible worlds semantics:

> A Kt-interpretation is an ordered quadruple (W, a, U , I) in which:
>
> W is a non-empty set—the set of possible worlds;
>
> a is a member of W—the actual world
>
> U is a function that assigns a non-empty set to each ordered pair (n, w), where n is an Arabic numeral of a non-negative integer, and w is a member of W; U(n, w) can be regarded as the set of things of level n that exist in w;
>
> I is a function that assigns:
>
> (a) to each individual constant of level 0, some element of U(0, a);

> (b) to the S-predicate of level 0, some subset of the Cartesian product of U(0, a) with Z[1] = the union of all the U(1, w), where $w \in W$;
>
> (c) to the I-predicate of level 0, some subset of the Cartesian product of U(0, a) with itself;
>
> (d) to each individual constant of level n (n > 0), some element of Z[n];
>
> (e) to the S-predicate of level n (n > 0), some subset of the Cartesian product $Z[n] \times Z[n + 1]$;
>
> (f) to the I-predicate of level n (n > 0), some subset of $Z[n] \times Z[n]$.
>
> (Chihara 1990, p. 57).

Actually, this seems to me like a major own goal for Chihara's strategy. It really is an intellectual cartwheel. He explains *only too well* the meaning of his "nominalistic" system by using a standard *mathematical explanation*, talking of sets, functions and membership, and of such things as possible worlds. If I were to *explain* to someone why iron rusts, say, invoking the valence properties of atoms and molecules, it would be nothing more than transparent sophistry to add, "By the way, atoms and the like don't exist!".

Chihara pours some undeserved scorn on the standard (Platonic, static) possible worlds model theory for metaphysical modality.

> It should be emphasized again that the above appeal to possible worlds was made to relate the constructibility quantifiers to familiar and heavily studied areas of semantical research. I, personally, do not take possible worlds semantics to be much more than a device to facilitate modal reasoning. Still, I hope to convince most philosophers by means of such analyses that the predicate calculus I shall be using is at least consistent and that it has a kind of coherence and intelligibility that warrants the study of such systems.
>
> (Chihara 1990, p. 38).

> Remember that, for me, this whole possible world structure is an elaborate myth, useful for clarifying and explaining modal notions, but a myth just the same. It would be a mistake to take this myth too seriously and imagine that we are exploring real worlds, finding there real open-sentence tokens that have puzzling features.
>
> (Chihara 1990, p. 60).

But this raises a major problem. For Chihara, metaphysical modality is to be understood as a *primitive notion*, not to be explicated using Kripkean model-theoretical semantics. This kind of stance is sometimes known as *modal primitivism* and is problematic:

> On one side, the modal nominalist must defend *intensionalism*, or acceptance of modal logical notions, against *extensionalism*, according to which modal notions should be avoided as obscure and confused. On the other side, modal nominalism must defend *primitivism*, acceptance of modal logical distinctions as undefined, against *reductivism*, which would reconstrue modal logical notions in terms of an apparatus of unactualized possibilia.
>
> (Burgess & Rosen 1997, p. 124).

The problem is that the standard (mathematical) possible worlds analysis *does in fact provide* an *explanation* for the validity and invalidity of "informally valid and invalid" modal arguments. We seem to have a conception of informal validity and invalidity of modal arguments (and of truth and falsity for modal assertions). The possible worlds analysis *clarifies* and *explains* our intuitions. Noone would learns modal logic can fail but to be *impressed* by the fact that, when modal statements are analysed as implicit quantifications, modal validity transforms into quantificational validity (roughly, "it is necessary that ..." means "in every possible world w, ...")[182]. Chihara, in arguing that modality be taken as basic (primitive and indefinable) is simply *repudiating* a powerful standard *explanation* (or at least, a powerful set of competing explanations). And not only does he purport to repudiate the explanation. He is prepared to *use* it to clarify what Cx means!

On this topic of primitivism, consider for a doctrine that might be called *Ontic Primitivism.* Namely, that the *concept of existence*, expressed using 'there is ...' or $\exists$, is ultimately primitive and indefinable. Of course, we can "explain" existence by taking other notions as primitive. Obviously, we may take $\forall$ and $\neg$ as primitive. Alternatively, within arithmetic, we may take 'the number of ...' and 'zero' as primitive and define $\exists x\varphi$ thus:

$$\exists x\varphi \leftrightarrow \#\varphi \neq 0$$

Or, within set theory, we may take 'the set of ...' and 'empty set' as primitives and define $\exists x\varphi$ thus:

$$\exists x\varphi \leftrightarrow Ext(\varphi) \neq \varnothing$$

182 See Forbes 1985.

But there are compelling reasons for regarding such "explanations" of existence as simply being the wrong way round. Rather, the number *0* is the number of non-self-identical things, *because there do not exist* any such non-self-identical things; and the empty set ∅ is the set of such non-self-identical things, again because *there do not exist* any such things. If this is correct, then the concept of existence *is* primitive and indefinable in simpler terms. As Quine put it,

> Existence is what existential quantification expresses. There are things of kind *F* if and only if $(\exists x)Fx$. This is as unhelpful as it is undebatable, since it is how one explains the symbolic notation of quantification to begin with. The fact is that it is unreasonable to ask for an explication of existence in simpler terms.
>
> (Quine 1969c, p. 97).

Now turn to Modal Primitivism. The modality expressed by such operators as □ and @ is to be taken as primitive and undefinable, and not reducible to other notions (e.g., to existence simpliciter, and notions expressed by 'w is the actual world', 'w_1 is accessible from w_2', and so on). Might something like this be argued in the modal case?

Consider an application of the possible worlds analysis of modal reasoning. We have the (metaphysical) analysis of,

(1) Quine might have been a chef

via,

(2) There is an accessible possible world at which Quine is a chef

Now can we argue that (1) is true *because* (2) is true?

Consider a modal argument like,

(3) If Quine had not been at Harvard, he wouldn't have met B.F. Skinner

Quine might not have been at Harvard

Quine might not have met B.F. Skinner

This can be analysed (for validity) using possible worlds:

(4) For any possible world w, if Quine is not at Harvard at w, then Quine doesn't meet B.F. Skinner at w

For some possible world w', Quine is not at Harvard at w'

For some possible world w', Quine does not meet B.F. Skinner at w'.

This argument has the form:

(5) $\forall w(F(w) \rightarrow G(w)), \exists wF(w)$; therefore $\exists wG(w)$.

Now the important thing is that this argument is *quantificationally valid* in first-order logic. It is via the quantificational validity of (5) that we explain the modal validity of (3). In short, we have reduced the explanation of *modal validity* to standard quantificational validity, where the quantification ranges over possible worlds (This is one the standard examples of trading "ideology" for "ontology").

One might argue that, irrespective of our scepticism about such peculiar metaphysical entities as "possible worlds", the possible worlds explanation of the validity of modal arguments *explains* their informal validity, just as the quantificational analysis of Aristotelian syllogisms *explains* their validity, and the analysis of tensed arguments as involving quantificational over temporal instants *explains* their validity.

> So far, I have described the situation between realism and fictionalism [modalism] as a balanced stand-off, but there is an important asymmetry. The fictionalist proposes that we rejects the realist's system—regarding it as no more than a work of fiction—whereas the realist accommodates the modal language, via the model-theoretic explications. The important point is that once the model-theoretic explication is in place, the realist has a lot to say about logical possibility and logical consequence. It is a gross understatement to point out that mathematical logic has been a productive enterprise. The challenge to the fictionalist is to show how she can use the results of model theory, as they bear on the *primitive* modal notion.
>
> (Shapiro 1997, p. 227).

> ... both Chihara and Hellman occasionally invoke a possible-worlds semantics, but they regard it as a heuristic, not to be taken literally. Neither of them believes that possible worlds exist. The role of the semantics is to help the reader grasp the intended logic of the formulas and to see what does and what does not follow from what. ... If the structure [of possible worlds] is just a myth, then I do not see how it explains anything. One cannot, for example, cite a story about Zeus to explain a perplexing feature of the natural world, such as the weather. ...

> Intuitively, to explain something is to give a reason for it or, according to Webster's *New Twentieth Century Unabridged Dictionary*, to clear from obscurity and make intelligible. In everyday life, a purported explanation must be true, or approximately true, in order to successfully explain. I take it that frictionless surfaces and the like, are parts of respectable scientific explanations of physical phenomena because they approximate actual physical objects. If they did not, then it is hard to see any explaining. It is not clear what, if anything, possible worlds approximate vis-à-vis the modal notions at hand—if not the possibilities themselves or the structure of the possibilities. In any case, I propose that the burden is on Chihara to tell us more about the modal notions and more about explanation before we can see how possible worlds can clear the modal notions from obscurity and make them intelligible.
>
> ... the fact that a myth of possible worlds happens to produce the correct modal logic is itself a phenomenon in need of explanation. That is, from the antirealist perspective, the success of possible worlds adds to the philosophical puzzle.
>
> (Shapiro 1997, pp. 232-233).

If the above comments about recent technical work on modal logic and tense logic are correct, then there is a relatively clear sense in which modality and tense have been, in some sense, partially reduced to, or analysed in terms of, demodalized and untensed Platonic mathematics. That is, to ordinary objectual quantification over possible worlds and temporal instants. In other words, those who claim that modality is "basic", or that tense is "basic", have got things the wrong way round. Rather, it is the structure of the set of possible worlds that is basic, and the structure of the set of temporal instants that is basic. Modal and tensed locutions within natural language are, on this view, highly derivative and non-basic[183].

8.8 Criticism III: Possible Existence = Existence

Putnam in his 1967b suggests that assertions like,

(1) there is an integer which is between 10 and 12 and is prime

[183] Can modal arguments be analysed using the metalogical conception of modality? This was a problem discussed long ago by Quine. See Quine 1953d, 'Three Grades of Modal Involvement'. The metalogical or *de dicto* conception is grade 1 (roughly, $\Box$ is attachable only to *closed* formulas). Quine argued that quantification *into* modal contexts requires a more involved, *de re* modality. But much ordinary modal discourse, when regimented, ends up with quantifiers binding variables *inside* modal contexts. (Indeed, this is connected to the analysis of *scope distinctions*, which explains how statements like 'I thought you were taller than you are' can make sense). I lack the space to discuss this controversial matter.

(2) there is a homomorphism from SU(2) to SO(3)

(3) there is a fixed point sentence G_T such that $T \vdash G_T \leftrightarrow \neg Prov_T(\ulcorner G_T \urcorner)$

are *equivalent* to saying things like,

$(1)^\Diamond$ it is possible to select an integer between 10 and 12 which is prime

$(2)^\Diamond$ it is possible to find a homomorphism from SU(2) to SO(3)

$(3)^\Diamond$ it is possible to construct a fixed point sentence G_T such that $T \vdash G_T \leftrightarrow \neg Prov_T(\ulcorner G_T \urcorner)$

I agree (although the demodalized versions (1)-(3) are preferable[184]). The problem is that it can be argued the application of metaphysical *de re* possibility and necessity to *abstract entities* is pointless. Abstracta (if there are any) have the properties they do by *necessity*. So, if it is possible that there is an *x* such that ..., then *there is* such an *x*, and, indeed, it is necessarily the case that there is such an *x*. In other words, "modalities collapse" upon application to purely mathematical assertions. If φ is a mathematical assertion, and the modal operators express metaphysical modality then $\varphi \leftrightarrow \Diamond\varphi$ and $\varphi \leftrightarrow \Box\varphi$ are both true, and indeed, both necessarily true.

> ... why was it that the possible existence of concrete numerals rather than that of abstract numbers was assumed? The reason is perhaps mainly that the idea of a distinction between actual and possible existence makes questionable sense in application to pure mathematicalia like numbers and sets of numbers ...; and whether or not it makes sense, the assumption that mathematical entities could perfectly well have existed and just happen not to is one few nominalists have found attractive. Indeed, even the appeal to the possible existence of new sorts of entities that are concrete in that they causally interact with one another, but that do not causally interact with actually existing sorts of entities, including human beings, might be though nominalistically repellent.
>
> (Burgess & Rosen 1997, p. 140).

This is why if one wants to re-develop mathematics as "modal logic", one must either

i. Invoke the metalogical *de dicto* construal of modality in application to closed and purely mathematical sentences. The function of such modality

184 However, the *proof* of the Fixed Point Theorem actually shows how to construct a fixed point sentence G given any formula $A(x)$. See Boolos & Jeffrey 1989, p. 173.

is to provide not a reconstrual of mathematical assertion *per se*, but of *metalogical assertions*. For example, one can re-express the assertion that φ is a consequence of M by saying that $\Box(\mathsf{M} \rightarrow \varphi)$ is true; and one can re-express the logical consistency of M by saying that $\Diamond\mathsf{M}$ is true;

or,

ii. One must invoke the metaphysical *de re* version of modality, reconstruing mathematical assertions about numbers and sets as modal assertions about non-abstract non-mathematicalia, such as numeral tokens or open sentence tokens (or whatever).

In other words, one must be very careful if one wants to say (as Putnam did) that,

(4) ω-sequences are possible

This may be taken to mean that,

(5) it is possible to find an *abstract structure* (X, f), consisting of a set X and a function f which is an injection from X to $X - \{e\}$, for some $e \in X$, and which satisfies the induction principle (so that, for any subset Y of X, if $e \in Y$ and if $f(x) \in Y$ whenever $x \in Y$, then $Y = X$).

Of course, such a structure (X, f), an ω-sequence, is an abstract set-theoretical entity. By our comments above, if it is possible to find such a structure, then *it exists*. Period. Such a construal of (4) is of no use to nominalists.

However, (4) may mean merely that,

(5) there could have been a *concrete exemplification of an ω-sequence.*

Namely, a *concrete entity* with denumerably many parts, which are related in a certain (physically definable) way. Such a construal of (4) is the one that appeals to nominalism. Talk about the existence of abstracta is to be re-interpreted as talk about the possible existence of certain kinds of (possibly unactualized) concreta.

Thus, it seems that the sort of modal *de re* reconstrual advocated by Chihara is what modal nominalists should be seeking. Chihara's open sentence tokens and his numerals function as concrete surrogates for bona fide mathematicalia, and *de re* modality makes sense in application to these (unless one is an extreme extensionalist, like Quine). If it is possible to construct a numeral or an open sentence token of such-and-such a kind, then such a numeral or open sentence token need not actually exist.

Thus, there are really two modal strategies at work:

A. The Modalization of Metalogic

The purpose of this strategy is the "modal nominalization" of metalogic. According to this, all metalogical assertions about consistency, derivability, satisfiability and so on, are to be reconstrued as *de dicto* modal assertions. This, in fact, is a component of Field's deflationist programme discussed in the next Chapter.

B. The Modalization of Mathematics and Mathematicized Science

The purpose of this strategy is the "modal nominalization" of mathematical assertions, including the mixed assertions used within science (scientific laws like Tim, Maxwell's Laws, Schrödinger's Law and so on).

We have seen that there are prima facie problems with programme (B) for the sorts of reconstrual proposed by Putnam and Hellman. A proper modalization of mathematics, rather than metalogic, will be one that can properly reconstrue mathematicized laws of Nature. It seems to me that the best prospect for such a reconstrual—that is, a reconstrual that doesn't face the *prima facie* adequacy problems as explained above—is a Chihara-style modalization using constructibility quantifiers, or some appropriate modification of modal logic.

8.9 Criticism IV: The Epistemology of Modality

The final criticism is perhaps the strongest. What is the main attraction of modalism within mathematics? Presumably, that it provides a quasi-nominalistic resolution for epistemic *Benacerrafitis*, the problem of "access". Apparently, because abstract mathematicalia are causally isolated from the mechanisms of sensory input, it follows that knowledge of, or epistemic access to, mathematicalia like numbers or sets is impossible. Apparently, so the argument goes, *modal knowledge* of "what might have been", "what it is possible to construct", and so on, is nevertheless possible *on the very same causal empiricist assumptions about sensation.* I simply disagree.

It is a standard assumption that *possibilia* are causally isolated, just as mathematicalia are. Indeed, if the Quinian remedy for epistemic *Benacerrafitis* briefly outlined in Chapter 6 is correct[185], then there is a very good holistic epistemology for mathematicalia: abstract entities are indispensable posits of our best high-level explanatory theories of the Universe, and we have at least as much evidence for mathematicalia as we have for neutrinos[186]. Perhaps, one might extend (*pace* Quine) this kind of epistemology to *possible worlds*[187]. When ontological matters are made explicit (by translation into first-order logic) such mysterious entities as mathematicalia, possibilia and temporal instants are recognized as values of variables, posits, of our over-all theory of the world.

In contrast, there is no such thing as a standard naturalistic, or empiricist, or causal, epistemology for modality. Perhaps the easiest epistemology would be for the metalogical *de dicto* conception of modality, as standardly explained using mathematical

[185] See also Hart 1979.

[186] Quine's argument that evidence for the existence of mathematical entities is of the *same kind* as evidence for neutrinos horrifies some authors (e.g., Parsons 1983b).

[187] Indeed, Quine does just this for temporal instants, proposing that tensed statements be regimented via quantification over times, rather than using Arthur Prior's primitive tense operators. To be sure, Quine's argument for this preference is based on the fact that modern physics already speaks of temporal instants (spatial hypersurfaces). If possible worlds turn up in *serious physics*, then the objection that they are mere "metaphysical fictions" would be significantly undercut. Now, as I noted earlier, some interpretations of Quantum Mechanics, and even some serious physicists, do take the idea of possible worlds seriously.

notions, like consistency, satisfiability or provability. It follows that an epistemology for *de dicto* modality might ride "piggy back" on a suitable epistemology for mathematics. If one thinks that mathematical knowledge is impossible, we have a problem even explaining knowledge of *de dicto* modal assertions, despite the fact that we are clearly capable of using and understanding modal assertions understood in this way.

Furthermore, there is no standard naturalistic, or empiricist, or causal, epistemology for metaphysical *de re* modality. How can we know, for example, that Quine mightn't have gone to Harvard? That is, how can we know that *there exists an accessible possible world* where Quine didn't go to Harvard? The obvious and natural question is a perfect analogue of Benacerraf's question about mathematical knowledge: *how is knowledge of metaphysical modality possible.* Since this is a copy of the problem for mathematics, let us call this worry *Benacerrafitis*$_{\text{Modal}}$:

Benacerrafitis$_{\text{Modal}}$: Access to Modal Knowledge

How can we know modal truths? Possible worlds and objects in them are not spatio-temporally related to the actual world; possible objects, like Lewis-ian modal counterparts, do not cause anything in the actual world; they do not influence our sense organs and we cannot point to any of them or perceive any of them. Our presumptions about what possibilia *are* seems to rule out any means by which we could *know* about them.

It seems to me that this problem is at least as great as epistemic *Benacerrafitis* is for ordinary mathematical knowledge. Unless Chihara and other modalists can give any reason for thinking that modal knowledge (*de re* modality, metaphysically construed) is more tractable than mathematical knowledge, then the main motivation for a modal reconstrual of mathematics collapses:

> The fact that there are such smooth and straightforward transformations between the ontologically rich language of the realist and the supposedly austere language of the fictionalist indicates that neither of them can claim a

> major epistemological advantage over the other. There is a positive side to the equivalence and a negative side. ...
>
> My contention is that with the translations, the major philosophical problems with realism get "translated" as well. ... There is, after all, no acclaimed epistemology for *either* language. In short, it is hard to see how adding primitive possibility operators to the formation of epistemic problems can make them any more tractable, and consequently, it is hard to see how the fictionalist has made any progress over the realist on the sticky epistemic problems.
>
> (Shapiro 1997, pp. 225-226).

Finally, we might even suggest that an epistemology for mathematical knowledge is likely to be *more* tractable than one for *de re* modal knowledge. This is Quine's current position. Mathematical knowledge is tractable on Quine's view via his *scientific pragmatic holism*: extensional (non-modal) mathematics is fully integrated within and indispensable from our overall scientific theory. Modal assertions are, on Quine's view, highly derivative and non-basic. It certainly is not legitimate to introduce *primitive* modality as a foundation for mathematics, and to claim that this is innocent of problematic consequences.

CHAPTER 9

Deflationism About Mathematics

Let us see how, or to what degree, natural science may be rendered independent of platonistic mathematics.

W.V. Quine 1948 (1980), 'On What There Is', p. 19.

... any inference from nominalistic premises to a nominalistic conclusion that can be made with the help of mathematics could be made (usually more long-windedly) without it.

Hartry Field 1980, *Science Without Numbers*, p. x.

9.1 Field's Programme

Along with modalism, the second major strategy for eliminating reference to abstract mathematicalia is due to Field, first presented in his 1980, *Science Without Numbers* and further developed in a series of papers anthologized in his 1989, *Realism, Mathematics and Modality*. Let us refer to this programme as *revolutionary deflationism* or just *deflationism*. There can be little doubt that Field's programme is by far the most carefully developed, and widely discussed of the various nominalist strategies. Partly as a result of this careful precision in Field's own development of his programme, and partly because of the detailed criticism this programme has received, the basic problems with the programme have been made very clear.

