Chapter 2
Philosophy of Mathematics and Philosophy of History

In this chapter, I argue for the historicity of mathematics, and thus for the pertinence of the study of the history of mathematics to the philosophy of mathematics. I also argue that we need to look carefully at what constitutes our philosophy of history, and the nature of historical explanation. As I hope should be evident from my arguments in the preceding chapter, this doesn’t mean that I think that what we say about mathematics is contingent or empirically based. Here I invoke Plato, the great dialectician of classical antiquity. The *Meno* is a dramatic dialogue in which two analyses take place, one mathematical and one philosophical. In the first, Socrates leads the slave boy to correct, refine, and extend his intuition of a pair of geometrical diagrams by examining his presuppositions in the light of certain background knowledge (about arithmetic), as they solve a problem together. In the second, Socrates leads Meno to correct, refine, and extend his intuition about virtue as well as about method by examining his presuppositions in the light of certain background knowledge, as they engage in philosophical dialectic together. In both cases, an intelligible unity is apprehended, but imperfectly, and analysis leads to increased understanding via a search for the conditions of intelligibility (Plato 1961: 353–384). The form of dramatic dialogue expresses something essential to the process of analysis, for the dialogue may be read both as a set of arguments (in all of which *reductio ad absurdum* plays a pivotal role) that uncover the logical presuppositions of various claims, but it may also be read as a narrative, a process in time and history.

And now I invoke Aristotle, the great logician of classical antiquity. A narrative, as Aristotle tells us in the *Poetics*, has a beginning, middle, and end. The beginning introduces us to characters who act in a certain situation, one or a few of whom have a special claim on our empathy and interest; the middle introduces one or more surprising contingencies or reversals, and the skill of the storyteller lies in maintaining the continuity of the story not in spite of but by means of those discontinuities, deepening our understanding of the characters and their actions; and the end draws us along throughout the tale: how will events ultimately turn out for these characters? (Aristotle 1947: 620–667). By contrast, an argument has the structure of
a series of premises that support, in more or less materially and formally successful ways, a conclusion. The priority of premises with respect to a conclusion is logical, not temporal or historical; an argument has no beginning or end, no before and after. The logician takes the cogency of an argument to hold atemporally: if such and such is the case, then such and such must follow, universally and necessarily. Pragmatic or rhetorical considerations may superimpose an historical dimension on the argument, but that is not the concern of the logician.

1 W. B. Gallie on History

Analysis, as I characterized it in the last chapter, is the search for conditions of the intelligibility of existing things; in mathematics, this often takes the form of solving a problem. It is an ampliative process, that increases knowledge as it proceeds. From an analysis, an argument can be reconstructed, as when Andrew Wiles finally wrote up the results of his seven-years-long search in the 108 page full dress proof in the May 1995 issue of *Annals of Mathematics* (Wiles 1995: 443–551). As we will see in the following chapter, Carlo Cellucci contrasts the analytic method of proof discovery rather starkly with the axiomatic method of justification, which reasons “downwards”, deductively, from a set of fixed axioms, as well as with algorithmic problem-solving methods (see Cellucci 2013; Ippoliti 2008). He argues that the primary activity of mathematicians is not theory construction but problem solution, which proceeds by analysis, a family of rational procedures broader than logical deduction. Analysis begins with a problem to be solved, on the basis of which one formulates a hypothesis; the hypothesis is another problem which, if solved, would constitute a sufficient condition for the solution of the original problem. To address the hypothesis, however, requires making further hypotheses, and this “upwards” path of hypotheses must moreover be evaluated and developed in relation to existing mathematical knowledge, some of it available in the format of textbook exposition, and some of it available only as the incompletely articulated “know-how” of contemporary mathematicians (see Breger 2000). Indeed, some of the pertinent knowledge will remain to be forged as the pathway of hypotheses sometimes snakes between, sometimes bridges, more than one domain of mathematical research (some of which may be axiomatized and some not), or when the demands of the proof underway throw parts of existing knowledge into question.