Field's programme may be summarized as follows. The whole of mathematicized science (including physics and logic) is to be disbelieved and renounced. One is to adopt a sceptical or fictionalist stance towards standard mathematics and standard mathematicized assertions of science. However, this destructive attitude must be complemented by a *positive, constructive* programme. One must show how mathematics is dispensable in applications and, indeed, one must provide workable *replacement theories* which are genuinely consistent with the renunciation of abstract mathematical entities.

i. Fictionalism About Mathematicalia:

Field calls his position "fictionalism". Mathematical theories, like arithmetic, analysis and set theory, are merely "good stories" which have certain properties that make them useful in science. Correspondingly, mathematicalia are "convenient fictions"[188]. According to Field, the primary philosophical *motivations* for this renunciation of mathematicalia[189] are the problems for platonism explained in Benacerraf 1965 and Benacerraf 1973 (the problems of identification[190, 191] and "epistemic access"[192]).

ii. Acceptance of Quine's "Science Entails Platonism" Argument

Despite advocating nominalism, Field *accepts* that standard mathematicized science *is* platonistic. On standard assumptions about truth, such mathematicized theories are ontologically committed to mathematicalia, and Field is prepared to accept these assumptions[193].

---

[188] Field 1989, p. 2: "A fictionalist about mathematics-taken-at-face-value is someone who does not literally believe mathematical sentences. (Or, if you prefer to 'semantically ascend', a fictionalist is someone who does not regard such sentences, taken at face value, as literally true.)".

[189] *ibid.*, p. 6: "The negative content of fictionalism is that it avoids having to answer some questions that need answering on a platonistic view. This is the main *motivation* for fictionalism."

[190] *ibid.*, p. 20-21: "A noteworthy feature of mathematics is that there is a tremendous amount of arbitrariness as the identification of different types of mathematical objects. The most famous example of this is the one highlighted by Paul Benacerraf 1965: if one wants to identify natural numbers with sets, it seems rather arbitrary which sets one picks. But of course, this example is just the tip of the iceberg. ... It seems absurd to suggest that in each such case there is a fact of the matter as to which sets the mathematical objects in question are. ... I think that the most natural conclusion is that topological spaces, and numbers, and ordered pairs and functions are neither definitely sets nor definitely not sets: *there is no fact of the matter* about whether they are sets or not, in addition to there being no fact of the matter as to which sets they are if they are sets."

[191] *ibid.*, p. 22: "I think that by far the best explanation of the pervasive arbitrariness (an explanation persuasively developed in Wagner 1982) is the fictionalist one: we have a good story about natural numbers, another good story about sets, and so forth; and in these stories it is completely unimportant whether one identifies numbers with sets, and unimportant which sets one does identify them with if one does want an identification."

[192] *ibid.*, pp. 25-26: "Benacerraf's challenge ... is to provide an account of the mechanisms that explain how our beliefs about these remote [mathematical] entities can so well reflect the facts about them. The idea is that *if it appears to be in principle impossible to explain this*, then that tends to *undermine* the belief in mathematical entities, *despite* whatever reason we might have for believing in them."

[193] *ibid.*, p. 17: "After all, the theories that we use in explaining various facts about the world not only involve a commitment to electrons and neutrinos, *they involve a commitment to numbers and functions and the like*. ... I think that this sort of argument for the existence of mathematical entities (the Quine-Putnam argument, I'll call it) is an extremely powerful one, at least *prima facie*."

iii. Anti-Hermeneuticism:

Any conception of non-literal hermeneutic "mathematical truth" (e.g., of the sort discussed in Chapters 6-8 above) is, at best, philosophically irrelevant[194]. If it is possible to *eliminate* mathematics from science fully, then it is unimportant to develop a novel understanding of "truth", according to which mathematics is still "true in some sense", although still not true[195].

iv. Conservativeness Claim

At the heart of Field's deflationism is the basic idea that mathematics is "insubstantial", "dispensable" or "redundant". In *Science Without Numbers*, Field argued that the insubstantiality of mathematics can be formulated more precisely as a *conservativeness claim*[196]. The result of "adding" mathematics to any mathematics-free theory must always yield a conservative extension. For example, Field argues that the result of expanding a any mathematics-free theory **N** by adding the axioms of applicable set theory (say, **ZFU**) always yields a conservative extension.

Field imposes the following as a constraint on any proposed mathematical theory. Such a theory is "good" if it is conservative with respect to all non-mathematical theories. But, Field stresses, mathematics needn't be *true* to be good:

> Let us call a mathematical theory that is consistent with every internally consistent theory about the physical world conservative . . .

Also, Field 1988 (1989), pp. 249-250: "... after all, *the existence of mathematical entities follows from the mathematical theory itself*, not just from the claim that the mathematical theory is true in the correspondence sense."

[194] Field 1989, p. 2: "The fictionalist may believe that there is some non-face-value construal of mathematical sentences under which they come out true; he or she may even believe that some such construal gives the 'real meaning of' the mathematical sentence, despite its departure from what the mathematical sentence appears to mean on the surface. My own view, though, is that the second of these additional claims is an uninteresting verbal one insofar as it goes beyond the first; and that the first, though of some interest, unnecessarily constricts the fictionalist."

[195] *ibid.*, p. 3: "The fictionalist can say that the sense in which '2 + 2 = 4' is true is pretty much the same as the sense in which 'Oliver Twist lived in London' is true; the latter is true only in the sense that it is true *according to a certain well-known story*, and the former is true only in that it is true *according to standard mathematics*."

[196] This is connected to one of the ideas of Chapter 5 above. A deflationist about truth should analogously hold that truth-theoretical principles are similarly dispensable, and this may be formulated as a conservativeness claim.

> A mathematical theory M is conservative if and only if for any assertion A about the physical world and any body N of such assertions, A doesn't follow from N + M unless it follows from N alone.
>
> Our modified anti-realism, then, says that besides being sufficiently comprehensive to be interesting, a mathematical theory must be conservative, but need not be true.
>
> (Field 1982 (1989), p. 58).

In the Appendix to Chapter 1 of his 1980, Field proves a theorem to the effect that if N is a "mathematics-free" theory (i.e., not containing $\in$), then any model of N may be expanded to a model of N $\cup$ ZFU. (There is a subtlety involved here, involving any *axioms schemes* that appear in the axiomatization of N, to which we return later). From this it follows that any mathematics-free theorem of N $\cup$ ZFU is already a theorem of N. In particular, it also implies that any mathematics-free theorem of ZFU alone must be a logical truth[197]. This settles a question which Quine recently asked,

> The accepted wisdom is that mathematics lacks empirical content. ... I do ... accept the accepted wisdom. No mathematical sentence has empirical content, not does any set of them. No conjunction or class of purely mathematical truths, however large, could ever imply a synthetic observation categorical. It seems obvious and I accept it (though I can't picture how a proof would look)
>
> (Quine 1995, p. 53).

v. Utility of False-But-Conservative Mathematics

If mathematics is not true, but merely conservative, then what is its utility in science? According to Field, conservative mathematics can be useful in drawing nominalistic conclusions from nominalistic premises.

Field gives an example[198] analogous to the following. Suppose that there are 7 women in the room and that there are 93 men in the room (and that no men are women). Then, defining 'person' as 'man or woman', we immediately infer that there are 100 persons in

---

[197] In Chapter 5, I showed this same "contentlessness" effect in relation to DT, the disquotation theory of truth composed of all the T-sentences.

[198] See Field 1980, Chapter 2, 'First Illustration of Why Mathematical Entities are Useful: Arithmetic'. [The recursive definition Field gives on p. 21 of the numerical quantifiers is slightly mistaken]. Field's example uses aardvarks, and involves multiplication. Field's idea here is influenced by a similar argument in Putnam 1967a, which I cited earlier in Chapter 7 (Section 7.5 on "Abstract Counterparts"). Indeed, I suspect that this idea has its origins in some of the remarks of the logical empiricists, notably Hans Reichenbach (Putnam's teacher). In fact, the claim that mathematics is nothing but a useful instrument is defended in Hempel 1945.

the room. But to infer this we need the *arithmetical premise* '7 + 93 = 100' (plus a few other premises[199]). Field argues that this mathematical premise is actually redundant, because the mixed arithmetical premises can be reformulated using numerically definite quantifiers. Instead of,

(1) The number of women in the room = 7

we could (in principle) say something like,

(2) There is a woman x_1, ... and a woman x_7 (all distinct) such that ...

(and likewise for the assertions about the number of men and the number of persons). Then, the conclusion (that there are 100 persons in the room) follows from these mathematics-free premises by ordinary first-order logic. In short, in this application of arithmetic, the mathematics used is dispensable. It merely simplifies the reasoning[200].

Field generalizes this to a strong claim that mathematics facilitates or simplifies logical reasoning:

> ... any inference from nominalistic premises to a nominalistic conclusion that can be made with the help of mathematics could be made (usually more long-windedly) without it.
>
> (Field 1980, p. x).

> ... the conclusions we arrive at [by adding the mathematical theory] are not genuinely new, they are already derivable in a more long-winded fashion from the [nominalistic theory] without recourse to mathematical entities
>
> (Field 1980, pp. 10-11).

> We can use [the mathematical theory] as a device for drawing conclusions . . . much more easily than we could draw them by a direct proof ...
>
> (Field 1980, p. 28).

> ... it might be much easier to see that φ follows from N + M than it is to see that φ follows from N alone ...
>
> (Field 1982 (1989), p. 58).

[199] In particular: if $\#F = n$ and $\#G = m$ and $\neg\exists x(F(x) \wedge G(x))$, then $\#(F \vee G) = n + m$.

[200] To *actually perform* such "mathematics-free" deductions in first-order logic, even involving 'there are exactly two', actually consumes vast amounts of space. The complexity of the deductions increases at a vast rate. But Field's point is that the reformulation is always possible *in principle*.

It will emerge that these claims of Field are ambiguous. Indeed, on the important reading (the deductive reading), they are false.

vi. Rejection of the Indispensability Thesis

Field's plan is to argue, against Quine and Putnam, that platonistic mathematicized science *is* dispensable. Roughly, for any standard platonistic theory P, one looks for a nominalistic replacement N which is "just as good" as P. The constraints on being "just as good" involve a standard *accessibility* demand (recursive axiomatization with finitely many primitives) plus the requirement that the nominalistic replacement N be able to express and prove just as much about the concrete domain as P. Technically,

i. P must be a deductively conservative extension of N;

ii. N must re-express all classifications amongst concreta expressed within P.

vii. Anti-Instrumentalism:

In contrast with the purely *destructive*, *instrumentalist* form of nominalism[201], Field urges that nominalistic replacements for mathematicized theories are *required.* He argues that unless such replacements are provided, then the advocacy of nominalism or fictionalism is sheer "double-think"[202].

viii. Modal Analysis of Metalogic;

In Chapters 1 to 3, we pointed out that metalogic is just as mathematical as mathematical physics is. Field proposes (see below) that physics be nominalized using *geometry.* He argues that it is inappropriate to introduce modal notions for this purpose. However, he does argue later that metalogical notions of implication and consistency are

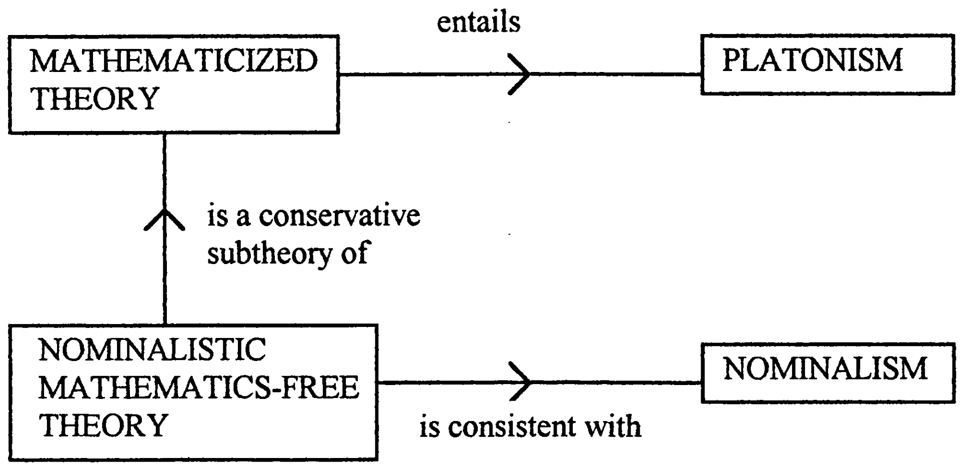
[201] These are Burgess' terms (Burgess 1983, Burgess & Rosen 1997). Burgess argues that this form of nominalism is purely *destructive,* offering no positive or constructive proposal which might *explain* why, if there are no mathematicalia, the Universe behaves "as if" mathematicized theories were true. Burgess compares any such destructive idea with the view advocated in van Fraassen 1980, where the mystery that the Universe behaves "as if" General Relativity is true is left unexplained.

[202] I cited Field's accusation of "double-think" in Chapter 3 (Section 3.2). See Field 1980, p. 2.

to be analysed *modally*. The basic notion of nominalistic metalogic is a primitive concept of "logical possibility". We lack the space to discuss this important topic[203].

ix. The Aim: Nominalistic Science

To summarize, Field's deflationary programme aims to replace platonistic science with mathematics-free nominalistic science:



The question is then: can this programme be fulfilled?

9.2 Geometrical Nominalism

The revolutionary aim of deflationism is to provide *replacement non-mathematical theories* for standard accepted mathematicized theories of Nature. If **N** is the platonistic theory to be replaced, then the basic constraints on a suitable replacement **N** are:

i. **N** must be consistent with nominalism

ii. **N** must be axiomatized by finitely many primitives.

[203] See Field 1984a (1989) and the Introduction to Field 1989. See Chihara 1990, Appendix, for a critique. Malament 1982 contains an important criticism that Field requires substantial mathematics in the metatheory to prove his metatheorems that "mathematics is dispensable". See also Burgess & Rosen 1997 (pp. 192-194), where they question Field's use of mathematics for metalogical purposes.

iii. N must have the same "physical content" as P. (N must re-express and prove everything about the concrete domain as does P).

Field 1980 contains an explicit development of an allegedly mathematics-free replacement theory for an important mathematicized theory, namely Newtonian Gravitational Theory (NG). Let us call this nominalistic replacement theory NG*. Field argued that NG* has all the non-mathematical content of NG, but that it does not postulate mathematicalia, like numbers, functions, etc. According to Field, NG* has the three crucial properties required (i) - (iii).

Property (iii) would guarantee that NG and NG* have all the same non-mathematical theorems (and that any factual classifications amongst concreta expressed in NG are re-expressed within NG*). We shall see below that (iii) actually fails. NG is not a conservative extension of NG*.

NG* it is formulated in a mathematics-free or nominalistic language L_N which is an *extension* of first-order logic, incorporating individual variables, *x*, *y*, ...[204] plus monadic second-order variables *X*(*x*), ...[205] (informally true of *regions* of such points). For the purposes of characterizing the geometry of space-time, L_N contains primitives:

Bet(*x*, *y*, *z*): *x*, *y* and *z* lie on a straight line and *x* lies between *y* and *z*;

Cong(*x*, *y*, *z*, *w*): the line segment *xy* is congruent to the line segment *zw*.

An axiomatization of flat Euclidean geometry using these primitives was given long ago by Hilbert[206]. The resulting axiomatic geometry is sometimes called *Synthetic Euclidean Geometry*, since all notions from mathematical analysis (real numbers) are absent from the finished product[207]. This axiomatic presentation is, in effect, a culmination of the two thousand year programme initiated in Euclid's *Elements*. We'll call the resulting axiom system SEG. (In our discussions below we shall, for simplicity, consider that a version

[204] Intuitively ranging over spacetime points.

[205] Informally true of *regions* of spacetime.

[206] See Hilbert 1899.

[207] For a comprehensive discussion of the geometrical background, including a wealth of technical material, see Burgess & Rosen 1997, Chapter II.A.

appropriate to just 2-dimensional plane geometry). In fact, Hilbert's axiomatic presentation of SEG precipitated an interesting controversy around 1901 between Frege and Hilbert about the role of mathematical theories as "implicit definitions"[208].

Hilbert 1899 proved two important metatheorems about this axiom system SEG, which in modern parlance amount to model-theoretic results:

i. Representation Theorem

Any model $M = (D, Bet, Cong)$ of SEG can be represented by the standard mathematical field structure of analysis $F = (\mathbf{R}, 0, 1, +, \times, <)$. Technically, the there is a injection $\tau: D \rightarrow \mathbf{R}^2$ such that the relations $Bet \subseteq D^3$ and $Cong \subseteq D^4$ can be "defined" in the field F[209]. For geometry, the homomorphism is a *co-ordinate system*[210];

ii. Uniqueness Theorem

Any pair of such representations τ_a and τ_b are related in a certain way: they are "unique" up to certain transformations. For SEG, these transformation between representing homomorphisms are, in fact, *Euclidean co-ordinate transformations*[211].

These metatheorems shed considerable light on how mathematics (that is, analysis) is applied to geometry. Any representation function τ is a co-ordinate scale, assigning real

---

208 See Shapiro 1997, pp. 161-165. One of Hilbert's colleagues, Otto Blumenthal, reported a remark of Hilbert's in a Berlin train station in 1891. According to Hilbert, in a proper axiomatization of geometry "one must always be able to say, instead of 'points, straight lines and planes', 'tables, chairs, and beer mugs'". This position simply infuriated Frege, whose interests were ontological and for whom it was obvious that a *plane in physical space* is not the same thing as a *beer mug*!

209 More precisely, let $d((x_1, y_1), (x_2, y_2))$ be the Euclidean metric on $\mathbf{R}^2$, i.e., $\sqrt{[(x_2 - x_1)^2 + y_2 - y_1)^2]}$. Then suppose that $(E, Bet, Cong)$ is a model of SEG. Then,

$$(e_1, e_2, e_3) \in Bet \Leftrightarrow d(\tau(e_1), \tau(e_2)) + d(\tau(e_1), \tau(e_3)) = d(\tau(e_2), \tau(e_3))$$

and $$(e_1, e_2, e_3, e_4) \in Cong \Leftrightarrow d(\tau(e_1), \tau(e_2)) = d(\tau(e_3), \tau(e_4))$$

210 In the more general case, such a representing homomorphism τ is called a *measurement scale*. See Krantz et al. 1971.

211 E.g., suppose that $(D, Bet, Cong)$ is a model of SEG and τ_a and τ_b are such homomorphisms from D to $\mathbf{R}^2$. Then $\tau_a = \Phi \circ \tau_b$, where Φ is a Euclidean transformation of $\mathbf{R}^2$ (that is, some combination of translation, reflection, rotation or dilation). See Field 1980, pp. 50-51 (or any decent physics textbook, say Longair 1984 or Eisberg 1961).

numbers (more generally, *n*-tuples of reals) to points in any model *M* of **SEG**. After Hilbert 1899, various other re-axiomatizations of **SEG** and extensions have been introduced. In particular, important facts have been established concerning matters of completeness and decidability. An important contribution is the monograph by Tarski 1951.

We have quickly surveyed the geometrical parts of the theory **NG*** that Field 1980 proposes. But there is a further, and highly non-trivial, component of **NG*** involving the re-axiomatization of the laws of motion (for particles) and the field equation governing the gravitational potential (usually known as Poisson's Law: schematically, $\nabla^2\Phi = \rho$). In a manner analogous to the synthetic axiomatization of geometrical laws, Field shows how to re-axiomatize these kinematical and dynamical laws in the mathematics-free language $\mathbf{L_N}$. The required non-mathematical primitives in $\mathbf{L_N}$ are:

Mass-Bet(*x*, *y*, *z*):	the mass density at *x* lies between that at *y* and that at *z*
GPot-Bet(*x*, *y*, *z*):	the gravitational potential at *x* lies between that at *y* and that at *z*

The axioms governing these physical primitives are analogous to those governing the space-time primitives, *Bet* and *Cong*. The hard work is to obtain axioms such that an Extended Representation Theorem is forthcoming, which, given any model *M* of the theory **NG***, proves the existence of generalized representation functions Φ and ρ from dom(*M*) to $\mathbf{R}^3$ such that $\nabla^2\Phi = \rho$ (so that Poisson's Law is satisfied). This is all carefully worked out in great detail in Chapters 7 and 8 of Field 1980.

The central philosophical issues concern **NG***'s consistency with nominalism and its adequacy as a replacement for **NG**. But there are some further aspects of conceptual and philosophical interest concerning **NG***.

i. "Synthetic" Characterization of Spacetime Structure

Field argues that the structure of (flat, Euclidean) space-time may be characterized semantically within **NG*** in terms of non-mathematical predicates, which, in effect, determine the topological and metric structure of the set of space-time points. Because there is no direct reference to (quantification over) real numbers or functions, Field considers this a major advance[212]. There are no *preferred* quantitative measurement scales or co-ordinates. In Field's terminology, **NG*** explains the motion of test bodies due to gravity "intrinsically". This is interesting, although the "cost" is the *second-order* nature of **NG***. This ought to be intolerable for a nominalist because, as we shall see, the construction of these explanations within **NG*** (that is, the deduction of nominalistic consequences) actually requires the reintroduction of mathematics.

ii. The Conventionality of Mathematical Representation

The Representation Theorem for **NG*** shows how for any model M of **NG***, there is an appropriate representation τ of M within F, and there are functions Φ and ρ which represent the nominalistically expressible facts about the gravitational potential and the density potential (that is, are such that $\nabla^2\Phi = \rho$). These representations τ, Φ and ρ are unique up to certain transformations. These transformations yield precisely the Galilean transformations of co-ordinate systems and the scale transformations for the field potentials. As Field stresses, this explains the conventionality of measurement scales and co-ordinate scales and clarifies some issues concerning invariance:

> This result gives an *explanation* of the fact that the laws of Euclidean geometry, when stated in terms of co-ordinates, are invariant under a shift of origin, reflection, rotation and multiplication of all distances by a constant factor: if we assume that the *genuine facts* about Euclidean space are just the facts about betweenness and congruence laid down in Hilbert's axioms, and that the function of co-ordinates is simply to facilitate the deduction of facts about betweenness and congruence and the relations definable in terms of these, then it *follows* that in an extrinsic formulation of the laws of geometry in terms of co-ordinates, the laws will be invariant up to Euclidean transformations and no further.
>
> (Field 1980, p. 51)

Perhaps this has implications for conventionalism elsewhere in the philosophy of science. One might argue that it shows how we can be realists about the "contentful" parts of our theories, but anti-realists about those parts that are "purely conventional". This can, however, be misleading. The *conventionality* of co-ordinate charts does not

---

[212] But Shapiro 1997 urges that because NG* characterizes the *structure* of spacetime correctly, it simply *is* mathematics (in sheep's clothing, maybe). See below, Section 9.4, Part (d).

entail the *non-existence* of co-ordinate functions or real numbers. It simply entails that there is some "slack" in our representation of the world. The "genuine facts" are, loosely speaking, what remain *invariant* under all admissible variations of these possible choices (something Einstein put to great effect). But it does not follow from this alone that we can *eliminate* implicit reference to these choices from our mathematicized theories.