The great scholar Gregory Vlastos made a career out of reconstructing the arguments in Platonic dialogues, abstracting them from the dramatic action. And yet Platonic analysis is also a process of enlightenment in the life of an individual as well as a culture: the slave boy and Meno (but not Anytus) are changed in some fundamental way by their encounter with Socrates. They come to understand something that they had not understood before, and their success in understanding could not have been predicted. There is no way to cause understanding, nor can virtue be caused, even by a virtuous father desperately concerned about his wastrel son, as Socrates shows at the end of the dialogue. The middle of the plot is a crisis.
where opinions reveal their instability and things their problematicity, and it is marked by a reductio argument; but it is also marked by anxiety and curiosity on the part of Socrates’ interlocutors, emotions that might lead them equally to pursue or to flee the analysis. And at the end of the dialogue, even though it is aporetic, a shared illumination occurs that bestows on the reader a sense of the discoverable intelligibility of things, how knowledge unfolds. To understand an argument and to follow a story are two different things, but the reader must do both in order to appreciate a Platonic dialogue.

What does this have to do with philosophy of mathematics? The dialogue Meno shows that the philosophy of mathematics must stand in relation to the history of mathematics, and moreover this relationship must be undergirded by a philosophy of history that does not reduce the narrative aspect of history to the forms of argument used by logicians and natural scientists. History is primarily narrative because human action is, and therefore no physicalist-reductionist account of human action can succeed (see Danto 1965: Chap. 8). So philosophy must acknowledge and retain a narrative dimension, since it concerns processes of enlightenment, the analytic search for conditions of intelligibility. Indeed, the very notion of a problem in mathematics is historical, and this claim stems from taking as central and irreducible (for example) the narrative of Andrew Wiles’ analysis that led to the proof of Fermat’s Last Theorem. If mathematical problems have an irreducibly historical dimension, so too do theorems (which represent solved problems), as well as methods and systems (which represent families of solved problems): the logical articulation of a theory cannot be divorced from its origins in history. This claim does not presume to pass off mathematics as a series of contingencies, but it does indicate in a critical spirit why we should not try to totalize mathematical history in a formal theory.

W. B. Gallie begins his book Philosophy and the Historical Understanding by rejecting the Hegelian/Marxist doctrine of an inescapable order, a rational purposive pattern in history that might lead us to predict the end of history, when all social contradictions shall be resolved. He reviews the partial insights of Cournot, Dilthey, Rickert, and Collingwood, and then remarks upon a set of logical positivist studies on the subject of historical explanation (see Gardiner 1961). Gallie writes, “These studies are, broadly speaking, so many exercises in applied logic: their starting point is always the general idea of explanation, and they tend to present historical explanations as deviant or degenerate cases of other logically more perfect models” (Gallie 1964: 19). Later in the book, he elaborates on this point: “There has been a persistent tendency, even in the ablest writers, to present historical explanations as so many curiously weakened versions of the kind of explanation that is characteristic of the natural sciences. To speak more exactly, it is claimed or assumed that any adequate explanation must conform to the deductivist model, in which a general law or formula, applied to a particular case, is shown to require, and hence logically to explain, a result of such and such description” (Gallie 1964: 105).

By contrast, Gallie argues that history is first and foremost narrative, recounts of human actions that concern the historian and his readers; and that the ending of a story is “essentially a different kind of conclusion from that which is synonymous
with ‘statement proved’ or ‘result deduced or predicted’” (Gallie 1964: 23). A good story always contains surprises or reversals that could not have been foreseen. “But whereas for a scientist a revealed discontinuity usually suggests some failure on his part, or on the part of his principles and methods and theories, to account for that aspect of nature which he is studying, for the man who follows a story a discontinuity may mean… the promise of additional insights into the stuff a particular character is made of, into the range of action and adaptation which any character can command” (Gallie 1964: 41). Indeed, Gallie adds, this is the logical texture of everyday life, where the unforeseen constantly puts to the test our intellectual and moral resources, and where our ability to rise to the occasion must always remain in question: the insight of tragedy is that anyone can be destroyed by some unfortunate combination of events and a lapse in fortitude or sympathy.

Thus, the most important relation between events in a story is “not indeed that some earlier event necessitated a later one, but that a later event required, as its necessary condition, some earlier one. More simply, almost every incident in a story requires, as a necessary condition of its intelligibility, its acceptability, some indication of the kind of event or context which occasioned or evoked it, or, at the very least, made it possible. This relation, rather than the predictability of certain events given the occurrence of others, is the main bond of logical continuity in any story” (Gallie 1964: 26). Unless one is reading a detective novel, which is not so much a story as a puzzle in narrative clothing, following a narrative is emphatically not a kind of deduction or even induction in which successively presented bits of evidence allow the reader to predict the ending: “Following is not identical with forecasting or anticipating, even although the two processes may sometimes coincide with or assist one another” (Gallie 1964: 32). Following a story is more like the process of analysis. The intrusion of contingency into a story makes it lifelike; the way we retrospectively make sense of or redeem contingencies in our lives, as in the way we understand stories, is a good example of how we lend intelligibility to things. The command to search for conditions of intelligibility is not just a theoretical but also a practical or moral command; and we honor it not just by constructing arguments but also by telling stories.

This view of history has two important, and closely related, philosophical consequences. The first is that although historians and human beings in their attempts to come to moral maturity must pursue objectivity, *wie es eigentlich gewesen war*, in order to escape the provincialism and prejudices that mar histories, they cannot give this ideal any descriptive content: it is a regulative ideal, to borrow a Kantian term. “The historian is committed to the search for interconnectedness and is thus drawn on by an ideal demand that expresses his ideal of the whole, of the one historical world. But at the same time, because of the inevitable selectiveness of all historical thinking, it is impossible that he should ever reach… that ideal goal” (Gallie 1964: 61). There can be no such thing as the Ideal Chronicle.

The second is that the physicalist-reductionist account of human action which supposes such an Ideal Chronicle recording events as they happen, in complete and accurate detail, must be renounced, as Danto argues. There are two ways to run the thought experiment. Suppose that the chronicle is written in a language rich enough
to include the ways in which historians normally pick out, characterize, and link events. (This is a generous concession to the physicalist-reductionist, who is not really entitled to it.) This language contains a whole class of descriptions that characterize agents and actions in terms of their future vicissitudes as well as words like ‘causes’, ‘anticipates’, ‘begins’, ‘precedes’, ‘ends’, which no historian could forego without lapsing into silence. But such descriptions and words are not available to the eyewitness of events who describes them on the edge of time at which they occur. The description of an event comes to stand in different relations to those that come after it; and the new relations in turn may point to novel ways of associating that event with contemporary and antecedent events or indeed to novel ways of construing the parts of a spatio-temporally diffused event as one event. The historian’s way of choosing beginnings and endings for narratives is her prime way of indicating the significance of those events.

Thus to allow a sufficiently rich language for the Ideal Chronicle violates the original supposition of how the chronicle is to be written. So let us hold the physicalist-reductionist to his own restrictions: the chronicle must be written in terms impoverished enough to meet the stringent conditions of its writing on the edge of time. Then we find that it is reduced to an account of matter in motion, and the subject matter of history, people and their actions, has dissolved (as it dissolves in Lucretius’ epic De rerum natura) (Lucretius 2007). The chronicle is no longer about history, and the chronicler is doing descriptive physics if he is doing anything at all. If we want to write history, an event must have a beginning, middle, and end; must be related to its past and future; and must be construed as significant.