This is analogous to the Quine-Benacerraf point about the conventionality available in identifying mathematicalia with sets. The conventionality or "non-facticity" in identifying 2 as either $\{\{\emptyset\}\}$ or $\{\emptyset, \{\emptyset\}\}$ does not entail the *non-existence* of numbers. Similar non-facticities in identifying real numbers, tensor products, or topological spaces with sets does not entail their *non-existence.* And similar Quinian non-facticities[213] between ordinary common sense discourse (about mountains, desks, galaxies and molecules, say) and tough "mereological" physics (about regions of curved space-time and quantum fields, say) does not entail the *non-existence* of mountains, desks, galaxies or molecules.

iii. Burgess: Synthetic Mechanics

Field's geometrical approach of 1980 is reconstructed in Burgess 1984 (and more fully in Burgess & Rosen 1997, Chapter II.A). Burgess begins with analytically formulated Euclidean plane geometry. Call this theory AEG. He then shows that AEG has the Tarskian "elimination property" (mentioned in Chapter 7 above). That is, AEG may be extended to a deductively and expressively conservative extension AEG^+ in an expanded notation L^+, and AEG^+ is fully conservative over its primary restriction $(\mathsf{AEG}^+)^\circ$ in $(\mathsf{L}^+)^\circ$.

Indeed, the primary restriction $(\mathsf{AEG}^+)^\circ$ is equivalent to synthetic geometry, SEG. The reconstruction provides a re-interpretation which maps all analytic assertions within analytic geometry AEG (mentioning real numbers) to synthetic assertions about points in synthetic geometry SEG.

---

213 Quine's classic discussions of "non-facticity" and "inscrutability of reference" appear in Quine 1969b and Quine 1977. See also Davidson 1977 and Putnam 1988 for Putnam's "model-theoretic" version of Quine's arguments.

The crucial element of the reconstruction is the existence of representation formulas for the real numbers. That is, there are existence and uniqueness formulas in the language of **AEG**,

i. $\forall X \exists x_1 \ldots \exists x_n R(x_1, \ldots, x_n, X)$

ii. $\forall x_1 \ldots \forall x_n \forall X_1 \forall X_2[(R(x_1, \ldots, x_n, X_1) \wedge R(x_1, \ldots, x_n, X_2)) \rightarrow X_1 = X_2]$

which are deducible in **AEG**. Burgess applies what he calls "Tarskian elimination" to generate the nominalistic reconstruction **SEG**. I refer to reader to Chapter II.A of Burgess & Rosen 1997.

9.3 Representation Theorems

In general, a Representation Theorem of the kind desired by Field connects a given non-mathematical axiomatic theory **N** and a given platonistic theory **P** and says, roughly,

> If $(D, N_1, N_2, \ldots) \models$ **N**, then there is a injection τ: $D \mapsto D_M$ (a domain of mathematicalia) such that,
>
> i. for any x_i in D, $(x_1, x_2, \ldots) \in N_i$ iff $(\tau(x_1), \tau(x_2), \ldots.) \in P_i$; and
>
> ii. the expansion $(D \cup D_M, N_1, N_2, \ldots, P_1, P_2, \ldots) \models$ **P**.

(where the P_i are purely mathematical relations on the mathematicalia $\tau(x_1)$, $\tau(x_2)$, etc. in the domain D_M).

For Field, there are two important applications of such Representation Theorems:

i. Introduction of "Abstract Counterparts";

ii. Proof that **P** is (Semantically) Conservative over **N**.

Illustrating (i), the Representation Theorem entails:

$$\mathsf{M} \vdash \exists\tau\forall x_1 \ldots [N_i(x_1, x_2, \ldots) \leftrightarrow P_i(\tau(x_1), \tau(x_2), \ldots)]$$

This means that any non-mathematical statement φ involving the non-mathematical predicates N_i can be "translated" into an abstract counterpart $\varphi_{abs}(\tau)$, involving mathematical predicates P_i. For example, suppose *Less-Mass*(x, y) is a predicate meaning, intuitively, 'x is less massive than y', as applied to physical bodies. Then equipped with the Representation Theorem for a *Less-Mass*-representing measurement scale τ we can take the contingent nominalistic statement:

(1) $\exists x \exists y(\textit{Less-Mass}(x, y))$

and "translate" it as:

(2) $\exists x \exists y(\tau(x) < \tau(y))$

This is because we have the following statement, which is entailed by **M** *alone*.

(3) $\exists \tau[$"τ represents *Less-Mass*" $\wedge \forall x \forall y(\textit{Less-Mass}(x, y)) \leftrightarrow (\tau(x) < \tau(y))]$

Thus, when we have a Representation Theorem for **N** we can immediately derive an abstract counterpart for each nominalistic statement in the language of **N**. However, the reverse need not obtain: it doesn't immediately follow that for each *platonistic* statement with physical content in the language of **P**, we can find a nominalistic counterpart for it.

Illustrating (ii), Field invokes the appropriate Representation Theorem for his nominalistic theory **NG*** to argue that this particular theory is in fact a suitable replacement for the standard platonistic theory **NG**. Indeed, what Field shows is that **NG** is *semantically conservative* over **NG***. The idea, briefly, is that any model of **NG*** may be expanded to a model of **NG**. Let us see how this works. What we want to show is that,

If:

if $(D, N_1, N_2, ...) \models$ **N**, then there is a representing function τ into a domain D_M of mathematicalia such that,

i. for any x_i in D, $(x_1, x_2, ...) \in N_i$ iff $P_i(\tau(x_1), \tau(x_2), ...)$, and

ii. the expansion $(D \cup D_M, N_1, N_2, ..., P_1, P_2, ...) \models$ **P**, Then:

if $\mathsf{P} \models \varphi$, then $\mathsf{N} \models \varphi$, for any $\varphi \in {}_0\mathsf{L}^\circ$.

Proof: Assume that if $(D, N_1, N_2, ...) \models \mathsf{N}$, then there is a representing function τ into a model $(D^*, N_1, N_2, ..., P_1, P_2, ...)$ of P such that $N_i(x_1, ..., x_n) \leftrightarrow P_i(\tau(x_1), ..., \tau(x_n))$, where the N_i are the extensions of non-mathematical predicates and the P_i* are the extensions of pure mathematical predicates. Now, also assume $\mathsf{P} \models \varphi$ and not:-($\mathsf{N} \models \varphi$). So, there is a model $\mathfrak{I}$ of N such that $\mathfrak{I} \models \neg\varphi$. There is a representing function τ into model $\mathfrak{I}^+ = (D^*, N_1, N_2, ..., P_1, P_2, ...)$ of P such that $S_i(x_1, ..., x_n) \leftrightarrow P_i(\tau(x_1), ..., \tau(x_n))$. Clearly, $\mathfrak{I}^+$ is an expansion of $\mathfrak{I}$. So, $\mathfrak{I}^+ \models \neg\varphi$. But $\mathfrak{I}^+ \models \mathsf{P}$. Therefore, $\mathfrak{I}^+ \models \varphi$. So $\mathfrak{I} \models \varphi$. Contradiction. ■

This is how one uses the Representation Theorem for a nominalistic theory N with respect to a platonistic theory P to derive a *semantical* conservativeness theorem. Note that one cannot conclude *deductive* conservativeness unless one has a complete derivability relation (e.g., the theories in question are first-order).

For a more concrete example, consider our toy theory of time Tim considered several times earlier. Tim just says that the set of temporal instants under the "before" relation is isomorphic to $(\mathbf{R}, <)$. Imagine we are looking for a nominalistic replacement N for Tim. The required Representation Theorem for N would be simply:

(*) If (D, Bef) is a model of N then (D, Bef) is isomorphic to $(\mathbf{R}, <)$.

Now this would then guarantee that,

(**) If (D, Bef) is a model of N then $(D \cup \mathbf{R}, Bef, <)$ is a model of Tim.

Imagine we have such a nominalistic replacement N. The proof of semantical conservativeness proceeds as follows. Suppose that for some $\varphi \in {}_0\mathsf{L}^\circ$, Tim $\models \varphi$ and not:-($\mathsf{N} \models \varphi$). So there is a model (D, Bef) of N and $(D, Bef) \models \neg\varphi$. By the Representation Theorem (**), there is a model $(D \cup \mathbf{R}, Bef, <)$ of Tim. Thus, $(D \cup \mathbf{R}, Bef, <) \models \varphi$. Thus, $(D, Bef) \models \varphi$. Contradiction. ■

It is worth noting for later that a Representation Theorem for a nominalistic replacement Tim* for Tim would require that Tim* be *categorical* and, further, that any such model be of cardinality $\mathfrak{c}$. Given that almost all interesting applications of mathematics involve non-denumerable domains, this constraint effectively *rules out Representation Theorems for first-order nominalistic theories*. For suppose that Tim* is a first-order theory and has an infinite model. Then, by the downwards Löwenheim-Skolem Theorem, Tim* also has an $\aleph_0$-model. But, by Cantor's Theorem, this cannot be isomorphic to **R**. Hence, we conclude that *there is no first-order replacement theory* Tim* that satisfies the required Representation Theorem for Tim.[214]

9.4 Criticisms of Geometrical Nominalism

a. Second-Order Formulation

Field 1980 offers two non-mathematical replacements for Newtonian space-time and gravitational field theory. The main one, for which the Representation Theorem is actually proved, uses the apparatus of second-order logic.[215] For example, Field uses the following second-order axiom to say that "space-time is continuous":

DedCont: Any bounded line segment has a closest bound.

He calls this postulate the *Dedekind continuity postulate*[216]. It is the mereological version of the usual axiom for order-completeness for a linear ordering:

DedCont: Any bounded set has a least upper bound.

Some critics have argued that the use of second-order logic transcends the limits of nominalism. One basis for this argument is Quine's analysis of the ontological commitments of statements containing quantified predicate letters. As Quine quipped, "second-order logic is set-theory in sheep's clothing"[217].

---

[214] Of course, there may be a *second-order* theory which satisfies a Representation Theorem.

[215] A Representation Theorem for the *first-order* version cannot be proved.

[216] See Field 1980, p. 37.

[217] See Quine 1970, pp. 66-68.

But, the argument may be too quick. Not *all* applications of second-order logic *must* receive a set-theoretical interpretation. Briefly, there are two ways of avoiding sets, if one restricts oneself to just monadic second-order logic

i. Plural Quantification (Boolos)

Boolos 1984, 1985 has supplied an interpretation of *monadic* second-order logic based on the "logic of plurality" and argued that second-order variables need not be interpreted as taking *sets* as values, but rather interpreted as referring to individuals "plurally".

ii. Mereological Interpretation

The nominalistic interpretation given by Field to the second-order quantifiers is *mereological*: the second-order variables range over mereological space-time regions (not *sets* of space-time points). That is, if $X(x)$ is a monadic second-order formula in the language of non-mathematical physics and σ is such a mereological valuation, then $\sigma \models X(x)$ iff $\sigma[x] \subseteq \sigma[X]$, where $\sigma[X]$ is a space-time region, not a set of points (and $x \subseteq y$ iff x is a part of y)[218].

Setting aside Boolos' controversial plural interpretation, we are left wondering whether Field's *mereological* interpretation of monadic second-order logic is legitimately nominalistic. Of course, the *metatheory* is mathematical, but the *values* of object-level second-order variables in the non-mathematical object-level theory are nominalistically acceptable *regions*, rather than nominalistically unacceptable *sets* of points.

In any case, the *proof theory* for such an extension of first-order logic remains the same as the proof theory for second-order logic (that is, one has rules of instantiation and generalization for second-order variables, plus the *comprehension axiom scheme*: $\exists X \forall x(X(x) \leftrightarrow \Phi)$, where the variable X is not free in the formula Φ, plus a second-order version of the axiom of choice[219]). However, for second-order logic with the standard

[218] The symbol $\subseteq$ is often used in mereology, mainly because mereology is "almost" isomorphic to Boolean algebra of class inclusion. The difference is that there is no empty or null region $\varnothing$ which is a part of all regions. See Simons 1987.

[219] See Shapiro 1991, pp. 65-70 for the "Deductive System" for second-order logic.

semantics[220], the Completeness Theorem fails. Indeed, the class of second-order logical truths is not recursively enumerable[221].

b. Geometrical Substantivalism

NG* postulates "geometricalia": *space-time points* and *regions*. It might be argued that space-time points are unobservable theoretical entities, better thought of as abstracta. Allegedly, any such reification of unobservable spatio-temporal entities transcends the limits of acceptable nominalism[222]. Again, it seems to me that such a criticism is very weak.

First, it is not clear in what sense quantification over space-time points is meant to transcend the ontological constraints of nominalism. One tentative definition of 'concretum' makes essential reference to space-time:

> A *concretum* is an entity embedded in space-time.

Thus, it would be very odd to claim that the mereological atomic parts of space-time (if there are such atomic parts) are not themselves concreta[223].

Second, although there has been a long tradition in metaphysics and philosophy of science of denying the existence of space-time regions and points, it seems to me that this tradition is neither particularly nominalist in motivation, nor even *compatible with current physical theory* (specifically, general relativistic space-time theory). For current

---

220 See Shapiro 1991, pp. 70-76. Intuitively, for a given domain D, n-adic predicate variables range over *all n-place relations* on D. That is, if X is an n-place predicate variable, it ranges over $P(D^n)$.

221 If φ is a formula in the language of arithmetic, then $\mathfrak{N} \models \varphi$ iff $PA^2 \models \varphi$. Thus, φ is an arithmetical truth iff $\models PA^2 \rightarrow \varphi$. Let V^2 be the class of second-order logical truths in the second-order language of arithmetic. Suppose V^2 were r.e. Then, by Craig's Theorem, there would be a proof procedure for membership in V^2. Thus, all logical truths of the form $PA^2 \rightarrow \varphi$ could be proved. From this it would follow that all arithmetical truths could be proved (i.e., if the procedure proves $PA^2 \rightarrow \varphi$, then take φ as proved), and this contradicts the incompleteness of any proof procedure for arithmetical truth. See Boolos & Jeffrey 1989, pp. 203-204.

222 See Malament 1982. Malament argues that space-time points are "not concrete physical objects in any straight-forward sense. They do not have mass-energy content (unlike, for example, the Klein-Gordon field itself). They do no suffer change. It is not even clear in what sense they exist *in* space and time" [p. 532].

223 The reason is simple. If one added the term 'concretum' to mereology. then it would be natural to suppose *closure principles*: if x and y are concrete, then all *parts* of them are concrete and all *aggregates* of them are concrete. It is, of course, debatable whether aggregation is itself closed (that is, whether "unnatural" aggregates exist. E.g., is there such a thing as the aggregate of The Eiffel Tower and The Chrysler Building? If there were, I'd call it concrete, not abstract).

physical relativistic space-time theory quantifies over space-time points: GR says things like:

i. the *set of space-time points* is the domain of a topological manifold;

ii. the open sets of *space-time points* (in the physical topology of space-time) can be assigned co-ordinates in a continuous way.

iii. *each space-time point e* has a Ricci curvature tensor proportional to its energy-momentum tensor

Geometrical objects of importance in physics, like tensor fields, are (defined as) functions from space-time points to mathematicalia, such as tangent vectors in the tangent space $T_e(M)$ at each space-time point *e*. To deny this is not to comprehend how modern relativistic physics works, irrespective of nominalistic scruples.

Third, it is debatable to claim that space-time regions are unobservable. Quite the opposite. We can point to and observe (fairly large and ostensively vague) regions of space-time. We can point to the spatial region between the legs of a chair (adding, perhaps "I mean the *empty gap*").

Fourth, it is implausible to claim that space-time points and regions are causally inert. The crucial physical field functions defined over space-time are not causally inert: according to the Lorentz force law, the acceleration a_μ of an electron at a space-time point *e* is a function of the electromagnetic field A_μ at *e*. The Lorentz Force Law states that, at each space-time point *e*, $ma_\mu(e) = e_0 v^\nu(e)\partial_\nu A_\mu(e)$ (where e_0 is the electron charge and v^μ is the velocity 4-vector). It seems to make sense to say that in assigning a causally potent electromagnetic field A_μ at each point *e*, we are effectively assigning "causal powers" to each point. After all, the way an electron "moves through" the space-time manifold depends upon the "causal powers" of the points it happens to move through: if an electron is currently located at *e*, its subsequent motion is determined by the "causal powers" of *e*. More exactly, the space-time trajectory of an electron is determined by the electromagnetic field A_μ (and, more generally, by the Ricci curvature tensor $R_{\mu\nu}$ of space-time as well).

c. Non-Finitism

Another potential criticism of NG* is that it postulates *infinitely* many space-time points (the domain of a model of NG* must contain precisely continuum many elements). It is not clear that this is an argument against nominalism, as Field himself construes the term. It seems to me that a nominalist of Field's stripe can quite consistently claim that there are infinitely many concreta. The person who criticizes Field's account in this way is conflating *anti-platonistic* nominalism (which says that there are no abstract mathematicalia) with the completely different and much stronger thesis, *finitistic* nominalism (which says that there are only finitely many entities):

> ... the nominalistic objection to using real numbers was not the grounds of their uncountability or of the structural assumptions (e.g., Cauchy completeness) typically made about them. Rather the objection was to their abstractness: even postulating *one* real number would have been a violation of nominalism. ... Conversely, postulating uncountably many *physical* entities ... is not an objection to nominalism; nor does it become any questionable when one postulates that these physical entities obey structural assumptions analogous to the ones that platonists postulate for the real numbers.
>
> (Field 1980, p. 31).

The deflationist about mathematics needn't advocate finitism. Irrespective of past usage of the term 'nominalism', deflationism is quite distinct from finitism. The deflationary fictionalist simply rejects abstracta mathematicalia. Indeed, in order to get any nominalistic theory of the ground, Quine, Church and others have argued that it is *necessary* to assume the existence of (or the modal constructibility of) infinitely many concreta. Field and deflationists seem to recognize this.

d. Structural Assumptions

The final objection to Field's programme that we shall consider is related to the previous one, and has recently been articulated by Shapiro:

> There is a revealing error in Hartry Field's *Science Without Numbers* [1980] ... Whether abstract or concrete, Field's Newtonian space-time is Euclidean, consisting of continuum-many points and even more regions. Space-time exemplifies most (but not all) of the structure of $\mathbf{R}^4$ ... So something like addition and multiplication, as well as the calculus of real-valued functions, can be carried out in this nominalistic theory. All of this is supposed to be consistent with the nominalistic rejection of *abstracta*.
>
> (Shapiro 1997, pp. 75-76).

Field himself was aware of this, considering the objection that "there doesn't seem to be a very significance difference between postulating such a rich physical space and postulating the real numbers" (Field 1980, p. 31). I have quoted Field's reply above, where he says that "even postulating *one* real number would have been a violation of nominalism". To this, Shapiro replies:

> The structuralist balks at this point. For us, a real number is a place in the real-number structure. It makes no sense to "postulate one real number", because each number is part of a large structure. It would be like trying to imagine a shortstop independent of an infield, or a piece that plays the role of the black queen's bishop independent of a chess game. Where would it stand? What would its moves be? One can, of course, ask whether the real-number structure is exemplified by a given system (like a collection of points). ... But it is nonsense to contemplate numbers independent of the structure they are part of.
>
> (Shapiro 1997, p. 76)

But this seems an over-reaction. Field would probably agree that it doesn't make sense to postulate the existence of, say, just thirteen real numbers, without the rest. I expect that Field would probably say something like this: "if *there were real numbers*, then they would have to satisfy the usual axioms for an order-complete field" (He has said something exactly analogous to this in his critique of Wright 1983).[224]

However, Shapiro thinks that mathematics just *is* "the deductive study of various kinds of structure":

> Field concedes that nominalistic physics makes substantial "structural assumptions" about space-time, and he articulates these assumptions with admirable rigor. Although Field would not put it this way, the "structural assumptions" characterize a structure, an uncountable one. This is a consequence of the fact that (the second-order version of) Field's theory of space-time is categorical—all of its models are isomorphic. Field's nominalistic physicist would study this structure as such, at least sometimes. Field himself proves theorems about this structure. As I see it; he thereby engages in mathematics. The activity of proving things about space-time is the same kind of activity of proving things about real numbers. Both are the deductive study of a structure, no more and certainly no less.
>
> (Shapiro 1997, p. 77).