2 Kitcher and Cavaillès on History

In short, history is not science, historical explanation is not scientific explanation, historical method is not scientific method, and human action is not (merely) an event in nature. If philosophers of science and mathematics wish to make serious use of the history of their subjects, they must take these distinctions into account; indeed, this accounting will be helpful because philosophy that takes history seriously cannot pretend to be a science. In fact, philosophy, like mathematics, is neither history nor science, but involves both narrative (like history) and argument (like science), because the process of analysis involves both; the philosophical school that tries to banish either dimension will not succeed.

Up until the last couple of decades, history has been a term conspicuously absent in Anglo-American philosophy of mathematics (see Della Rocca 2016 for a compelling explanation). The philosophy of mathematics has seemed to have little to do with the philosophy of history, and the way in which history holds together the modalities of possibility, contingency, and necessity in human action. Even recent works that bring the history of mathematics to bear on the philosophy of mathematics in one way or another contain little philosophical discussion of history as such: they instead take their lead from current discussion in philosophy of science,
where historical episodes are taken to be evidence for philosophical theses, or instantiations of them. That is, the relation between episodes in the history of mathematics to philosophical theses about mathematics is understood in terms of a scientific model of the relation between scientific theory and empirical fact. The problem with this model is that it is itself timeless and ahistorical, that is, it construes the history of mathematics as a set of facts that can support or instantiate (or falsify) a theory, and be explained (or predicted) by a theory; thus the very historicity of mathematics is lost to philosophical view (see Brown 2008; Gillies 1993; Kitcher 1983; Maddy 2000).

This difficulty is compounded by the epistemology that most philosophers of mathematics have chosen to work with to undergird their accounts of mathematics. The first, an empiricist ‘naturalized epistemology’, begins with an appeal to perceptual processes that, while temporal, are not historical. They are construed as processes in nature, like those described by physics or biochemistry: the impingement of light on organs, the transmission of electrical impulses in neurons, etc., so although they happen in time, they take place in the same way universally, in all times and locations. They are not historical because they do not share in the peculiarity, idiosyncrasy, and irreversibility of historical character. The other, rationalist alternative, truth as formal proof, evidently stands not only outside of history but of time as well.

Mathematics is famous for lifting human beings above the contingencies of time and history, into a transcendent, infinitary realm where nothing ever changes. And yet philosophy of mathematics requires philosophy of history, because the discovery of conditions of intelligibility takes place in history, and because we lend ourselves to the intelligibility of things in undertaking the search. If we can borrow Plato’s simile of ascending from the cave and Leibniz’s simile of attaining a wider and more comprehensive point de vue of a city without the totalizing supposition of The Good, or of God, we can begin to see how the search for conditions of intelligibility takes place in history. If we can borrow Hegel’s dialectic or Peirce’s creative evolution of the universe without the totalizing supposition of the end of history, then we can begin to see how we lend ourselves to the intelligibility of things. And if we can invoke reason or intelligibility without trying to set pre-established limits for it, then we might arrive at a philosophy of history useful for understanding mathematics.

What use should the philosophy of mathematics make of history? Philip Kitcher, in his book The Nature of Mathematical Knowledge, just cited, tries to bring the philosophy of mathematics into relation with the history of mathematics, as a quarter of a century earlier Kuhn, Toulmin, and Lakatos aimed to do for the philosophy of science. Yet his approach is quite ahistorical. The epistemology he offers has no historical dimension: it is, he claims, a defensible empiricism. “A very limited amount of our mathematical knowledge can be obtained by observations and manipulations of ordinary things. Upon this small basis we erect the powerful general theories of modern mathematics”, he observes, hopefully, and adds, “My solution to the problem of accounting for the origins of mathematical knowledge is to regard our elementary mathematical knowledge as warranted by ordinary sense
perception”. He does admit that “a full account of what knowledge is and of what types of inferences should be counted as correct is not to be settled in advance…” especially since most current epistemology is still dominated by the case of perceptual knowledge and restricted to intra-theoretic reasoning (Kitcher 1983: 92–97).

However, his own epistemological preliminaries seem to be so dominated and restricted: “On a simple account of perception, the process would be viewed as a sequence of events, beginning with the scattering of light from the surface of the tree, continuing with the impact of light waves on my retina, and culminating in the formation of my belief that the tree is swaying slightly; one might hypothesize that none of my prior beliefs play a causal role in this sequence of events…. A process which warrants belief counts as a basic warrant if no prior beliefs are involved in it, that is, if no prior belief is causally efficacious in producing the resultant belief. Derivative warrants are those warrants for which prior beliefs are causally efficacious in producing the resultant belief”. A warrant is taken to refer to processes that produce belief in the right way. Then “I know something if and only if I believe it and my belief was produced by a process which is a warrant for it” (Kitcher 1983: 18–19). This is an account of knowledge with no historical dimension. It also represents belief as something that is caused, for a basic warrant is a causal process that produces a physical state in us as the result of perceptual experience and which can (at least in the case of beliefs with a basic warrant) be engendered by a physical process.

In a later article, “Mathematical Progress”, Kitcher makes a different, rather more pragmatic claim about mathematical knowledge, characterizing ‘rational change’ in mathematics as that which maximizes the attainment of two goals: “The first is to produce idealized stories with which scientific (and everyday) descriptions of the ordering operation that we bring to the world can be framed. The second is to achieve systematic understanding of the mathematics already introduced, by answering the questions that are generated by prior mathematics”. He then proposes a concept of “strong progress”, in which optimal work in mathematics tends towards an optimal state of mathematics: “We assume that certain fields of mathematics ultimately become stable, even though they may be embedded in ever broader contexts. Now define the limit practice by supposing it to contain all those expressions, statements, reasonings, and methodological claims that eventually become stably included and to contain an empty set of unanswered questions” (Kitcher 1988: 530–531). So there are two assumptions that render Kitcher’s account ahistorical. One is that mathematical knowledge has its origins in physical processes that cause fundamental beliefs in us (and these processes, while temporal, are not historical). The other is that mathematics should optimally end in a unified, universal, axiomatized system where all problems are solved and have their place as theorems. This unified theory has left history behind, like Peirce’s end of science or Hegel’s end of history, so that history no longer matters.

The intervention of history between the ahistorical processes and objects of nature, and the ahistorical Ultimate System (related to the former as theory to model, thus by the ahistorical relationship of instantiation) seems accidental. Kitcher puts all the emphasis on generalization, rigorization, and systematization,
processes that sweep mathematics towards the Ultimate System, with its empty set of unanswered questions. The philosophy of mathematics of Jean Cavaillès, provides an instructive contrast here. The method of Cavaillès, like that of his mentor Leon Brunschvicg, is historical. He rejects the logicism of Russell and Couturat, as well as the appeal of Brouwer and Poincaré to a specific mathematical intuition, referring the autonomy of mathematics to its internal history, “un devenir historique original” which can be reduced neither to logic or physics (Cavaillès 1937). In his historical researches (as for example into the genesis of set theory), Cavaillès is struck by the ability of an axiomatic system to integrate and unify, and by the enormous autodevelopment of mathematics attained by the increase of abstraction. The nature of mathematics and its progress are one and the same thing, he thinks, for the movement of mathematical knowledge reveals its essence; its essence is the movement.

Cavaillès always resists the temptation to totalize. History itself, he claims, while it shows us an almost organic unification, also saves us from the illusion that the great tree may be reduced to one of its branches. The irreducible dichotomy between geometry and arithmetic always remains, and the network of tranversal links engenders multiplicity as much as it leads towards unification. Moreover, the study of history reminds us that experience is work, activity, not the passive reception of a given (see Sinaceur 1994: 11–33; Sinaceur 2013: Chap. 6; see also Wagner 2016: Chap. 1). Thus for Cavaillès a mathematical result exists only as linked to both the context from which it issues, and the results it produces, a link which seems to be both a rupture and a continuity.