[224] Roughly, Field 1984b replies to Wright that, *if there were such things as numbers*, then they would satisfy Hume's Principle. But Field denies that there are numbers. Very much in the spirit of fictionalism, he could also add that it probably wouldn't make much sense to postulate Holmes without Watson, or Little Red Riding Hood without the Big Bad Wolf, or Santa Claus without Rudolph, his red-nosed reindeer. And so on.

This position seems to lead to a *reductio*. *Any* assumption (about, say, the relative masses of the professors as the LSE, or the relative extremity of varying advocates of post-structuralist cyber-feminism) can be thought of as a "structural assumption". But if one makes such a "structural assumption", and investigates its deductive consequences, one is surely *not doing mathematics*. One is just *thinking* (about LSE professors or cyber-feminists). Shapiro's position here leads to a *reductio*. Indeed, any satisfiable assumption about the physical world characterizes a class of structures. Suppose the class of structures characterized by the set of (non-indexical) statements I say to my wife in a certain week happens to be isomorphic to some interesting mathematical structure. Does it follow that I am doing mathematics when I speak to my wife? The *absurdum* of the *reductio* which Shapiro's argument yields is that all consistent (satisfiable) discourse about the physical world is mathematics! This cannot be right.

Field's own position is, in a sense, a mirror image of Shapiro's. Field argues that some of our "intuitions" about what mathematical structures there are, actually arise out of a kind of abstraction from considerations about aspects of the physical world. For example,

> The reasons for finding certain mathematical claims natural (even if not literally believable) will vary somewhat from one mathematical claim to another. For instance,
>
> (a) {Human females} $\cup$ {human non-females} = {humans}
>
> is natural to accept largely because of its intimate association with the logical truth '$\forall x$(x is human if and only if either x is human and female or x is human and not female)'. Similarly,
>
> (b) 1 + 1 = 2
>
> is natural to accept largely because of its intimate association with logical truths like 'If there is exactly one apple on the table and exactly one green thing on the table and no apple on the table is green then there are exactly two things on the table which are either apples or green'; ... the situation is similar with,
>
> (c) Between any two real numbers there is another real number
>
> this is natural to accept because of its intimate association with an analogous claim about points on a line in physical space. ...,
>
> (d) For any physical objects x and y, there is a set containing x and y as its only members
>
> draws some of its naturalness from the claim that there is an aggregate of x and y; ...
>
> There are of course mathematical claims that unlike (a)-(d) are not intimately connected to non-mathematical claims, and which seem natural even when we hear them, so that education is not directly a factor: an example is,

(e) There are inaccessible cardinals

(Field 1989, Introduction, pp. 9-10).

Burgess & Rosen 1997 illustrate this neatly with the quick summary:

> For all its wealth of results, and for all the power of its applications, mathematics as of 1800 dealt with only a handful of mathematical structures, all closely connected with the models of time and space used in classical physics: the natural, rational, real, and complex number systems; the Euclidean spaces of dimensions one, two, and three. Indeed, mathematics was widely held to deal directly with the structure of physical space and time, and to provide an example of pure thought arriving at substantive information about the natural world. ... All that changed completely during the nineteenth century with the introduction of more and more novel mathematical structures, beginning with the first non-Euclidean spaces.
>
> (Burgess & Rosen 1997, p. v).

One might argue, *with* Shapiro, that our "access" to these basic mathematical structures, of time and space, is somehow grounded in their concrete exemplifications in the physical world.

e. Extensions: General Relativity and Quantum Theory

The (allegedly) nominalistic reconstruction given by Field is a reconstruction of just one interesting theory from mathematical physics. It is completely unknown whether such a reconstruction is possible other theories.

David Malament expresses scepticism on this matter, arguing that,

> Field's nominalization strategy, even if successful in some cases, almost certainly fails when applied to other physical theories of interest. His example (Newtonian gravitational theory) and mine (the theory of the Klein-Gordon field) are very special. ...
>
> ... Suppose Field wants to give some physical theory a nominalistic reformulation. Further suppose the theory determines a class of mathematical models, each of which consists of a set of "points" together with certain mathematical structures defined on them. Field's nominalization strategy cannot be successful unless the objects represented by the points are appropriately physical (or non-abstract). But in lots of cases the represented objects *are* abstract. In particular, this is true of all "phase space" theories.
>
> (Malament 1982, p. 533).

Malament's point is that certain physical theories represent the possible states of a physical system by the points inside a certain structure, called a "phase space". The problem for Field would be that these *possible states* are themselves abstract.

There are two other groups of theories of central importance within modern mathematical physics. First, theories along the lines of Einstein's *General Theory of Relativity*, the theory of curved space-time and gravitation[225]. Second, the various *Quantum Field Theories*, the theories of quantized fields in space-time[226].

Burgess & Rosen 1997 briefly discuss this matter, concluding,

> ... whether the obstacles enumerated can be surmounted is an open research problem. As a consequence of nominalism's being mainly a philosopher's concern, this open research problem is moreover one that has so far been investigated only by amateurs—philosophers and logicians—not professionals—geometers and physicists; and the failure of amateurs to surmount the obstacles is no strong grounds for pessimism about what could be achieved by professionals.
>
> (Burgess & Rosen 1997, p. 118).

It is impossible to know how this issue will develop. Most serious mathematicians and physicists are completely disinterested in the question of nominalism, and, consequently, it is unlikely that someone will even attempt a "nominalistic" version of, say, Kaluza-Klein supergravity theory or quantized Yang-Mills gauge theory.

But it seems to me that the complex mathematical structure of theories used in modern theoretical physics (Lie groups, fibres bundles, cohomology groups, path integrals over function spaces, and so on) cannot be reduced in any straightforward way to simple

---

225 Classical General Relativity in 4 dimensions can be generalized in several ways. The most obvious, introduced by Theodore Kaluza as early as 1919, is to introduce extra dimensions. With this theory in 5 dimensions, one can derive Maxwell's Equations for the extra components of the curvature tensor! Such theories are still of interest, and are now called Kaluza-Klein theories (Klein introduced the idea that the extra fifth dimension is "curled up", and thus hard to see, and its small size determines the electric charge). Another generalization is to introduce extra fields on spacetime, in particular spinor fields. The symmetry operation that transforms between spinor fields and tensor fields is called "supersymmetry" and the resulting theories are called "supergravity theories". Kaluza-Klein supergravity theories were very popular in the 1970s and early 1980s until the arrival of superstring theories (which only work in specific dimensions, like 26 or 10).

226 The basic quantum field theories currently studied or advocated as fundamental explanations are all examples of the second-quantized Yang-Mills theory with "gauge-invariance". A gauge field is, roughly, the force field between particles of matter (fermions). Quantum Electrodynamics (QED) is the simplest such theory, with electrons as the matter and photons as the gauge field. The standard Weinberg-Salam-Glashow (1967) Electro-Weak Model is also such a theory, incorporating something fancy called "spontaneous symmetry breaking" (SSB). Also Quantum Chromodynamics (QCD) is a gauge theory, this time with quarks as the matter and gluons as the gauge field.

polyadic relations between space-time points, like betweenness and congruence (or their generalization to deal with scalar fields on space-time).

9.5 Non-Conservativeness Within Mathematics

The above criticisms are somewhat inconclusive. But there is a very powerful criticism of Field's programme which attacks the main technical claim that *constitutes* deflationism, the alleged property of conservativeness of adding or applying standard applicable mathematical theories (such as standard set theory with ur-elements) to non-mathematical theories. This central claim is statable simply enough:

> **Conservativeness**
>
> The result of adding applicable mathematical axioms M to a nominalistically formulated theory N always yields a conservative extension.

Field 1980 construes conservativeness in the semantical sense, model-theoretically, using the (Tarskian) definition of logical consequence. But the theory he constructs is second-order, and because the consequence relation for second-order logic is not complete, a strong case can be made for the claim that we can only know facts about this relation by *presupposing* mathematics[227].

Indeed, it is plausible to argue that, on the sort of *epistemology* favoured by a nominalist, the proof-theoretic conception of *deductive* conservativeness should be

[227] Malament 1982 makes a related point: "… a nominalist cannot understand the assertion that sentence S_L is a logical consequence of theory *T*. (What could it mean to say that S_L is true in all set-theoretic models of T?)" [p. 530]. E.g., let G be a fixed point Gödel sentence (for the negation of the provability predicate) of PA^2. We can prove, using set theory, that $PA^2 \models G$, but we cannot finitely derive G from PA^2. Presumably, the nominalist cannot understand the assertion that G is a logical consequence of PA^2.

basic[228]. That is, like all of us, the nominalist thinks that our ability to *recognize* "what follows from what" and of "what is consistent" is somehow grounded in actually performing sound derivations involving linguistic tokens (or perhaps mental tokens), while he or she argues that the more "mathematical notion" of semantic logical consequence is some kind of unacceptable abstraction, to be dispensed with.

A mathematical theory M is *deductively conservative* if and only if, for *any* non-mathematical theory N and any non-mathematical statement φ, if $\mathsf{M} \cup \mathsf{N} \vdash \varphi$ then $\mathsf{N} \vdash \varphi$. This is where things get very sticky for the deflationist claim that mathematics is "insubstantial". For Gödel's Incompleteness Theorems entail the phenomenon of *non-conservativeness* (e.g., for consistent recursively axiomatized extensions of Q). Consider PA, axiomatic first-order Peano Arithmetic. Gödel showed in 1931 in effect that there are true sentences G_{PA} and $\mathsf{Con}_{\mathsf{PA}}$ in the language of arithmetic such that,

$$\text{not:-}(\mathsf{PA} \vdash \mathsf{G}_{\mathsf{PA}}) \qquad \text{and} \qquad \text{not:-}(\mathsf{PA} \vdash \mathsf{Con}_{\mathsf{PA}})$$

(Indeed, if T is *any* recursively axiomatized sound extension of PA, then there are arithmetical truths which T does not imply). Now focus on $\mathsf{Con}_{\mathsf{PA}}$ (that is, the formula $\neg Prov_{\mathsf{PA}}(\ulcorner 0 \neq 0 \urcorner)$ of L, the language of arithmetic). In its "oblique" reading, this formula "asserts" that a contradiction is not derivable in PA, and thus that PA is consistent. Gödel's Second Incompleteness Theorem tells us that $\mathsf{Con}_{\mathsf{PA}}$ is *not* a theorem of PA[229] (of course, PA *is* consistent).

Now consider axiomatic set theory. It is possible to prove in axiomatic set theory facts about models. In particular, one can prove that PA has a model (and so is consistent), that PA is true in the standard model $\mathfrak{N}$, and even that any model of PA^2 is isomorphic to $\mathfrak{N}$. Thus,

$$\mathsf{ZFC} \vdash \text{"There is a model of PA"}$$

---

[228] N.B. The reader should be aware that Field actually advocates neither approach as basic. Instead, in recent papers he turns to *modality*, introducing a (non-essentialist) *primitive modal operator* $\Diamond$ to analyse logical notions. Roughly: φ implies χ just in case $\neg\Diamond(\varphi \wedge \neg\chi)$ is (disquotationally) true.

[229] **Proof**. Suppose $\mathsf{PA} \vdash \mathsf{Con}_{\mathsf{PA}}$. So, $\mathsf{PA} \vdash (0 \neq 0) \vee \neg Prov(\ulcorner 0 \neq 0 \urcorner)$. Thus, $\mathsf{PA} \vdash Prov(\ulcorner 0 \neq 0 \urcorner) \rightarrow (0 \neq 0)$. Thus, by Löb's Theorem, $\mathsf{PA} \vdash 0 \neq 0$. So, PA is inconsistent. ■

ZFC $\vdash$ "There is a (unique up to isomorphism) model of PA^2"

Thus,

ZFC $\vdash Con_{PA}$

To summarize, adding axiomatic set theory to axiomatic arithmetic allows one to derive *more true theorems* about natural numbers. This is non-conservativeness. The situation is quite general. Standard axiomatic set theory ZFC is not conservative over weaker axiomatic theories, and ZFC itself, being incomplete of course, can be non-conservatively strengthened by adding further "higher-level" set-theoretical axioms (e.g., about the existence of inaccessible cardinals). These new axioms have the power to settle undecidable assertions "lower down":

> Because of this phenomenon of *non-conservativeness* of richer mathematical theories with respect to lower-level theories, there arises the prospect of justifying the richer theories indirectly in virtue of their power to decide questions at the lower (more "observational") level that otherwise would remain undecided. It was just this prospect that led Gödel to some of his well-known speculations concerning the possible justification of some strong axioms of infinity for set theory.
>
> (Hellman 1989, p. 121).

Gödel fully recognized these consequences of his Incompleteness Theorems, and urged the importance of such considerations in attempted resolutions of presently undecidable assertions:

> First of all the axioms of set theory by no means form a system closed in itself, but, quite on the contrary, the very concept of set on which they are based suggests their extension by new axioms which assert the existence of still further iterations of the operation 'set of'. These axioms can be formulated also as propositions asserting the existence of very great cardinal numbers. ... The simplest of these strong "axioms of infinity" asserts the existence of inaccessible numbers > $\aleph_0$. ... These axioms show clearly, not only that the axiomatic system of set theory as used today is incomplete, but also that it can be supplemented without arbitrariness by new axioms which only unfold the content of the concept of set explained above.
>
> It can be proved that these axioms also have consequences far outside the domain of very great transfinite numbers, which is their immediate subject matter: each of them, under the assumption of its consistency, can be shown to increase the number of decidable propositions even in the field of Diophantine equations.
>
> (Gödel 1947+1964 (1983), pp. 476-477).

Indeed, we have invoked such a non-conservativeness property of a Tarskian truth theory in Chapter 5 above to argue against the adequacy of deflationary truth theories. Tarski's theory of truth/satisfaction TS is such that,

$$\mathsf{PA} \cup \mathsf{TS} \vdash \mathsf{G}_{\mathsf{PA}} \quad \text{and} \quad \mathsf{PA} \cup \mathsf{TS} \vdash \mathsf{Con}_{\mathsf{PA}}$$

That is, PA ∪ TS is not a conservative extension of PA.

Many technical results about the non-conservativeness of axiomatic set theory, and the non-conservativeness of various independent new axioms for set theory[230], are well-known.

It transpires that these results about non-conservativeness strike a direct hit on Field's deflationism. When these results are applied to Field's geometrical nominalism, two central facts emerge:

i. adding *mathematics* to an axiomatic non-mathematical theory N may be deductively non-conservative (and thus, if N is true, then abstract mathematics is deductively indispensable for deducing the *true* (semantic) consequences of N); and,

ii. standard (set-theoretical) space-time theory cannot possess first-order "natural" nominalistic sub-theories over which it is a conservative extension (in short, a nominalistic replacement N for such platonistic mathematicized theories simply do not exist).

In the light of these facts, the whole motivation of the deflationist programme disintegrates. In particular, the analysis of the *utility* of mathematics (i.e., it *dispensably* helps to derive nominalistic conclusions from nominalistic premises) is wrong, and the central dispensability claim (i.e., that mathematics is *dispensable* from our mathematicized theories of Nature) is untenable.

---

[230] Gödel 1947+1964 (1983) points out several consequences of Cantor's Continuum Hypothesis.

9.6 Non-Conservativeness: The Demise of Deflationism

From now on, let N be the second-order nominalistic theory developed by Field in his 1980. Consider N°, the "first-order weakening" of N. That is, the result of reformulating N in first-order mereology, using the "part-of" relation, and replacing any second-order axioms by first-order schemes.

i. Second-Order Deflationism

It transpires, for logical reasons, that Field's conservativeness claim with respect to the second-order theory N is true on a semantical reading but false on a deductive reading. (Indeed, semantical conservativeness follows from nothing more than the categoricity of N and its relative consistency with set theory). It turns out that this has considerable impact on the central destructive claims of the deflationist. In particular, the crucial theses analysing *utility* of false-but-conservative mathematics, and asserting the *dispensability* of mathematics are undermined.

These basic problems with Field's programme were first explained in Shapiro 1983a. This paper develops some Gödelian ideas actually mentioned[231] in the final Chapter of Field 1980.

Using Shapiro's notation, let S be standard applicable axiomatic set theory. Shapiro argues (correctly, I think) that, as a nominalist, Field ought to require a *deductive* conservativeness metatheorem:

(*) if $N \cup S \vdash \varphi$, then $N \vdash \varphi$

After all, the second-order logical consequence relation is not even recursively enumerable (this relation $\models$ thoroughly transcends the limits of finite derivability). A second-order theory T (say, PA^2) may have a logical consequence φ (a Gödel sentence, say, the consistency of PA^2) which is not *deducible* in T. In this case, even though $T \models \varphi$,

[231] The ideas originated from comments made to Field (*before* writing his 1980) by John Burgess and Yiannis Moschovakis.

it requires full-blown mathematics to *prove* that T $\models \varphi$[232]. To be sure, if N is a first-order axiomatic theory, or if any axiom schemes in N are treated as *lists*, rather than as *rules* (for this terminology, see Burgess & Rosen 1997), then the result (*) is true, because of Gödel's Completeness Theorem (for first-order logic, $\models$ and $\vdash$ are co-extensive).

However, many interesting theories N (in particular, the geometrical theory employed by Field 1980) include *axiom schemes* (e.g., a scheme asserting the existence of space-time regions satisfying any arbitrary non-mathematically expressible condition). The natural formulation of N is second-order and, in this case, the deductive conservativeness theorem actually *fails*.

We need to suppose two things:

i. One may prove in S that N is consistent;

ii. There is an interpretation of axiomatic Peano Arithmetic within N.

Both assumptions are fulfilled. After all, the Representation Theorem for N says, in effect, that the reals R *form a model* of N: the construction of such a model of N can, of course, be formalized within axiomatic set theory. So, S can prove the consistency of N. (Indeed, S can prove that any pair of models of N are have cardinality $\mathfrak{c}$). The second assumption, again, is fulfilled. Indeed, Field himself shows in his 1980 that it is possible to define a formula $\psi(R, p)$ meaning intuitively '*p* is an end-point of a region *R* of equally spaced space-time points'.

In fact, N proves the non-mathematical assertion $\exists R \exists p \psi(R, p)$, saying there is such a region. Interpreting '0' as *p* and the successor predicate '*s*' as the next-point-in-the-region-relation, we can model Peano Arithmetic within N. This is how Shapiro sketches the construction:

232 In first-order logic, if T $\models \varphi$, then T $\vdash \varphi$, and thus, by the very explanation of deducibility $\vdash$, a finite subset of T $\cup \{\neg\varphi\}$ generates a formal contradiction (using a tree say). In second-order logic, Completeness fails, so *even if* T $\models \varphi$, it needn't follow that T $\vdash \varphi$; and thus it needn't follow that T $\cup \{\neg\varphi\}$ is inconsistent (and thus formally refutable). E.g., $PA^2 \models Con_{PA}$ but $PA^2 \cup \{\neg Con_{PA}\}$ is not *formally* inconsistent (although it is unsatisfiable). In this case, in order to *prove* that T $\models \varphi$ one requires *substantial* mathematical assumptions (not just symbol manipulation).

> Field indicates the definition in N of a spatio-temporally equally spaced region, a class of (discrete) points all of which lie on a single straight line and such that the distance between adjacent points is uniform. Let $\psi(R, p)$ be a formula equivalent to 'R is an infinite, spatio-temporally equally spaced region containing p as end-point'. We have $\mathsf{N} \vdash \exists R \exists p \psi(R, p)$. Such pairs (R, p) are models of the natural numbers. Relative to R and p, one can construct a formula $\Sigma(x, y)$ equivalent to 'x and y are both in R, there is no point in R strictly between x and y, and if $x \neq p$, then x is strictly between p and y'. The formula represents the 'successor relation' of (R, p). The analogues of addition and multiplication can also be defined. It follows from N that if $\psi(R, p)$ holds, then (R, p) under Σ satisfies the axioms of Peano arithmetic.
>
> (Shapiro 1983a (1996), p. 230)[233].

This means that we can formulate within the language of N a Gödel sentence $\mathsf{Con_N}$ expressing the consistency of N. Since N is an *axiomatic* (consistent) theory, Gödel's Second Incompleteness Theorem applies, so,

a. $\mathsf{Con_N}$ is *not* a theorem of N.

But we agreed at the outset that,

b. $\mathsf{Con_N}$ *is* a theorem of $\mathsf{S} \cup \mathsf{N}$.

Hence, $\mathsf{N} \cup \mathsf{S}$ is not a deductively conservative extension of N.

Let me emphasize. There are "non-mathematical" geometrical logical consequences of N which are *not* derivable in N, but which *are* derivable in $\mathsf{N} \cup \mathsf{S}$. It seems that the set-theoretical mathematics involved in $\mathsf{N} \cup \mathsf{S}$ is *not dispensable*. It is indispensable for deriving these physical facts about space-time:

> Since (presumably) N is recursively axiomatized, Gödel's incompleteness theorems apply. Let $\mathsf{Con_N}(R, p)$ be the formula asserting the 'consistency' of N in terms of the points of R (analogous to the usual formulation using the natural numbers). Finally, let θ be the sentence:
>
> $\forall R \forall p[\psi(R, p) \rightarrow \mathsf{Con_N}(R, p)]$.

---

[233] For completeness here are the definitions (see Shapiro 1983a, footnotes 8 and 9):

(I) R is an spatio-temporally equally spaced region: "any three points in R are co-linear and for every point x in R which lies strictly between two points of R, there are points y and z in R such that (a) exactly one point in R lies strictly between y and z and that point is x, and (b) the distance between y and x is equal to the distance between x and z".

(II) $\psi(R, p)$: "R is a spatio-temporally equally spaced region, p is in R, p does not lie strictly between any two points in R, and, for every x in R, if $x \neq p$, then there is a point y in R such that x lies strictly between p and y".

The relevant version of the incompleteness theorem entails that not:-(N $\vdash \theta$). Notice that θ is sentence in the language of nominalistic physics. Its variables range over space-time points and regions.

A second presumption is that the consistency of N is provable in the set theory S. Let $Con_N(\omega, 0)$ be the usual sentence asserting the "consistency" of N in terms of the natural numbers. We have S $\vdash Con_N(\omega, 0)$. Notice that it is provable in N + S that the formula $\psi(R, p)$ is "categorical" and that any pair (R, p) satisfying ψ is isomorphic to the natural numbers $(\omega, 0)$. The latter follows from the existence of a representing homomorphism from the points of space-time to R^4. Thus, we have

$$N + S \vdash \forall R \forall p[\psi(R, p) \rightarrow (Con_N(\omega, 0) \leftrightarrow Con_N(R, p))].$$

Hence, N + S $\vdash \theta$. This refutes the deductive conservativeness of S over N.

(Shapiro 1983a (1996), p. 230)[234].

Shapiro further comments on the *semantical conservativeness* of set theory over N:

A few brief remarks may be in order. The preceding result shows that for second-order theories, deductive conservativeness is not coextensive with semantic conservativeness. The gap can be substantial. Notice that if T is any categorical theory and S + T is satisfiable, then S is semantically conservative over T. Thus, for example, set theory is semantically conservative over second-order arithmetic. As is well known, however, in set theory one can deduce many arithmetical statements that are not theorems of arithmetic alone.

(Shapiro 1983a (1996), p. 231).

ii. First-Order Deflationism

Turning to the first-order theory N°, the crucial changes are these. First, any formula with a monadic second-order variable, e.g., $R(x)$, is to be replaced by a first-order mereological formula $x \subseteq r$, meaning "*x* is a part of the region *r*". Second, geometrical second-order axioms are replaced by axiom schemes.

The *general* non-conservativeness of mathematics is established, as above, via Gödel's Theorems. But didn't Field show in the Appendix to Chapter 1 of his 1980, via an Expansion Lemma, that set theory is *semantically* conservative over first-order nominalistic theories? Well, yes and no. In the situation where it does hold, it follows, by Gödel's Completeness Theorem for first-order logic, that if N is a first-order theory, then

[234] The reprint in Hart (ed.) 1996 contains a typographical mistake, omitting the remainder of the sentence after the appearance of '$Con_N(\omega, 0)$', and the following sentence.

the $N \cup S$, the simple *union* of the axioms of N and those of S, is also a deductively conservative extension of N. But now an important subtlety intrudes.

There is a crucial ambiguity involved in "adding" a pair of theories, where one theory is formulated in an *expanded* language (i.e., the extra terminology in our case is mathematical, namely $\in$). Normally, by "axiom" of a combined theory $T_1 \cup T_2$ we simply take the *union* of all the axioms of T_1 and T_2. However, if T_1 contains *axiom schemes*, are we to count the occurrence of new terminology in such axiom schemes as axioms of the combined theory? It turns out that it makes a difference to what can be proved.

From now on, let us write,

$T_1 \cup T_2$: the result of simply pooling all the axioms of T_1 and T_2.

And let us write,

$T_1 + T_2$: the result of simply pooling all the axioms of T_1 and T_2, *and* taking as additional axioms, any that arise from axiom schemes of T_1 by including formulas containing the new terminology.[235]

In the terminology of Burgess & Rosen 1997, $T_1 \cup T_2$ is the result of treating these schemes as *lists*, and $T_1 + T_2$ is the result of treating these axiom schemes as *rules*.

The Expansion Lemma proved in the Appendix to Chapter 1 of Field 1980 remains invariant under both proposals, since it does not concern any particular *theory* in L°. However, in applying the Expansion Lemma, we want to show that, for any model of some particular non-mathematical theory N, there is an expansion which is a model of $N \cup ZFU$. This, again, is fine. But is it true that, for any model of N, there is an expansion which is a model of N + ZFU? The answer is "Not necessarily". In short, although, any interpretation of L° may still be expanded to a model of axiomatic set theory ZFU itself, it does not follow that, for any *nominalistic theory* N, any model of N may be expanded

[235] N.B. This is implicitly what happens when T_1 is a *second-order theory* containing a second-order axiom $\forall X[...X...]$. In deriving theorems within $T_1 \cup T_2$, one permits formulas expressed in T_2 vocabulary as instances of the predicate variable X.

to a model of N + ZFU. Indeed, this cannot be true, for we can show that N + ZFU is *not* in general a conservative extension of N.

This is how Burgess & Rosen report the technical situation:

> The technical issues in part turn on a subtle distinction pertaining to axiom schemes. When a theory T in a language L involves a scheme, we have said that we conceive of the scheme as a rule to the effect that:
>
> (i) for every formula Q, ---Q--- is an axiom
>
> An alternative would be to conceive of it as a list of all the axioms:
>
> (ii) ---Q_0---, ---Q_1---, ---Q_2---, . . .
>
> that result when this rule is applied to all formulas:
>
> (iii) Q_0, Q_1, Q_2, . . .
>
> of the language L. The difference in conception makes a difference only when the theory T in the language L is extended to or incorporated in some stronger theory T′ in some richer language L′ which will have new formulas Q′ not on the list (iii), and for each such formula a new formula ---Q′--- not on the list (ii). To conceive of the scheme as a rule (i) means that these new formulas ---Q′--- are taken as new axioms; to conceive of the scheme as a list (ii) means that they are not.
>
> (Burgess & Rosen 1997, pp. 194-195).

With this in mind, consider N°, the *first-order* theory obtained by weakening all second-order axioms in N to axiom schemes. Again, it is possible to show that the standard platonistic theory of flat space-time P is *still not conservative* over N°. Shapiro 1983a explained why. The argument is proof-theoretic. Again, N° has a gödel sentence, which can be proved in P.

We assume that P is equivalent to N° + S, obtained by applying the set theory to the nominalistic theory N°. In this case, the undecidable assertion is still $Con_{N°}$. Again, by Gödel's Theorem, N° does not imply the Gödel sentence $Con_{N°}$. However, the platonistic space-time theory N° + S does prove $Con_{N°}$, as long as axiom schemes are expanded. In short, N° + S is again *not* a conservative extension of N°.

Burgess & Rosen 1997 summarize the situation:

> Returning to the description of Field's strategy, his intermediate theory may be described as follows. The synthetic theory T° involves a scheme of continuity. The intermediate theory $T^{\dagger\circ}$ Field considers is the result of adding the apparatus of set theory while treating that scheme as a list (and contrasts with $T^{\ddagger\circ}$ considered in article II.A.5.b, the result of adding the apparatus of set theory

> while treating the scheme as a rule). ... More is provable from the standard axioms of set theory about the real numbers as standardly reconstructed in set theory than is provable about real numbers in analysis. ...
>
> Field shows that $T^{\dagger\circ}$ is conservative over T° (as $T^{\ddagger\circ}$ is not, since adding set theory gives new geometric results, and adding proposed further axioms beyond standard set theory gives further geometric results, with the majority and minority proposals leading to incompatible geometric results). In fact, adding the apparatus of set theory to any theory produces a conservative extension, provided any and all schemes in the original theory are treated as lists, not rules.
>
> (Burgess & Rosen 1997, p. 195-196).

That is, in my notation,

i. $\mathsf{N}^{\circ} \cup \mathsf{S}$ is a conservative extension of N°, but

ii. $\mathsf{N}^{\circ} + \mathsf{S}$ is *not* a conservative extension of N°.

This proof theoretic result has an important model-theoretic consequence. The theory P asserts that there is an isomorphism from the set of space-time points to $\mathbf{R}^4$. Now, $\mathsf{N}^{\circ} + \mathsf{S}$ is not conservative over N°, so there must be non-standard models of N° that *cannot* be expanded to models of $\mathsf{N}^{\circ} + \mathsf{S}$. Indeed, since N° has countable models, it is impossible to prove the sort of Representation Theorem (that Field requires) to the effect that there is a representing homomorphism from any model of N° to the reals **R**.[236]

9.7 The Indispensability of Mathematics

What are we to make of this? The above argument suggest that mathematics is indispensable in deducing facts about space-time (that is, in deducing mathematics-free theorem about space-time that logically follow from currently accepted space-time theory). For any sufficiently powerful axiomatic platonistic theory P, a proposed mathematics-free replacement N must lose some mathematics-free theorems. In general, N is not "just as good" as the platonistic theory P which it is meant to replace.

In our introduction to Field's deflationary programme, we highlighted three major claims about mathematics:

[236] See Field 1985 (1989), p. 131: "It is immediately clear that if the extended representation theorem ... is to hold, N cannot be formulated in first-order logic".

The Conservativeness Claim

> The result of adding standard mathematics M to a body N of non-mathematical nominalistic assertions is a conservative extension.

The Utility of Mathematics

> Any inference from nominalistic premises to a nominalistic conclusion that can be made with the help of mathematics could be made (usually more long-windedly) without it.

Eliminability of Mathematics in Science

> For each platonistic mathematicized theory of Nature P, there is a "nice" nominalistic theory N axiomatized using finitely many primitives such that P is a deductively conservative extension of N.

We now see that each claim is false. Adding mathematics is not conservative. There are bodies of nominalistic assertions such that adding mathematics yield new nominalistic conclusions. So, the Conservativeness Claim is false. This implies that, for some inferences between nominalistic statements involving mathematics, the mathematics is not dispensable. So, the Utility Claim is false. Finally, there exist platonistic theories incorporating set theory for which there do not exist "nice" nominalistic replacements. So, the Eliminability Claim is false.

There are two further points to be made:

i. Perhaps, the undecidable consequences of N are somehow "recherché": Field suggests that, although such consequences are nominalistic witnesses to non-conservativeness, they are perhaps not "physically significant". In short, the sort of Gödelian assertions undecided by N, but provable in N ∪ S, simply *don't matter.*

ii. Standard axiomatic platonistic space-time theory P is itself "Gödelizable", and *its* consistency assertion, Con_P is likewise not a theorem of P. Furthermore, P itself still cannot be "nominalistically complete". There will always be sentences in the nominalistic notation that are undecided, no matter how many mathematical axioms are added (in an effective way)

As for point (i), Shapiro himself notes:

> The sentence θ, taken as a statement about space-time points and regions, is rather obscure—it is not likely to form an essential part of the account of any phenomenon to be explained by physics. It may be the case that every interesting or scientifically relevant theorem of N + S is a theorem of N. ... The latter, however, remains to be shown (provided that a notion of 'scientific relevance' can be formulated).
>
> (Shapiro 1983a (1996), p. 231)

However, this is an exact replica of the debate *within the philosophy of pure mathematics* concerning the significance of non-conservativeness. On this topic, it has emerged recently that a whole series of important mathematical facts are undecided by certain weak theories: there are important facts about the natural numbers, expressible in PA but not decided by PA. The most famous example is the demonstration by Paris & Harrington 1977 of the unprovability of a variant of the finite Ramsey Theorem in PA. As Isaacson 1987 puts it,

> Attention has focused in recent years on some extremely interesting examples of arithmetical truths unprovable in Peano arithmetic which are thought of as much more genuinely and purely mathematical than the Gödel sentences. They include the study by Kirby and Paris of Goodstein's Theorem, the Paris-Harrington variant of the finite Ramsey theorem, and Friedman's finite version of Kruskal's theorem.
>
> (Isaacson 1987 (1996), p. 215).

Indeed, there are important facts about the real numbers not decided by low-level set theories which become decidable by extending the set theory,

> A whole series of remarkable results along similar lines [to Paris-Harrington] has been obtained by Harvey Friedman, in which the lower-level set theory is already some axiomatic set theory (e.g., some natural modification of Z or ZF) and the higher-level set theory is a richer set theory (e.g., ZFC or ZFC + a large

cardinal axiom, etc.). But the statement in question is not an "esoteric" set-theoretical statement, but a statement pertaining to functions of sort encountered in "normal" mathematics "lower down" (e.g., Borel functions from $R^n \to R$).

(Hellman 1989, p. 122).

There may be important geometrical facts about (flat) space-time expressible in axiomatic synthetic geometry which may be settled by adding strong enough set-theoretical axioms. (See Burgess & Rosen, pp. 120-123).

As for (ii), this does not lessen the failure of the deflationist programme: no-one expects that standard axiomatic theories should be *deductively complete* (with respect to the "physical" facts), and it is an intriguing fact that decisions about the structure of space-time may depend upon extensions of the so-called "standard framework" for mathematics (axiomatic Zermelo-Fraenkel set theory, ZFC).

Hellman makes a related point,

> ... granted that as much mathematics as can be formulated within Z^+ is to be included in a framework for applied mathematics, is this really enough? Might we not require, for example, the full power of ZF, or ZFC, or ZFC plus large cardinal axioms? Just because we haven't encountered the need for such stronger axioms yet doesn't mean that we won't, much less that we couldn't. ...
>
> Now it may seem on the face of it outlandish that large cardinal axioms, or even the Axiom of Replacement, should ever be needed in scientific applications. However, remarkable recent investigations of Harvey Friedman suggest that the idea is not so outlandish after all.

(Hellman 1989, p. 120).

This puts the Indispensability Argument in a new light. Although Quine and Putnam themselves never argued this way using such Gödelian considerations, it certainly seems to have occurred to Gödel:

> ... even disregarding the intrinsic necessity of some new axiom, and even in case it has no intrinsic necessity at all, a probable decision concerning its truth is possible in another way, namely inductively by studying its "success". Success here means fruitfulness in consequences, in particular, in "verifiable" consequences, i.e., consequences demonstrable without the new axiom, whose proofs with the help of the new axiom, however, are considerably simpler and easier to discover, and make it possible to contract into one proof many different proofs. ... There might exist axioms so abundant in their verifiable consequences, shedding so much light upon a whole field, and yielding such powerful methods for solving problems (and even solving them constructively, as far as that is possible) that no matter whether or not they are intrinsically necessary, they would have to be accepted at least in the same sense as any well-established physical theory.

(Gödel 1947+1964 (1983), p. 477).

The central claim of Field's deflationism is that mathematics is "insubstantial"[237], adding no new physical content to any purely nominalistic theory. The Gödelian argument explained by Shapiro shows that this claim is false (if arithmetic can be modelled in the nominalistic theory).

Similarly, if a mathematicized scientific theory P can be represented as N + ZFC, where N is an underlying "mathematics-free" theory, then *in some important cases* N is not "just as good" as P, because P will have "mathematics-free" theorems that N does not have.

This concludes our discussion of deflationism about mathematics. Mathematics is "substantial" in the sense that mathematics is indispensable to our scientific description of the world[238].

---

[237] Just as deflationism about truth claims that "truth is insubstantial", adding no new content. I argued against this claim in Chapter 5.

[238] I have ignored the natural question of just "how much" mathematics is indispensable to science. The reason is that I am sceptical that any clear-cut answer can be given. In contrast, Burgess 1984 (p. 386) claims that *mathematical analysis* is "probably sufficient to develop, making use of coding devices, all the mathematics that has found scientific applications up to the present". But anyone familiar with the standard mathematics used in modern theoretical physics (e.g., differential geometry, topology, cohomology, fibre bundle theory, symplectic manifold theory, dimensional regularization in non-integral (!) dimensional spacetime, infinite-dimensional path integrals, and so on) will treat *any* view that all this machinery can be gotten down to *mathematical analysis* with extreme scepticism. For example, the "phase space" Ω for a scalar field Φ on spacetime contains *all* functions $f: \mathbf{R}^4 \rightarrow \mathbf{R}$. This phase space is larger than the continuum.

Quine discusses this matter briefly in his reply to Charles Parsons (Quine 1986), and Hellman discusses the matter in some detail in Hellman 1989, Chapter 3, §2.

CHAPTER 10

Conclusion: Mathematics and Truth

Renouncing classes means rejecting mathematics. That will not do. Mathematics is an established going concern. Philosophy is as shaky as can be. To reject mathematics on philosophical grounds would be absurd. ... I laugh to think how presumptuous it would be to reject mathematics for philosophical reasons.

David Lewis 1991, *Parts of Classes*, §2.8.

I find it quite amazing that it is possible to predict what will happen by mathematics, which is simply following rules which really have nothing to do with the original thing.

R.P. Feynman 1965 (1992), *The Character of the Physical Law*, p. 171.

It is positively spooky how the physicist finds the mathematician has been there before him or her.

S. Weinberg 1986, 'Lecture on the Applicability of Mathematics', p. 725.

10.1 Science Entails Platonism

The burden of Chapters 1, 2 and 3 was to clarify how mathematics is integrated within scientific theories of the world, and how this integration entails ontological commitment to mathematical entities. In short, "Science Entails Platonism".

The point is, in a sense, quite trivial, as soon as one is clearly about what it means for an interpreted theory to be *true*. A theory T is true if and only if the entities it quantifies over—the range of its bound quantificational variables—are related exactly as the theorems of T say they are. This remains the case whether T is a nominalistic theory, or a pure mathematical theory, or a mixed "impure" mathematicized theory of Nature. Part of the burden of Chapter 4 was to explain how our modern understanding of the concept of truth, grounded on the so-called T-sentences, but developed within a full Tarskian theory of satisfaction, explicates this informal talk of entities "being related as T says they are".

Chapter 6 defended a Tarskian conception of mathematical truth and criticized attempts to *repudiate* such a Tarskian conception. In particular, Tarski's conception of truth is one which defines truth in terms of the *referential relation* of satisfaction. Loosely speaking, the language and mind-independent facts in the abstract world of numbers, functions, sets and mathematical structures *make* mathematical statements true. *Non*-Tarskian semantics for mathematics is implausible, especially when we turn to the *applications* of mathematics. An assertion is not true *because* it is provable, or *because* it is justifiable, or *because* it has warranted assertibility, or anything like that. 'Tachyons travel faster than light' would be a true sentence if tachyons travelled faster than light. And this proposition would be true even if we *never found out*. After all, Gödel taught us that, for any sound axiomatic theory containing a fragment of arithmetic, there are truths it does not prove. And Tarski taught us that the class of true sentences in any sufficiently rich language is not computable.

10.2 The Possibilities of Nominalism

The various programmes of modern nominalism are primarily motivated by an *epistemological* concern, which we have christened *Benacerrafitis*, the problem of "epistemic access". Roughly, abstract mathematical entities are "unknowable abstractions", causally disconnected from the sensory processes of knowledge acquisition.

The most promising version of nominalism, one that does not immediately collapse when applications are considered[239], is some version of *modalism*, incorporating a substantial modal logical machinery of constructibility quantifiers and actuality operators. Statements about mathematicalia are to be reconstrued as statements about possible-but-non-actual entities, which might have existed or might have been constructed (like Chihara's type-theoretic hierarchy of unactualized open sentence

[239] Like if-thenism, as I argue in Chapter 6. Or simplistic modalizations using just the metalogical $\Diamond$, as I argue in Chapter 8.

tokens). One immediate problem is that many philosophers, like Quine, Putnam and Field, would simply say that this apparatus of modality is *highly anti-nominalistic*.

In any case, such modal apparatus is standardly analysed *using mathematics*, rather than the other way round! If intelligibility is conferred on modal notions by recent technical work in model theory for modal logic, then it is quite unacceptable to pretend that this explanation is just an "elaborate myth", as Chihara claims. Furthermore, an ontology of possibilia (or an ideology of modal notions: '*x* might have been *F*') is just as inaccessible or inexplicable (on the causal empiricist account of knowledge) as is a standard ontology of abstract mathematicalia, like numbers and sets.

I conclude that if anything resembling nominalism is to succeed, it must adopt a modal framework which is every bit as difficult to make epistemological sense of as standard mathematics is. Indeed, if we accept a Quinian *pragmatic holism* as our working epistemology, then (as Hart 1979 pointed out) there is no problem with justifying our acceptance of theories that reify or posit mathematicalia: such theories are conjectures which are justified holistically by their evidential and pragmatic virtues (simplicity, elegance, predictive and explanatory power, and so on). Indeed, they are *already* part of our working theory of the world. It would be odd, to say the least, to suggest that this mathematical ontology (and/or ideology) be replaced by a quite puzzling modal ontology (and/or ideology).

Modal nominalism is thus an inflated version of nominalism. It repudiates an abstract ontology in favour of a resplendent ideology of modal concepts (and even, a resplendent ontology of unactualized possibilia). Furthermore, Benacerraf's argument can be run through for modal knowledge. Thus, given the similar epistemological problems of modal nominalism, what is its conceptual *attraction*? Extreme constructivists apart, mathematics seems to make perfect sense as it is standardly practised. Mathematicians and mathematical scientists do not find anything bizarre about talk of functions of real numbers, fibre bundles, phases spaces, and so on. What would be the advantage of introducing a bizarre *modal* framework, which appears to have been invented in a thoroughly parasitic manner, using *standard* mathematical theories?[240]

10.3 The Substantiality of Truth and Mathematics

In Chapter 5, I argued that the deflationary conception of truth (erected upon a "deflationary truth theory" like DT or MT, constructed simply of the T-sentences) is *inadequate* as a theory of truth. One of the properties of such theories is *deductive conservativeness*, a property the deflationist might aspire to. For if a truth theory is conservative, then anything proved with it can be proved without it. In short, a conservative deflationary truth theory yields an "insubstantial theory of truth".