3 Reference in Mathematics

Current philosophical discussion of reference in mathematics is a bit hard to characterize. Sometimes the problem of model construction is substituted for the problem of reference; this move is favored by anti-realist philosophers. Thus theories are about models, though this leaves open the issue of how models refer, or if they do; and models are not individuated in the way that mathematical things typically are individuated. Bertrand Russell argued a century ago that the reference of a name is fixed by a proper definite description, an extensionally correct description which picks out precisely the person or natural kind intended (Russell 1905: 479–493). And W. V. O. Quine argued half a century ago that the ontological commitment of a first order theory is expressed as its universe of discourse (Quine 1953/1980: 1–19). But first order theories do not capture the objects they are about (numbers, figures, functions) categorically, and the ‘ontological commitments’ of higher order theories are unclear. Saul Kripke insisted that we need the notion of an initial baptism (given in causal terms), followed by an appropriate causal chain that links successive acts of reference to the initial act, for only in this case would the name be a ‘rigid designator’ across all possible worlds; a rigid designator picks out the correct person or natural kind not only in this world but in all possible worlds.
where the person or kind might occur (Kripke 1980). Hilary Putnam argued that the ability to correctly identify people and natural kinds across possible worlds is not possessed by individuals but rather by a society where epistemic roles are shared (Putnam 1975). And Paul Benacerraf argued in a famous essay that linking reference to causal chains makes an explanation of how mathematics refers seem futile (Benacerraf 1965).

In sum, it is not generally true that what we know about a mathematical domain can be adequately expressed by an axiomatized theory in a formal language; and it is not generally true that the objects of a mathematical domain can be mustered in a philosophical courtyard, assigned labels, and treated as a universe of discourse. What troubles me most about this rather logicist picture is that the difficulty of integrating or reconciling the two tasks of analysis and reference (as well as the epistemological interest of such integration) is not apparent, since it is covered over by the common logical notions of instantiation and satisfaction.

The assumption seems to be that all we need to do is assign objects and sets of objects from the universe of discourse (available as a nonempty set, like the natural numbers) to expressions of the theory. If we carry out the assignment carefully and correctly, the truth or falsity of propositions of the theory, vis-à-vis a ‘structure’ defined in terms of a certain universe of discourse, will be clear. In a standard logic textbook, the universe of discourse is the set of individuals invoked by the general statements in a discourse; they are simply available. And predicates and relations are treated as if they were ordered sets of such individuals. In real mathematics, however, the discovery, identification, classification and epistemic stability of objects are problematic; objects themselves are enigmatic. It takes hard work to establish certain items (and not others) as canonical, and to exhibit their importance. Thus reference is not straightforward. Moreover, of course, neither is analysis; the search for useful predicates and relations proceeds in terms of intension, not extension, and the search for useful methods and procedures cannot be construed extensionally at all. Analysis is both the search for conditions of intelligibility of things and for conditions of solvability of the problems in which they figure. We investigate things and problems in mathematics because we understand some of the issues they raise but not others; they exist at the boundary of the known and unknown.

My claim that mathematical objects are problematic (and so in a sense historical and in another sense strongly related to practices) need not lead to skepticism or anti-realism. We can argue, with Scott Soames inter alia, that natural language (English, French, German, etc.) provides us with an enormous amount of reliable information about the way the world is, what things are in it and how they turn up; and note that we act on this information as we reason (Soames 2007). Natural language, along with abstract mathematical structures, also allow us to bring the disparate idioms of mathematics (which are so different from natural language) into rational relation. So presumably mathematics enhances our information about the world, modifying without dismissing our everyday understanding. We can claim that discourse represents things well without becoming dogmatic, if we leave behind the over-simplified picture of the matching up of reference and analysis as the satisfaction of propositions in a theory by a structure.
Things, even the formal things of geometry and arithmetic, have the annoying habit of turning out to be irreducible to discourse. Things transcend discourse even though of course they lend themselves to it. But their irreducibility is what makes truth possible: we can’t just make things up or say whatever we want about things. A true discourse must have something that it is about. The history of both mathematics and science shows that things typically exhibit the limitations of any given discourse sooner or later, and call upon us to come up with further discourses. Because of their stubborn haeccty, things are, we might say, inexhaustible. Thus when we place or discover them in a novel context, bringing them into relation with new things, methods, modalities of application, theories, and so forth, they require revised or expanded discourses which reveal new truths about them.