But the crux of the matter is that an *adequate* theory of truth needs to be *non-conservative*. Our ability to recognize the truth of Gödel sentences is a complete mystery on a conservative truth theory. Indeed, a conservative truth theory cannot satisfy the simple "Equivalence Principle", the constraint that,

i. T + truth theory $\vdash$ 'all theorems of T are true'

In a sense, this is nothing more than a generalization of the constraint that a truth theory be such that,

ii. φ + truth theory $\vdash$ 'φ is a true sentence'

However, the constraint (i) immediately yields the conclusion that the truth theory must be *non-conservative*. The reason is that if T is sufficiently powerful (an extension of Q), it is able to formulate a fixed point (Gödel) sentence G_T such that T implies that G_T is not true if and only if G_T is a theorem of T. From (i), we infer that T + truth theory implies that if G_T is not true, then it is true; so it implies that G_T is true. If the truth theory were conservative, then T would imply G_T. But, Gödel's Incompleteness Theorem says that, if T is consistent, then T does not imply G_T. Thus, a truth theory that satisfies constraint (i) must be non-conservative.

[240] Chihara's modal constructibility theory is a *notational rewrite* of Russellian type theory.

Tarski's theory satisfies the constraint (i), as Tarski himself proved (Tarski 1936). This shows that Tarskian truth is adequate. In short, truth is substantial. It follows that deflationism about truth is false.

Analogously, the (Field-style) deflationist about mathematics asserts that mathematics is insubstantial: standard set theory is allegedly conservative over any internally consistent nominalistic theory. However, the logical results outlined in Chapter 9 about non-conservativeness indicate the *substantiality of mathematics* within (rich enough) scientific theories. The pure and mixed axioms within such a theory T cannot be simply dropped. The non-conservativeness of mathematics results totally undermines perhaps the most central idea of deflationism. Adding mathematics to a non-mathematical theory can increase its deductive power. So mathematics is not dispensable. It is required in theorem-proving. In general, a platonistic mathematicized theory P cannot be represented as N + S, where N is a first-order recursively axiomatized theory and S is axiomatic set theory, while demanding that P be a conservative extension of N[241]. Mathematicized scientific theories are not dispensable. In short, mathematicized platonistic theories cannot in general be replaced, *at zero cost*, by non-mathematical theories.

To summarize—*pace* deflationism—truth and mathematics are both *substantial*. In Quinian terminology, they make substantial contributions to our "over-all scientific system of the world". Nominalism (whether deflationary or modal) seems to offer no real advantages over standard platonistic mathematics. The deflationary version of nominalism, favoured by Field, seems to be riddled with insurmountable technical problems, concerning the basic Gödelian logical facts about non-conservativeness. The modal version, favoured by Chihara, seems to be riddled with epistemological problems concerning modal knowledge which are simply analogues of the problem for standard non-modal mathematics detected by Benacerraf in 1973.

[241] For, in the cases of interest (where N contains enough geometry), the result of applying mathematics, N + S is simply not a conservative extension of N.

As this work has been driven by the ideas of Quine (and to a lesser extent by those of Gödel), let me finish with a final quote from the author of the most promising argument for mathematical realism:

> I have explained early and late that I see no way of meeting the needs of scientific theory, let alone those of everyday discourse, without admitting universals irreducibly into our ontology ... Mathematics, moreover, and applied mathematics at that, is up to its neck in universals; we have to quantify over numbers of all sorts, functions and much else. I have argued that there is no blinking these assumptions; they are as integral to the physical theory that uses them as are the atoms, the electrons, the sticks, for that matter, and the stones. I have inveighed early and late against the ostrichlike failure to recognize these assumptions. ...
>
> Nominalism, ostriches apart, is evidently inadequate to a modern scientific system of the world.
>
> (Quine 1981c (1981a), pp. 182-183).

BIBLIOGRAPHY

ALSTON, WILLIAM 1958. 'Ontological Commitments', *Philosophical Studies* **9**, 8-17. In Benacerraf & Putnam (eds.) 1964.

ARISTOTLE *Metaphysics*. In *The Complete Works of Aristotle: The Revised Oxford Translation*, edited by J. Barnes. Princeton University Press, New Jersey, 1984.

AYER, ALFRED JULES 1936. *Language, Truth and Logic*. Pelican Books, London, 1971.

BELL, ERIC TEMPLE 1952. *Mathematics: Queen and Servant of Science*. G. Bell & Sons Ltd. London.

BENACERRAF, PAUL 1965. 'What Numbers Could Not Be', *Philosophical Review* **74**. In Benacerraf & Putnam 1983.

— 1973. 'Mathematical Truth', *Journal of Philosophy* **70**. In Benacerraf & Putnam (eds.) 1983 and in Hart (ed.) 1996.

— 1996. 'What Mathematical Truth Could Not Be - I'. In A. Morton & S. Stich (eds.) 1996, *Benacerraf and His Critics*. Blackwell, Oxford

BENACERRAF, PAUL & PUTNAM, HILARY (eds.) 1964. *Philosophy of Mathematics: Selected Readings*. Prentice-Hall, Englewood Cliffs, New Jersey.

BENACERRAF, PAUL & PUTNAM, HILARY (eds.) 1983. *Philosophy of Mathematics: Selected Readings*. Revised second edition. Cambridge University Press, Cambridge.

BERNAYS, PAUL 1935. 'On Platonism in Mathematics', in *L'Enseignement Mathématique* **34**, 52-69. Translated by C.D. Parsons. In Benacerraf & Putnam (eds.) 1983.

BETH, E.W. 1953. 'On Padoa's Method in the Theory of Definition', *Indagationes Mathematicae* **XV**, 330-339.

BIRKHOFF, G. & MACLANE, S. 1965. *A Survey of Modern Algebra*. Third edition, Macmillan, New York.

BLACKBURN, SIMON 1983. *Spreading the Word: Groundings in the Philosophy of Language*. Oxford University Press, Oxford.

BOOLOS, GEORGE 1971. 'The Iterative Conception of Set', *Journal of Philosophy* **68**. In Benacerraf & Putnam (eds.) 1983.

— 1984. 'To Be is To Be the Value of a Variable (or To Be Some Values of Some Variables)', *Journal of Philosophy* **81**, 430-449.

— 1985. 'Nominalist Platonism', *Philosophical Review* **94**, 327-344.

— 1987. 'The Consistency of Frege's *Foundations of Arithmetic*', in Hart (ed.) 1996.

— 1993. *The Logic of Provability*. Cambridge University Press, Cambridge.

BOOLOS, GEORGE & JEFFREY, RICHARD. 1989, *Computability and Logic.* Revised third edition. Cambridge University Press, Cambridge.

BOYER, CARL 1968. *A History of Mathematics.* Wiley, New York.

BURGESS, J.P. 1983. 'Why I am Not a Nominalist', *Notre Dame Journal of Formal Logic* **24**, 93-103.

— 1984. 'Synthetic Mechanics', *Journal of Philosophical Logic* **13**, 379-395.

BURGESS, J.P. & ROSEN, GIDEON 1997. *A Subject With No Object. Strategies for Nominalistic Interpretation of Mathematics.* Clarendon Press, Oxford.

CARNAP, RUDOLF 1950. 'Empiricism, Semantics and Ontology', *Revue Internationale de Philosophie* **4**, 20-40. In Benacerraf and Putnam (eds.) 1983.

CARTWRIGHT, NANCY 1983. *How The Laws of Physics Lie.* Clarendon Press, Oxford.

CHIHARA, CHARLES 1973. *Ontology and the Vicious Circle Principle.* Cornell University Press, Ithaca.

— 1990. *Constructibility and Mathematical Existence.* Oxford University Press, Oxford.

CHOQUET-BRUHAT, Y., DILLARD-BLEICK, M., & DEWITT-MORETTE, C. 1977 *Analysis, Manifolds and Physics.* North-Holland Physics Publishing, Amsterdam. Revised edition, 1982.

CHURCH, ALONZO 1958a. 'Ontological Commitment', *Journal of Philosophy* **55**, 1008-14.

— 1958b. Letter to Goodman (1 Dec 1958). In Goodman 1972.

CORCORAN, J. 1997. Entry on Tarski. *Cambridge Dictionary of Philosophy.* Edited by R. Audi. Cambridge University Press.

CRAIG, W. 1953. 'On Axiomatizability Within a System', *Journal of Symbolic Logic* **18**, 30-32.

DAVID, MARIAN 1994. *Correspondence and Disquotation: An Essay on the Nature of Truth.* Oxford University Press, Oxford.

DAVIDSON, DONALD 1967. 'Truth and Meaning'. In Davidson 1984.

— 1969. 'True to the Facts', *Journal of Philosophy* **66**, 748-764. In Davidson 1984.

— 1973. 'In Defence of Convention T', in H. Leblanc (ed.), *Truth, Syntax and Modality*, North Holland. In Davidson 1984.

— 1977. 'Reality Without Reference', *Dialectica* **31**, 247-253. In Davidson 1984.

— 1984. *Inquiries into Truth and Interpretation.* Clarendon Press, Oxford.

DAVIS, PHILIP J. & HERSH, REUBEN 1980. *The Mathematical Experience.* Paperback edition, Penguin Books, 1990.

DEDEKIND, RICHARD 1888. *Was sind und was sollen Die Zahlen*? Vieweg, Brunswick. Translation, *The Nature and Meaning of Numbers*, by W.W. Beman in Dedekind 1901, *Essays on the Theory of Numbers*, Open Court, Chicago. Paperback edition, Dover, New York, 1963.

DRAKE, F.R. & SINGH, D. 1996. *Intermediate Set Theory*. John Wiley & Sons, Chichester.

DUMMETT MICHAEL 1991. *The Logical Basis of Metaphysics*. Harvard University Press, Cambridge, Ma.

EARMAN, JOHN 1995. *Bangs, Crunches, Whimpers and Shrieks*. Oxford University Press, Oxford.

EARMAN, JOHN & NORTON, JOHN D. 1996. 'Infinite Pains: The Trouble With Supertasks'. In Morton & Stich (eds.) 1996.

EDDINGTON, SIR ARTHUR S. 1928. *The Nature of the Physical World*. Cambridge University Press, Cambridge.

EINSTEIN, A., LORENTZ, H.A, MINKOWSKI, H. & WEYL, H. 1952. *The Principle of Relativity*. Edited by A. Sommerfeld. Dover.

EISBERG, R.M. 1961. *Fundamentals of Modern Physics*. Wiley, New York.

ELLIS, BRIAN 1985. 'What Science Aims to Do', in P. Churchland and C. Hooker (eds.) *Images of Science*, University of Chicago Press, Chicago. In Papineau (ed.) 1996.

EUCLID *The Thirteen Books of Euclid's Elements*. Translated by T.L Heath. Paperback edition. Dover, New York, 1956.

FEYNMAN, RICHARD 1965. *The Character of the Physical Law*. Penguin Books, London, 1992.

FIELD, HARTRY H. 1972. 'Tarski's Theory of Truth', *Journal of Philosophy* **69**, 347-375.

— 1980. *Science Without Numbers: A Defence of Nominalism*. Princeton University Press, Princeton, N.J.

— 1982. 'Realism and Anti-Realism About Mathematics', *Philosophical Topics* **13**, 45-69. In Field 1989.

— 1984a. 'Is Mathematical Knowledge Just Logical Knowledge?', *Philosophical Review* **93**, 509-552. Revised version in Field 1989.

— 1984b. 'Review of Wright 1983', *Canadian Journal of Philosophy* **14**, 637-662. Appeared under title 'Platonism for Cheap? Crispin Wright on Frege's Context Principle', in Field 1989.

— 1985. 'On Conservativeness and Incompleteness'. *Journal of Philosophy* **81**, 239-260. In Field 1989.

— 1986. 'The Deflationary Conception of Truth', in G. MacDonald and C. Wright (eds.) 1986, *Fact, Science and Value: Essays on A.J. Ayer's Language Truth and Logic*. Blackwell, Oxford.

— 1988. 'Realism, Mathematics and Modality', *Philosophical Topics* **19**, 57-107. In Field 1989.

— 1989. *Realism, Mathematics and Modality*. Blackwell, Oxford.

FINE, ARTHUR 1984. 'The Natural Ontological Attitude', in J. Leplin (ed.) 1984, *Scientific Realism*, University of California Press, Berkeley. In Papineau (ed.) 1996.

FLAVELL, J. 1985. *Cognitive Development*. Prentice-Hall, Englewood Cliffs, New Jersey.

VAN FRAASSEN, B.C. 1976. 'To Save the Phenomena', *Journal of Philosophy* **73**, 723-32. In Papineau (ed.) 1996.

— 1980. *The Scientific Image*. Clarendon Press, Oxford.

FREGE, GOTTLOB 1884. *Die Grundlagen der Arithmetic*. Koebner, Breslau. Translation by J.L. Austin as *The Foundations of Arithmetic*, Blackwell, Oxford, 1950.

— 1892. 'Über Sinn und Bedeutung', *Zeitschrift für Philosophie und philosophische Kritik* **100**. Translated as 'On Sense and Meaning', in Geach and Black (eds.) 1980. Also, translated as 'On Sense and Reference', in A.W. Moore (ed.) 1993, *Meaning and Reference*, Oxford University Press, Oxford.

— 1893. *Grundgesetze der Arithmetic*, Vol. i., Pohle, Jena.

— 1903. *Grundgesetze der Arithmetic*, Vol. ii., Pohle, Jena.

FRIEDMAN, MICHAEL 1983. *Foundations of Space-Time Theories*. Princeton University Press, Princeton, New Jersey.

GALILEO GALILEI 1618. *The Assayer*. In Galileo 1960, *The Controversy of the Comets in 1618*, translated by S. Drake & C.D. O'Malley, University of Philadelphia Press, Philadelphia.

GARDNER, MARTIN 1996. *The Night is Large. Collected Essays 1938-1995*. Penguin, London.

GEACH, PETER & BLACK, MAX (eds.) 1952. *The Philosophical Writings of Gottlob Frege*. Barns and Noble Books, New Jersey. Third edition, 1980.

GJERTSEN, DEREK 1989. *Science and Philosophy: Past and Present*. Penguin Books, London.

GÖDEL, KURT 1930. 'Die Vollständigkeit der Axiome des logischen Funktionenkalküls', *Monatshefte für Mathematik und Physik* **37**, 349-360.

— 1931. 'Über formal unentscheidbare Sätze der *Principia Mathematica* und verwandter Systeme I', *Monatshefte für Mathematik und Physik*, **38**, 173-198.

— 1940. *The Consistency of the Axiom of Choice and of The Continuum Hypothesis with the Axioms of Set Theory. Annals of Mathematics Studies* No. **3**., Princeton University Press, Princeton.

— 1944. 'Russell's Mathematical Logic'. In P.A. Schilpp (ed.), *The Philosophy of Bertrand Russell*, Northwestern University, Evanston and Chicago. In Benacerraf & Putnam (eds.) 1983.

— 1947. 'What is Cantor's Continuum Problem?', *American Mathematical Monthly* **54**.

— 1947+64 . 'What is Cantor's Continuum Problem?', revised version of Gödel 1947, in Benacerraf & Putnam (eds.) 1964 and Benacerraf & Putnam (eds.) 1983.

— 1967. 'On Formally Undecidable Propositions of *Principia Mathematica* and Related Systems', Translation of Gödel 1931 by J. van Heijenoort in van Heijenoort 1967.

GOODMAN, NELSON 1972. *Problems and Projects.* The Bobbs-Merrill Company Inc., New York.

GOTTLIEB, DALE 1980. *Ontological Economy: Substitutional Quantification and Mathematics.* Clarendon Press, Oxford.

HAACK, SUSAN 1978. *Philosophy of Logics.* Cambridge University Press, Cambridge.

HAHN, L.E. & SCHILPP, P.A. (eds.) 1986. *The Philosophy of W.V. Quine.* Open Court, La Salle, Illinois.

HALE, BOB 1987. *Abstract Objects.* Blackwell, Oxford.

HARDY, G.H. 1940. *A Mathematician's Apology.* Cambridge University Press, Cambridge.

HART, W.D. 1979. 'Access and Inference'. *Proceedings of the Aristotelian Society*, **53**, 153-165. Also in Hart (ed.) 1996.

— (ed.) 1996. *The Philosophy of Mathematics.* Oxford University Press, Oxford.

HAWKES, TERENCE 1977. *Structuralism and Semiotics.* Methuen, London.

HAWKING, STEPHEN 1988. *A Brief History of Time: From the Big Bang to Black Holes.* Bantam Press, London.

VAN HEIJENOORT, J. (ed.) 1967. *From Frege to Gödel.* Harvard University Press, Cambridge, Ma.

HELLMAN, GEOFFREY 1989. *Mathematics Without Numbers: Towards a Modal Structuralist Interpretation.* Clarendon Press, Oxford.

HEMPEL, CARL 1945. 'On The Nature of Mathematical Truth', *American Mathematical Monthly* **52**, 543-556. In Benacerraf & Putnam (eds.) 1983.

HEYTING, AREND 1956. *Intuitionism: An Introduction.* North-Holland, Amsterdam. The section, "Disputation", is reprinted in Benacerraf & Putnam (eds.) 1983.

HILBERT, DAVID 1899. *Grundlagen der Geometrie.* Teubner, Leipzig.

— 1926. 'Über das Unendliche', *Mathematische Annalen* **95**. Translated as 'On The Infinite' by E. Putnam and G.J. Massey, in Benacerraf & Putnam (eds.) 1983.

— 1971. *Foundations of Geometry.* Translation of Hilbert 1899 (10th ed.) by L. Under. Open Court, Chicago and La Salle, Ill.

HINTIKKA, JAAKKO (ed.) 1969. *The Philosophy of Mathematics.* Oxford University Press, Oxford.

HODGES, WILFRID 1997. *A Shorter Model Theory.* Cambridge University Press, Cambridge.

HORWICH, PAUL 1990. *Truth.* Blackwell, Oxford.

ISAACSON, DANIEL 1985. 'Arithmetic Truth and Hidden Higher-Order Concepts', The Paris Logic Group (eds.), *Logic Colloquium '85.* North Holland, Amsterdam, 1987. In Hart (ed.) 1996.

JEANS, SIR JAMES 1930. *The Mysterious Universe.* Pelican Books, London.

JEFFREY, RICHARD 1996. 'Logicism: 2000: A Manifesto'. In Morton & Stich (eds.) 1996.

KANT, IMMANUEL 1781/1787. *Kritik der reinen Vernunft*, Konigsberg. Translation *The Critique of Pure Reason*, by J.M.D. Meiklejohn, Everyman Classics, J.M. Dent & Sons Ltd., London.

KIRKHAM, RICHARD L. 1992. *Theories of Truth.* Paperback edition, Bradford Books, MIT Press, Cambridge, Ma., 1995.

KNEALE, WILLIAM & KNEALE, MARTHA. 1962. *The Development of Logic.* Paperback edition, Clarendon Press, Oxford, 1988.

KRANTZ, D.H., SUPPES, P., TVERSKY, A. & LUCE, R.D. 1971. *The Foundations of Measurement* (Vol I). Academic Press Inc., San Diego.

KRIPKE, SAUL 1959. 'A Completeness Theorem in Modal Logic', *Journal of Symbolic Logic* **24**, 1-14.

— 1963. 'Semantical Considerations in Modal Logic', *Acta Philosophica Fennica* **16**, 83-94.

— 1975. 'Outline of a Theory of Truth', *Journal of Philosophy* **72**, 690-716. In Robert L. Martin (ed.) 1984, *Recent Essays on Truth and the Liar Paradox.* Oxford University Press, Oxford.

— 1976. 'Is There a Problem With Substitutional Quantification?', in G. Evans & J. McDowell (eds.) 1976, *Truth and Meaning*, Oxford University Press, Oxford.

— 1980. *Naming and Necessity.* Blackwell, Oxford.

LAUDAN, LAWRENCE 1981. 'A Confutation of Convergent Realism', *Philosophy of Science* **48**, 19-48. In Papineau (ed.) 1996.

LAVINE, SHAUGHAN 1994. *Understanding the Infinite.* Harvard University Press, Cambridge, Ma.

LEEDS, STEVEN 1978. 'Theories of Reference and Truth'. *Erkenntnis* **13**, 111-129.

LEWIS, DAVID 1979. 'Counterpart Theory and Quantified Modal Logic', in M.J. Loux (ed.) 1979, *The Possible and The Actual*, Ithaca.

— 1991. *Parts of Classes.* Oxford University Press, Oxford.

LONGAIR, MALCOLM 1984. *Theoretical Concepts in Physics.* Cambridge University Press, Cambridge.

LOSEE, JOHN 1993. *A Historical Introduction to the Philosophy of Science.* Oxford University Press, Oxford. Third edition.

LUCAS, J.R. 1961. 'Minds, Machines and Gödel' *Philosophy* **36**: 120-124.

MACHOVER, MOSHÉ 1996. *Set Theory, Logic and Their Limitations*, Cambridge University Press, Cambridge.

MADDY, PENELOPE 1980. 'Perception and Mathematical Intuition', *Philosophical Review* **89**, 163-196. In Hart (ed.) 1996.

— 1990. *Realism in Mathematics.* Clarendon Press, Oxford.

— 1992. 'Indispensability and Practice', *Journal of Philosophy* **89**, 275-289.

— 1996. 'The Legacy of Mathematical Truth', in Morton and Stich (eds.) 1996.

MALAMENT, DAVID 1982. 'Review of Field 1980', *Journal of Philosophy* **79**, 523-34.

MARTIN, R.M. 1963. *Intension and Decision: A Philosophical Study.* Prentice-Hall, Englewood Cliffs, NJ.

MENDELSON, ELLIOT 1987. *An Introduction to Mathematical Logic.* Wadsworth and Brooks, Monterey, Cal. Third edition.

MISNER, C.W., THORNE, K.S. & WHEELER, J.A. 1975. *Gravitation.* Freeman, San Francisco, Cal.

MOORE G.H. 1982. *Zermelo's Axiom of Choice: Its Origins, Development, and Influence.* Springer-Verlag, New York.

MORTON, ADAM & STICH, STEVEN (eds.) 1996. *Benacerraf and his Critics.* Blackwell, Oxford.

MOSTOWSKI, A. 1939. 'Über die Unabhängigheit des Wohlordnungsatzes vom Ordnungsprinzip', *Fundamenta Mathematicae* **32**, 201-252.

MUSGRAVE, ALAN 1989. 'NOA's Ark: Fine for Realism', *Philosophical Quarterly* **38**, 383-398. In Papineau (ed.) 1996.

OCKHAM, WILLIAM *Super Quattuor Libros Sententiarum.* Lyons 1495. See Kneale & Kneale 1988.

— *Summa Totius Logicae.* Oxford 1675. See Kneale & Kneale 1988.

Papineau, David (ed.) 1996. *The Philosophy of Science.* Oxford University Press, Oxford.

Paris, Jeff & Harrington, Leo 1977. 'A Mathematical Incompleteness in Peano Arithmetic', in Jon Barwise (ed.), *Handbook of Mathematical Logic*, North Holland, Amsterdam, 1977.

Parsons, Charles 1971. 'A Plea for Substitutional Quantification'. *Journal of Philosophy* **68**, 231-237. In Parsons 1983a.

— 1975. 'What is the Iterative Conception of Set', in R.E. Butts & J. Hintikka (eds.) 1975, *Logic, Foundations of Mathematics and Computability Theory* (Proceedings of the Fifth International Congress of Logic, Methodology and the Philosophy of Science: London, Ontario), Reidel, Dordrecht. In Benacerraf & Putnam (eds.) 1983.

— 1983a. *Mathematics in Philosophy: Selected Essays.* Cornell University Press, Ithaca, New York.

— 1983b. 'Quine on the Philosophy of Mathematics' in Parsons 1983a. Also in. Hahn & Schilpp (eds.) 1986.

Penrose, Roger 1989. *The Emperor's New Mind.* Vintage, London, 1990.

Plato *Phaedo.* In *The Collected Dialogues.* Edited by E. Hamilton and H. Cairns. Princeton University Press, Princeton, 1961.

Plochman, G.K. 1973. *Plato.* Dell, New York.

Pollard, Stephen 1990. *A Philosophical Introduction to Set Theory.* University of Notre Dame Press, Notre Dame.

Poincaré, Henri 1905. *Science et Hypothèse.* Translation *Science and Hypothesis*, by W.J. Greenstreet, Dover, New York, 1952.

Popper, Sir Karl R. 1982. *The Open Universe: An Argument for Indeterminism.* Hutchinson, London.

Putnam, Hilary 1967a. 'The Thesis That Mathematics is Logic', in R. Schoenman 1967 (ed.), *Bertrand Russell, Philosopher of the Century*, Allen and Unwin, London. Also in Putnam 1979.

— 1967b. 'Mathematics Without Foundations', *Journal of Philosophy* **64**. In Putnam 1979, in Benacerraf & Putnam (eds.) 1983, and in Hart (ed.) 1996.

— 1971. *Philosophy of Logic.* In Putnam 1979.

— 1975. 'What is Mathematical Truth?', in Putnam 1979.

— 1979. *Mathematics, Matter and Method. Philosophical Papers* (Vol 1). Cambridge University Press, Cambridge. Second edition.

— 1981. *Reason, Truth and History.* Cambridge University Press, Cambridge.

— 1988. 'Model Theory and the Factuality of Semantics'. In George, A. (ed.) 1988, *Reflections on Chomsky.* Clarendon Press, Oxford.

Quine, W.V. 1936. 'Truth by Convention', in O.H. Lee (ed.) 1936, *Philosophical Essays for Alfred North Whitehead*, Longmans, New York. In Quine 1976 and in Benacerraf & Putnam (eds.) 1983.

— 1939. 'Designation and Existence', *Journal of Philosophy* **36**.

— 1940. *Mathematical Logic*. Harvard University Press, Cambridge, Ma.

— 1947. 'Logic and the Reification of Universals', in Quine 1953a and 1980.

— 1948. 'On What There Is', *Revue of Metaphysics* **2**. In Quine 1953a and 1980.

— 1950. *Methods of Logic*. Harvard University Press, Cambridge, Ma.. Third edition, 1972.

— 1951a. 'Two Dogmas of Empiricism', *Philosophical Review* **60**. In Quine 1953a and 1980.

— 1951b. 'On Carnap's Views on Ontology', *Philosophical Studies* **2**. In Quine 1966 and Quine 1976.

— 1953a. *From a Logical Point of View: Nine-Logico-Philosophical Essays*. Harvard University Press, Cambridge, Ma.

— 1953b. 'Notes on the Theory of Reference', in Quine 1953a and 1980.

— 1953c. 'Reference and Modality', in Quine 1953a and 1980.

— 1953d. 'Three Grades of Modal Involvement', *Proceedings of the XIth International Congress of Philosophy*, Volume 14, North-Holland, Amsterdam. In Quine 1966 and 1976.

— 1954. 'Carnap and Logical Truth', in P.A. Schilpp (ed.) 1963, *The Philosophy of Rudolf Carnap*, Open Court, La Salle, Ill. Appeared in *Synthese* **12** (1960). In Quine 1966 and 1976.

— 1960. *Word and Object*. MIT Press, Cambridge, Ma.

— 1962. 'Reply to Professor Marcus'. *Synthese* **20**. In Quine 1966 and 1976.

— 1963a. *Set Theory and its Logic*. Belknap Press of Harvard University Press, Cambridge, Ma. Revised edition, 1969.

— 1963b. 'On Simple Theories of a Complex World', *Synthese* **15**. In Quine 1966 and 1976.

— 1966. *The Ways of Paradox*. Harvard University Press, Cambridge, Ma.

— 1966-74. 'On Multiplying Entities'. In Quine 1976.

— 1969a. *Ontological Relativity and Other Essays*. Columbia University Press, New York.

— 1969b. 'Ontological Relativity', in Quine 1969a.

— 1969c. 'Existence and Quantification', in Quine 1969a.

— 1970. *Philosophy of Logic*. Harvard University Press, Cambridge, Ma. Second edition, 1986.

— 1973. 'The Limits of Knowledge', Radio Talk. In Quine 1976.

— 1974. *The Roots of Reference*. Open Court, La Salle, Ill.

— 1976. *The Ways of Paradox*. Revised enlarged second edition. Harvard University Press, Cambridge, Ma.

— 1977. 'Facts of the Matter'. In R.W Shahan and K.R. Merrill (eds.), *American Philosophy from Edwards to Quine*. University of Oklahoma Press. Also in R.W. Shahan and C.V. Swoyer (eds.), *Essays on the Philosophy of W.V. Quine*. The Harvester Press, Hassocks, Sussex, 1979.

— 1978a. 'Success and Limits of Mathematization'. In Quine 1981a.

— 1978b. 'Kurt Gödel'. *1978 Year Book* for American Philosophical Society. In Quine 1981a.

— 1980. *From a Logical Point of View: Nine-Logico-Philosophical Essays*. Second revised edition. Harvard University Press, Cambridge, Ma.

— 1981a. *Theories and Things*. Belknap Press of Harvard University Press, Cambridge, Ma.

— 1981b. 'Things and Their Place in Theories', in Quine 1981a.

— 1981c. 'Soft Impeachment Disowned', *Pacific Philosophical Quarterly* **61**. Appeared as 'Responding to David Armstrong' in Quine 1981a.

— 1986. 'Reply to Charles Parsons'. In Hahn & Schilpp (eds.) 1986.

— 1987. *Quiddities*. Penguin, London.

— 1995. *From Stimulus to Science*. Harvard University Press, Cambridge, Ma.

QUINE, W.V. & GOODMAN, NELSON 1947. 'Steps Towards a Constructive Nominalism', *Journal of Symbolic Logic* **12**, 105-122. In Goodman 1972.

RAMSEY, FRANK PLUMPTON 1927. 'Facts and Propositions', *Proceedings of the Aristotelian Society*, supp. vol. 7., pp. 153-170. In Ramsey 1931 and Ramsey 1978.

— 1931. *The Foundations of Mathematics and Other Essays*. Edited by R.B. Braithwaite. Routledge, London, 1931.

— 1978. *Foundations*. Edited by D.H. Mellor. Routledge, London.

RESNIK, MICHAEL 1981. 'Mathematics as a Science of Patterns: Ontology and Reference', *Noûs* **15**, 529-550.

— 1982. 'Mathematics as a Science of Patterns: Epistemology', *Noûs* **16**, 95-105.

— 1988. 'Mathematics From the Structural Point of View', *Revue Internationale de Philosophie* **42**, 400-424.

Rindler, Wolfgang 1969. *Essential Relativity.* Springer-Verlag, New York. Revised second edition, 1979.

Robbin, Joel 1969. *Mathematical Logic: A First Course.* W.A. Benjamin, New York.

Russell, B.A.W. 1903. *The Principles of Mathematics.* Cambridge University Press. Paperback edition: Routledge, London, 1992.

— 1905. 'On Denoting', *Mind* **14**, 479-493. In Russell 1956.

— 1912. *Problems of Philosophy.*, Williams and Norgate, London. Paperback edition: Oxford University Press, Oxford, 1967.

— 1918-19. 'The Philosophy of Logical Atomism', *The Monist* **28**: 495-527; *The Monist* **29** (1919): 32-63, 190-222, 345-380. In Russell 1956.

— 1919. *Introduction to Mathematical Philosophy.* Allen & Unwin Ltd., London.

— 1945. *History of Western Philosophy.* Unwin Paperbacks, London, 1979.

— 1950. 'Logical Positivism', in Russell 1956.

— 1956. *Logic and Knowledge: Essays 1901-1950.* Paperback edition. Unwin Paperbacks, London, 1988.

Russell, B.A.W. & Whitehead, A.N. 1910-12. *Principia Mathematica.* Vols. I-III. Cambridge University Press, Cambridge.

Ryle, Gilbert 1949. *The Concept of Mind.* Hutchinson, London.

Saussure, Ferdinand de 1916. *Course in General Linguistics.* Translated by W. Baskin, McGraw-Hill Book Co., New York, 1966.

Shapiro, Stewart 1983a. 'Conservativeness and Incompleteness', *Journal of Philosophy* **80**, 521-531. In Hart (ed.) 1996.

— 1983b. 'Mathematics and Reality', *Philosophy of Science* **50**, 523-548.

— 1991. *Foundations Without Foundationalism.* Oxford Logic Guides - 17. Clarendon Press, Oxford.

— 1997. *Philosophy of Mathematics: Structure and Ontology.* Oxford University Press, Oxford.

Scheffler, Israel & Chomsky, Noam 1958. 'What is Said To Be', *Proceedings of the Aristotelian Society* **59**.

Simons, Peter 1987. *Parts: A Study in Ontology.* Oxford University Press, Oxford.

Smullyan, Raymond 1987. *Forever Undecided: A Puzzle Guide to Gödel.* Oxford University Press, Oxford.

— 1993. *Satan, Cantor and Infinity.* Oxford University Press, Oxford.

Spinoza, B. 1677. *Tractatus de Intellectus Emendatione.* (Published posthumously in a work called *Opera Posthuma*). Translation by R.H.M. Elwes as *On the Improvement* *of the Understanding. The Ethics. Correspondence.* Dover Publications Inc., New York, 1955.

STEBBING, SUSAN 1937. *Philosophy and the Physicists.* Methuen, London.

STEWART, IAN 1987. *The Problems of Mathematics.* Oxford University Press, Second edition (renamed: *From Here to Infinity*), 1992.

SUTHERLAND, W.A. 1975. *Introduction to Metric and Topological Spaces.* Clarendon Press, Oxford.

SYNGE, J.L. & GRIFFITH, B.A. 1942. *Principles of Mechanics.* McGraw-Hill Book Company Inc., New York, Third edition, 1959.

TAKEUTI, GAISI 1975. *Proof Theory* (Studies in Logic and the Foundations of Mathematics, Vol. 81). North Holland, Amsterdam.

TAIT, W.W. 1986. 'Truth and Proof: The Platonism of Mathematics', *Synthese* **69**, 341-370. In Hart (ed.) 1996.

TARSKI, ALFRED 1936. 'Der Wahrheitsbegriff in den formalisierten Sprachen', *Studia Philosophica* **I** (1936), 261-405.

— 1956. *Logic, Semantics, Metamathematics: Papers by Alfred Tarski from 1922-1938.* Clarendon Press, oxford. Edited and translated by J.H. Woodger.

— 1951. *A Decision Method for Elementary Algebra and Geometry.* Berkeley, Cal.

— 1959. 'What is Elementary Geometry?', in L. Henkin, P. Suppes and A. Tarski (eds.), *The Axiomatic Method with Special Reference to Geometry and Physics,* Amsterdam, 1959. Also in Hintikka (ed.) 1969.

TARSKI, A., MOSTOWSKI, A., & ROBINSON, R.M. 1953. *Undecidable Theories.* North Holland, Amsterdam

THOMAE, J. 1898. *Elementare Theorie der analytische Functionen einer complexen Veränderlichen.*

ULLIAN, JOSEPH S. 1986. 'Quine and the Field of Mathematical Logic'. In Hahn & Schilpp (eds.) 1986.

VAIHINGER, HANS 1913. *Die Philosophie des Als Ob.* Verlag von Reunther & Reichard, Berlin. Translated as *The Philosophy of "As If"*, by C.K. Ogden, Routledge and Kegan Paul, London, 1935.

WAGNER, STEVEN 1982. 'Arithmetical Fiction'. *Pacific Philosophical Quarterly* **63**, 255-269.

WALD, ROBERT 1984. *General Relativity.* University of Chicago Press, Chicago.

WEINBERG, STEVEN 1972. *Gravitation and Cosmology.* Wiley, New York.

— 1986. 'Lecture on the Applicability of Mathematics'. *Notices of the American Mathematical Society* **33**.

WEYL, HERMAN 1910, 'Über die Definitionem der mathematischen Grundbegriffe'. In Weyl 1968, *Gesammelte Abhandlungen.*

WIGNER, EUGENE P. 1960. 'The Unreasonable Effectiveness of Mathematics', *Communications in Pure & Applied Mathematics* **13**, 1-14.

WITTGENSTEIN, LUDWIG 1922. *Tractatus Logico-Philosophicus.* Translated by C.K Ogden. Paperback edition, 1981. Routledge, London.

WRIGHT, CRISPIN 1983. *Frege's Conception of Numbers as Objects.* Aberdeen University Press, Aberdeen.

ZERMELO, ERNST 1908. 'Untersuchungen über die Grundlagen der Mengenlehre I', *Mathematische Annalen* **65**. English translation by S. Bauer-Mengelberg, 'Investigations in the Foundations of Set Theory I', in van Heijenoort (ed.) 1967.

ADDENDUM

Indispensability and Constructivism

If number were an idea, then arithmetic would be psychology. ... It would be strange if the most exact of all the sciences had to seek support from psychology, which is still feeling its way none too surely.

Gottlob Frege 1884, *Grundlagen der Arithmetic* §27.

Pure mathematics, on the other hand, seems to me a rock on which all idealism founders: 317 is prime, not because we think so or because our minds are shaped in one way or another, but *because it is so*, because mathematical reality is built that way.

G.H. Hardy 1940, *A Mathematician's Apology*, p. 70.

... on a limited idealist view, one that views mathematical entities as some sort of human construction but makes no claim about the physical world, the application of mathematics to the physical world may turn out to be a mystery. The danger, in other words, is that in order to explain the applicability of mind-dependent mathematical entities to the physical world, the idealist about mathematics may have to become a full-blown idealist, and hold that even things like electrons and dinosaurs are somehow 'human constructions'. If this danger were realized, I would regard that as a *reductio ad absurdum* of the idea that mathematical objects were human constructions.

Hartry Field 1989, *Realism, Mathematics and Modality*, p. 27. Footnote 16.

A.1 Constructivism & Indispensability

I have argued in the main body of the thesis that *nominalism* cannot answer the Quine-Putnam indispensability argument. However, perhaps we have not paid sufficient attention to another school: *constructivism*. The constructivist conception of mathematics claims that mathematicalia themselves are *mental constructions*. Thus, constructivism involves a plainly idealistic premise that mathematicalia are dependent on our definitions and constructions. Such entities are alleged to be "produced" by mathematical constructions.[242] There are several ways to implement mathematical

[242] I find it exceedingly difficult to grasp this metaphor in the literal sense required by the constructivist (c.f., the notion of "construction" in Kant's (first) *Critique*, Russell's 1914 *Our Knowledge of the External World* and Carnap's 1934 *Aufbau*. What these authors intend (I think) is construction of *representations* of an object, not construction of the object itself. To conflate these is to commit the preposterous crime of idealism about the physical world). I can construct something

idealism: *intuitionism* (e.g., Brouwer, Heyting, Dummett), *definitionism* (c.f., Poincaré) or *constructivism* (e.g., Kronecker, Weyl, Bishop, Feferman). And there are various sub-schools and sub-programmes.

The guiding spirit of constructivism was neatly summarized by G.T. Kneebone:

> Many eminent mathematicians in fact, from Kronecker onwards, have held that a mathematical entity is not properly defined unless the definition that is offered permits the construction of the entity (or at least a controlled approximation to it) by a finite process.
>
> (Kneebone 1963, p. 274).

In Chapter 1 I argued that there is in fact not even a *prima facie* reason to suppose that an idealistic interpretation of mathematics is true. The objectivity (or inter-subjectivity) of mathematics is *prima facie* evidence that mathematical facts are *found* and not "made" or "constructed" (c.f., Hardy 1940). However, perhaps the constructivist can appeal to some universal human ability of "mental construction" to explain this inter-subjectivity. (C.f., the Chomskyan appeal to a universal (human) language acquisition ability for the natural biological development of language by infants).

A standard reason for rejecting constructivism was stressed by Hilbert.[243] Constructivism imposes severe constraints on what might count as intelligible reasoning *within* mathematics (or as an acceptable proof *within* mathematics) and, indeed, on the structure of the *mathematical universe* (for example, it cannot be non-countable).

In particular, constructivism counts as *unacceptable* various forms of reasoning involving impredicative existence assumptions, indirect existence proofs (e.g., by *reductio ad absurdum*), proofs requiring the law of excluded middle (LEM), quantification over *arbitrary* subsets of any denumerable infinity (and, indeed,

physical like a *chair* by taking some pieces of wood, some glue and some nails. Likewise, I can construct a physical sentence *token* by vibrating my larynx or moving my hands. But, to speak strictly, I have no idea how to "construct" a neutron or a galaxy, a transfinite number or a manifold. (Of course, we can make lots of neutrons by refining some uranium ore, forming a critical mass of U_{235} and dropping the ingenious result on innocent civilians).

243 Although Hilbert welcomed non-constructive axioms and proofs *within* mathematics as "ideal elements", he imposed the analogous demand for a highly constructive (finitary) *metamathematics*. In particular, Hilbert proposed that mathematicians attempt to find finitary metamathematical *consistency proofs* for (axiomatic) arithmetic, analysis and set theory. This proposal was of course scuppered by Gödel's incompleteness theorems (1931).

reference to *non-denumerable sets*). Hilbert compared these demands of constructivism to the absurd demand that a "boxer should fight with both hands tied behind his back". In particular, two severe limitations arise: one cannot develop standard (impredicative) *real analysis* and modern *impredicative set theory* (e.g., as codified in *ZFC*).

For similar reasons, very many working mathematicians who reflect upon the topic are opposed to constructivism, whatever its *a priori* philosophical defence. As noted by Gardner 1996 (and many others), the working mathematician thinks of mathematics as having an *objective mind-independent subject matter*, which would exist even if there were no such things as proofs, definitions and constructions.

One of the most important recent *philosophical defences* of mathematical platonism (via considerations largely internal to the practice of mathematics) was given by Kurt Gödel (Gödel 1944 and 1947+1964). Moreover, platonism (in some form or another) has been defended by a variety of mathematicians and physicists (e.g., Hardy, Jeans, Stewart, Gardner, Penrose, Smullyan and many others). Virtually all mathematicians are realists, and many are prepared to defend their mathematical realism by philosophical argument.

How does the Quine-Putnam indispensability argument relate to constructivism? Again, I briefly mentioned this in Chapter 1. In an important sense, constructivism is *prima facie* incompatible with modern spacetime theory, with its assumption of a non-countable set of spacetime points and quantification over arbitrary sets of spacetime points and functions on the set of spacetime points. The electromagnetic field A_μ, for example, is one such function. And such functions are presumed by physicists to exist *prior* to our *discovering* them. In other words, if one wants to remain a constructivist when considering the *application* of mathematics, one will have to claim that the electromagnetic field A_μ is a "mental construct". That leads to *idealism* in *physics*, and, like Field 1989, I take that to be a *reductio* of the position.

However, suppose that we are charitable to the constructivist and consider an idealistic interpretation of *sets* of spacetime points and *functions* on spacetime. To provide a constructivistic response to the Quine-Putnam argument, we will then search for *constructive replacements* for all scientifically applicable mathematics. The central issue concerns the question of how much mathematics is indispensable to science. The importance of the Quine–Putnam argument is that it provides a powerful weapon in the *anti-nominalist*'s armoury. Unless one advocates some *a priori* (and unscientific) scepticism about science itself (e.g., van Fraassen 1980, Cartwright 1983), at least *some* mathematical entities must *exist*. One might argue that this argument justifies the existence *only* of a "small universe" of mathematicalia that has found applications in *science*. In particular, it might be possible to subsume this small mathematical universe under some version of constructivism (i.e., idealism). This is a view that has recently been urged, for example, by Solomon Feferman.

But I shall now argue that this prospect is highly unlikely.

A.2 Non-Constructive Mathematics in Physics

How much mathematics does science need? If science only needs mathematics of a certain "minimal" kind, then perhaps all the "surplus" can be treated instrumentalistically, as some kind of convenient fiction. I will argue that this dispensability claim is false. In fact, I am sceptical that there is any sharp or interesting distinction between mathematics "needed" for applications and mathematics that "transcends" application. As well as the pure machinery of the various number systems (**N**, **Z**, **Q**, **R** and **C**), abstract algebras and so on, the modern theoretical physicist needs to *assume* that there exists a (non-countable) set E of spacetime points (to form the spacetime manifold) and that there exist uncountably-many arbitrary subsets of this set (e.g., there is no reason to suppose that the natural topology on E is *countable* subset of $\mathsf{P}(E)$; physical geodesics in E are ordered sets in E and there are uncountably-many such geodesics; and so on). From a realist viewpoint, it would be simply astonishing if the presumably ultra-complicated *mind-independent structure* of the external world of matter in physical spacetime were fully "constructible" by the all-too-limited human mind. There is no reason to believe that this is so. The constructivist about applied mathematics has to establish that *all* of the mathematics used in applications can be compressed into the Procrustean bed of constructive mathematics.

John Burgess has neatly summarized the situation:

> Even accepting the opinion that 95% of applications can be fairly easily handled by a moderately liberal constructivistic mathematics, if one is going to adopt the militant attitude of the early constructivists like Brouwer, and call for the outright abolition of classical mathematics, one has a responsibility to look at the more problematic 5% remainder. And the most problematic cases seem to be connected with highly theoretical science. Professor Hellman cited one such example in Uppsala, connected with quantum mechanics; Roger Penrose cited another from his own research, related to general relativity. Had I been a participant rather than a moderator, I would probably have said something about measurable selection theory, a cluster of results on the theoretical fringes of subjects whose cores are applied: optimization and control theory, probability and statistics, mathematical economics, operator theory.
>
> (Burgess 1992, pp. 436-437).

Further emphasizing the likely fragility of the constructivist case, Geoffrey Hellman has given fairly detailed analyses of the applicable mathematics required in both Quantum Mechanics and General Relativity. An important result along these lines is that the constructive statement of Gleason's Theorem, which plays an important role in foundational studies of Quantum Mechanics cannot be constructively proved (see Hellman 1993). He concludes that,

> The demand to respect scientifically applicable mathematics has important consequences tending to rule against the adequacy of Brouwer and Bishop constructivism. Classical, non-constructive concepts and reasoning would seem indispensable to some of our best science.
>
> (Hellman 1992, p. 462).

Let us quickly survey some of the mathematics used in modern physics.

A.2.1 Application of the real number structure

Constructivists do not admit, as a constructible reality, the uncountable real number structure of impredicative real analysis. However, we certainly want to *refer* to this structure in describing the physical world. The standard application involves *spacetime*, but an even simpler application of the ordered real number structure (**R**, <) involves just *time* itself.

Let *Tim* be the set of temporal instants and let *Before* be the before-after relation on *Tim*. Standard spacetime theory asserts that *there exists an isomorphism* ρ from (*Tim*, *Before*) to (**R**, <). It follows from this that the "impure structure" (*Tim*, *Before*) is non-denumerable and is order-complete (that each bounded set of temporal instants has a supremum[244]). This is *prima facie* unacceptable to a constructivist, for the supremum axiom is impredicative.

In short, this implies that, because the real number structure is (in a sense) "built-in" to the physical world, as soon as we discover this (objective) fact, we are *compelled* to use *impredicative non-constructive mathematics* to study this aspect of the structure of the real world (namely, the before-after ordering of time instants).

A.2.2 Important non-constructive spaces in physics

Burgess and Rosen 1997 suggest that *mathematical analysis* is "probably sufficient to develop, making use of coding devices, all the mathematics that has found scientific applications up to the present" (Burgess & Rosen 1997, p. 386). In footnote 328 above, I expressed strong reservations about this claim, citing differential geometry, topology, cohomology, fibre bundle theory, symplectic manifold theory, dimensional regularization in non-integral-dimensional spacetime and infinite-dimensional path integrals.

To be more specific, the following notions from differential geometry, "topological manifold", "metric tensor", "geodesic completeness", "affine connection", "time function", "Cauchy surface", "fibre bundle" are all firmly embedded within highly successful mathematicized theories from modern theoretical physics.

Can differential geometry be given a constructive formulation? It is true that a certain amount of differential geometry *can* be reproduced within real analysis, by coding techniques. This is the reductive technique that Burgess and Rosen allude to. However, as Hellman 1989 has stressed, it seems to be the case that the *general notions* above (e.g., of a topological manifold) and several other notions required in *GR*, for example, transcend even the (coding capacities) of analysis:

> The abstract theory of manifolds transcends the RA^2 framework[245], but essentially only at the earliest stages, namely in the abstract characterization of manifolds themselves.

[244] This is sometimes called Dirichlet's axiom (see Kneebone 1963, p. 256).

(Hellman 1989, p. 108).

(This is essentially because, according to classical set theory, there are more such structures (manifolds, functions on a given manifold, geodesics, etc.) than real numbers. You cannot code *every* manifold as a real).

Indeed, the modern theoretical physicist has no qualms about quantifying over arbitrary subsets of a given base set. Thus, "big" spaces are invoked in modern physics. In the standard case, the base space is usually related to spacetime and is uncountable itself (as big as the classical continuum, **R**). Examples of big spaces introduced via power set operations on the base space are,

(i) *phase spaces* for classical fields,

(ii) the space of *arbitrary functions* on a given hyper-surface in phase space or spacetime,

(iii) spaces of *trajectories* and *paths* on a configuration space or on spacetime.

A physical field on spacetime is a function $\varphi: E \rightarrow \mathbf{V}$, where E is the set of spacetime points and **V** is some vector space. The physics itself asserts that there is a bijection between E and **R**, but this bijection is not constructible. The *set* of *all* these fields constitutes the *configuration space* for the field φ in question. This set is bigger than the continuum. The same applies to a field φ specified on an boundary hypersurface Σ in space-time.

For example, the "initial value problem" asks whether the time-evolution of φ, constrained by the equation of motion for φ, is *uniquely determined* by the initial values on Σ at $t = 0$ and to what extent *arbitrary perturbations* of the initial state $\varphi(x, y, z, 0)$ affect the evolution at later times. The spaces involved are large: the class of *all* functions $\varphi(x, y, z, 0)$ on the boundary surface Σ.

An even more involved example is the "Path-Integral" formulation of Quantum Mechanics introduced by Richard Feynman. The quantum-mechanical "amplitude" for a particle to travel from a spacetime position (q, t) to position (q^*, t^*) is given by an *integral* over *all possible space-time trajectories* from (q, t) to (q^*, t^*):

---

245 RA^2 is (axiomatic) second-order analysis: the usual axioms for the ordered archimedean field with the second-order version of the order-completeness axiom (= l.u.b. axiom = Dirichlet's axiom).

$$<q, t; q^*, t^*> = \int \mathrm{D}f(q, t) \exp\{i \int dt\, S(f, {}^{df}/_{dt}, t)\}$$

The space of possible trajectories is enormous, since they are parametrized by all possible continuous maps $f(t)$: $[0, 1] \rightarrow E$ such that $f(0) = (q, t)$ and $f(1) = (q^*, t^*)$. The measure for the path integral $\mathrm{D}f(q, t)$ contains a function variable f which ranges over the (non-denumerable) totality of *all* paths, not just the definable ones.

A.2.3 Feferman's predicative set theory

Solomon Feferman has devised a *predicative* type-theoretic set theory, which treats the natural numbers as urelements (see Feferman 1988, 1992). Feferman calls this set theory *W* (after Hermann Weyl). Feferman proves two important facts about *W*:

(i) *W* is proof-theoretically reducible to *PA*

(ii) *W* is a conservative extension of *PA*

Feferman then argues that "by the fact of the proof-theoretical reduction of *W* to *PA*, the only ontology it commits one to is that which justifies acceptance of *PA*" (1992, p. 451). Since mainstream constructivists are happy to treat *PA* as an acceptable account of our arithmetical constructional abilities, it follows that the *W* similarly satisfies such constructivist demands. In his 1988, Feferman conjectures that,

> ... all scientifically applicable mathematics can be formalized in (a subtheory of) W, and hence does not require the assumption of impredicative set theory or of uncountable cardinal numbers for its eventual justification"
>
> (Feferman 1988, p. 89-90).

In his 1992, Feferman notes that mathematical physics "makes primary use of mathematical analysis on Euclidean, complex and Riemannian spaces, and of functional analysis on various Hilbert and Banach spaces" (p. 443). After a short discussion of which theorems of classical analysis can be proved in *W*, he concludes that,

> ... while there are clearly parts of theoretical analysis that cannot be carried out in W ..., the working hypothesis that all of scientifically applicable analysis can be developed in W has been verified in its core parts".
>
> (Feferman 1992, p. 449).

This is highly controversial, at best. It would be surprising if the basic axioms and assumptions of modern spacetime theory could be redeveloped within a predicative set theory which is conservative over *PA*. Here is an argument, analogous to Shapiro's 1983 argument against Field. Field showed how to find a rather natural replacement theory *N* such that $N + ZFC$ is equivalent to *P*, where *P* is a certain standard axiomatic spacetime theory (also talking about the gravitational field). Field required for his instrumentalist interpretation of the mathematics *ZFC* that *P* be a *conservative extension* of *N*. But Shapiro showed that *P* is *not* a conservative extension of *N*, since $P \vdash Con(N)$.

Now, let us apply a similar argument to *W*. We know from Field and Shapiro that *PA* is interpretable within the spacetime theory *N*. By theorem (i), Feferman's predicative set theory *W* is interpretable within *PA*. It follows that *W* is interpretable within *N*. It then follows that $N + W$ is a conservative extension of *N*. It then follows, via Shapiro's argument, that *N* is not "just as good as" the *original platonistic theory P*, for *P* can prove the consistency of *N* (coded as a statement in the nominalistic spacetime language L_N). In short, this indicates that the mathematics contained in Feferman's set theory *W* is (*prima facie*) *not sufficient* to do the sort of spacetime physics that our standard (platonistic) theory *P* does.

Indeed, the platonistic spacetime theory *P* states explicitly that there exist in physical spacetime both physical exemplifications of the *natural number structure* (thereby proving that *PA* is consistent) and, indeed, physical exemplifications of the *real number structure* (i.e., open line segments homeomorphic to **R** thereby proving that classical real analysis is consistent). This is the reverse of Hilbert's 1899 method of providing analytic models of geometry: contingent physical spacetime theory asserts the existence of (impure) *geometrical* models of arithmetic and analysis.

A.3 Conclusion: Naturalism and Constructivism

Einstein once remarked that the working physicist (and scientist generally) must "appear to the systematic epistemologist as an opportunist". As any working *philosopher* knows, scholarly epistemological battles about the *justification* of scientific principles are at best uninteresting to the scientist. Although in Einstein's own case (as he often stressed), epistemological considerations (especially the empiricism of Hume and Mach) were influential in his theorizing, his magnum opus (*GR*) provides little (if any) support for sceptical positions such as epistemological *positivism* (Mach) or ontological spacetime *relationism* (Leibniz).

Consider some venerable sceptical arguments from bygone days: "continuous motion is impossible" (Zeno), "infinitesimals are nonsense" (Berkeley), "the time series must have had a beginning" (Kant), "atoms and molecules are fictions" (Mach, Vaihinger), "the Axiom of Choice is unjustifiable" (Borel, Lebesgue), "completed infinities are nonsense" (Brouwer), "Dirac's delta function $\delta(x)$ is nonsense" (a common mathematical view around the time of Dirac's work in the 1920s) Although the *arguments* of venerable sceptics were indubitably clever, the sceptical *conclusions* were just *wrong*. The sceptical criticisms exhibit a certain level of rationality, but the sceptical instincts of these authors were utterly irrational. Is there any example in the history of science of the epistemological *sceptic* actually *winning* the argument? The conceptual intelligibility of *continuous motion* turned out fine (*pace* Parmenides and Zeno). Leibniz's *infinitesimals* and Newton's *moments* turned out fine (*pace* Berkeley). The possibility of an infinite time series turned out fine (*pace* Kant). Dalton's atoms and molecules turned out fine (*pace* Mach, Duhem and Vaihinger). The Axiom of Choice turned out fine (*pace* Lebesgue and Borel). Dirac's delta function turned out fine. And so on.

So, why advance such sceptical *conclusions* now? Criticism, even sceptical criticism, is healthy. But to claim that a concept *C* or a principle *P* is "meaningless" or untenable (on *a priori* epistemological grounds) is often little more than a *wild* claim. To be sure, continuous motion, infinitesimals, infinite ordered series, Dirac's function, and so on, all had their attendant conceptual problems. But some of these problems were *solved*, and almost *always*, it seems, in a way *quite opposite* to the *sceptic's intentions*.[246]

---

[246] New problems, of course, arise out of the solutions to the old problems. But, *c'est la vie*! This is why the so-called "Pessimistic Meta-Induction" from the past falsity of three or four scientific theories is so utterly muddle-headed. The proper "induction" from the real history of science is an *Optimistic* Meta-Induction. The creation of new explanatory conjectures, the solution of sceptical puzzles, the resolutions of anomalies, the daily solutions of problems—both tiny and deeply conceptual—by

A better philosophical strategy is *naturalism*, construed in the broad sense intended by Quine 1969. We already *own* our overall conceptual scheme of physical bodies and events, of mental and mathematical entities, indeed of abstract concepts and ideas: we cannot philosophize from a perspective *prior* to (or "external to") science itself. There is no *first philosophy.* Our duty as epistemologists is to analyse the concepts we already use and to understand, *from a scientific perspective*, how human knowledge of our *already accepted science* is possible. Such analysis may lead to small-scale revisions, but large-scale scepticism is not a rational option. For example, our attitude to science cannot be one of "uncompromising empiricism" (Russell's phrase, in his 1950 rebuttal of Logical Positivism), if the logical consequence of such empiricism is the sceptical abandonment of science. In Neurath's pleasing metaphor, we must rebuild the ship of knowledge while still at sea, for the epistemological *terra firma* of philosophical imagination is a mere fantasy.

In full accord with such naturalism, we conclude via the examples mentioned above that the mathematics used and required in modern theoretical science (especially theoretical physics) transcends the limits of constructivism. Constructive mathematics is (even in principle) not enough to do science. Putting it crudely, non-constructive mathematics is indispensable because the *physical universe itself* exemplifies non-constructive structures.

---

working scientists. All of these indicate *optimism* (a kind of Hegelian-Popperian optimism, perhaps, about the growth of objective knowledge).

SUPPLEMENTARY BIBLIOGRAPHY

BURGESS, JOHN P. 1992: 'How Foundational Work in Mathematics Can Be Relevant to Philosophy of Science', *Proceedings of the 1992 Philosophy of Science Association, Vol 2*, pp. 433-441.

FEFERMAN, SOLOMON 1988: 'Weyl Vindicated: "Das Continuum" 70 Years Later', *Atti del Congresso: Temi e Prospettive Della Logica e Della Filosofia Della Scienza Contemporanee (Cesena, 1987), Vol I - Logica*, pp. 59-94. Bologna: CLUEB.

—1992: 'Why a Little Bit Goes a Long Way: Logical Foundations of Scientifically Applicable Mathematics', *Proceedings of the 1992 Philosophy of Science Association, Vol 2*, pp. 442-455.

HELLMAN, GEOFFREY 1992: 'On the Scope and Force of Indispensability Arguments', *Proceedings of the 1992 Philosophy of Science Association, Vol 2*, pp. 456-464..

—1993: 'Gleason's Theorem is not Constructively Provable', *Journal of Philosophical Logic* **22**: 93-203.

KNEEBONE, G.T. 1963. *Mathematical Logic and the Foundations of Mathematics*. Paperback edition, 1965. London: Van Nostrand Co. Ltd.

QUINE, W.V. 1969. 'Epistemology Naturalized', in *Ontological Relativity and Other Essays*: New York: Columbia University Press.

WANG, HAO 1974: *From Mathematics to Philosophy*. London: Routledge.

CORRIGENDA

p. 6, ln. 8: "Hersh & Davis" should be replaced by "Davis & Hersh".

p. 7, ln. 12: Unless otherwise qualified, references to "Peano Arithmetic" are to be understood as references to *first-order* Peano Arithmetic.

p. 10, ftn. 2, ln. 2: "is nothing more that" should be "is nothing more than".

p. 30, ln. 24: "injective" should be omitted.

p. 31, ln. 1: the symbol "$\mapsto$" should be replaced by the symbol "$\rightarrow$".

p. 31, ln. 12: "$\exists!\rho(\rho(x_0) = 1$" should be replaced by "$\exists!\rho(\rho(x_0) = 1)$".

p. 32, ln. 16: The phrase "as well as the meaningless square $[\mathsf{dist}_m(x, y)]^2$" should be omitted.

p. 38, ln. 15: The term "non-finitistic" here refers to *strict* finitism (i.e., there is a fixed finite upper bound on the number of things).

p. 40, ftn. 33, ln. 2: The second occurrence of "$F(x, y)$" should be replaced by "$F(y, x)$".

p. 45, ln. 23: The first "made" should be replaced by "may".

p. 51, ftn. 44: The footnote is wrong. Cantor's writings and intuitions seem more in accord with a conception of set theory known as "limitation of size". Zermelo's discussion of the "iterative conception" does not appear in his orignal 1908 paper, but later (in the 1930s).

p. 52, ln. 14: Formula (3) should be "$\exists!x(\neg U(x) \wedge \forall z(z \notin x))$".

p. 52, ln. 16: "there is a set with no elements" should be "there is a *unique* set with no elements".

p. 54, ln. 7: "*PA*" refers to *first-order* Peano Arithmetic.

p. 60: The discussion here is restricted to Euclidean *plane* geometry.

p. 73, ln. 4: "expansion" should be replaced by "extension".

p. 90, ln. 13: "continuum many" refers to the *classical* (non-denumerable) continuum.

p. 106, ln. 35: The phrase "it seems to me that" should be omitted.

p. 108, ftn. 70: This footnote is too pessimistic. The mathematical complications described are in fact simple to avoid.

p. 122, ln. 4: The following parenthetical clause should be added after "the formula whose gödel number is *n*": "(if there is none, then let *diag*(*n*) = 1)".

p. 127, ln. 13: "expansion" should be replaced by "extension".

p. 127, ln. 23: "expansion" should be replaced by "extension".

p. 130, ln. 22: The phrase "= DT $\cap$ **A**" should be omitted.

p. 157, ln. 2: The symbol "$[A]_{||}$" means "The equivalence class of all lines x such that $x \,||\, A$".

p. 160, ln. 26: The second occurrence of "were" should be omitted.

p. 161, ftn. 111: "famous" should be "famously".

p. 169, ftn. 121, ln. 5: The phrase "recursively specifiable" should be omitted. The final sentence of the paragraph, beginning "Of course, if the "axioms" of M are ..." should be omitted.

p. 170, ln. 5: The phrase "(a refutation of $\neg\varphi$)" should be omitted.

p. 193, ln. 9: In statement (14), the initial phrase "There is a set ..." should be replaced by "there is a *non-empty* set ...".

p. 193, ln. 12: In the formula (15), the inequality "$y \neq z$" should be "$y \neq x$".

p. 214, ln. 3: The term "implication" here refers to the *relation* of *logical implication*, not to the syntactic operation of forming the conditional $\varphi \rightarrow \chi$ of the formulas φ and χ.

p. 222, ln. 3: The symbol "$\mapsto$" should be replaced by "$\rightarrow$".

p. 223, ftn. 151, ln.1: The term "full" should be replaced by "complete".

p. 237, ln. 20: The sentence "The negation of FLT is then, prima facie, an *existence* claim" should be replaced by "The negation of FLT is then classically equivalent to the following prima facie *existence* claim".

p. 255, ftn. 178: Implication is a relation. See above, for p. 214.

p. 280, ftn. 205: This footnote should be omitted, because it repeats the content of line 15 above.

p.281, 10:	"homomorphism" should (in this particular context) be replaced by "isomorphism".
p. 285, ln. 11:	The symbol "$\mapsto$" should be replaced by "$\rightarrow$".
p. 292, ln. 15:	The phrase "any questionable" should be replaced by "any more questionable".
p. 292, ln. 21:	"abstracta" should be replaced by "abstract".

Many thanks to Professor Moshé Machover for providing a list of errata.