The irreducibility of things should also lead us to pay closer attention to what I have called subject-discourses, how they are constituted and how they function (given that they must point beyond themselves). Given their function, they must be truer to the individuality or specificity of things than predicate-discourses need to be, since the function of predicate-discourses is to generalize. So they will be less systematic and well-organized, but more expressive, messy, precise, and surprising.

Another way of putting this is that any discourse that encompasses a broad domain of problems in mathematics or science will be internally bifurcated, since it must integrate a subject-discourse and a predicate-discourse. Due to the inherent disparity of such discourses, this unification will become unstable sooner or later. Moreover, any discourse fails to exhaust knowledge about the things it concerns; sooner or later those problematic, resistant things, and the novel (worldly and discursive) contexts that may arise with respect to them, will generate problems that the original discourse can’t formulate or solve. Indeed, the very desire to solve problems via problem-reduction may bring about the situations that force the revision, replacement and extension of discourse. We often place things in novel contexts, or come to recognize that novel contexts can be made to accommodate familiar things, in order to elicit new information. In sum, the model of theory reduction, examined in detail in Chap. 4, does not do justice to the internal bifurcation of both the reduced and reducing theories, and it does not capture the complex relations between them; and so it does not attend properly to the ways in which knowledge grows. Processes of problem-reduction should be examined by philosophers alongside processes of theory-reduction; and the ways in which these processes are ampliative should be recognized.

### 4 Wiles’ Proof of Fermat’s Last Theorem

Here I return to the story of Andrew Wiles and his proof of Fermat’s Last Theorem. On the one hand, we have a narrative about an episode in the life of one man (in a community of scholars) who, inspired by a childhood dream of solving Fermat’s Last Theorem, and fortified by enormous mathematical talent, a stubborn will, and the best number theoretical education the world could offer, overcame various
obstacles to achieve truly heroic success. Indeed, the most daunting and surprising obstacle arose close to the end, as he strove to close a gap discovered in the first draft of his proof. On the other hand, we have a proof search which can be mapped out and reversed into the full dress proof, though it is important to recall that the proof is not located in any single axiomatized system. It makes use not only of the facts of arithmetic and theorems of number theory (both analytic number theory and algebraic number theory), but also results of group theory (specifically Galois theory); and it exploits the system of $p$-adic numbers (offspring of both topology and group theory), representation theory, deformation theory, complex analysis, various algebras, topology, and geometry. This proof search has its own location in history, which must be distinguished from that of Wiles’ life, for it constitutes a path backwards through mathematical history (where earlier results make later results possible, and where new results bring earlier results into new alignments) and a leap that is also a rupture opening onto the future, making use of older techniques in novel ways to investigate a conjecture that many number theorists in fact worried could not be proved by the means available at the time.

Wiles’ proof of Fermat’s Last Theorem relies on verifying a conjecture born in the 1950s, the Taniyama-Shimura conjecture, which proposes that every rational elliptic curve can be correlated in a precise way with a modular form. (It is a nice example for Carlo Cellucci’s philosophical approach, discussed in Chap. 3, because Fermat’s Conjecture is true, if the Taniyama-Shimura conjecture is true, and this turns out to be a highly ampliative reduction.) It exploits a series of mathematical techniques developed in the previous decade, some of which were invented by Wiles himself. Fermat wrote that his proof would not fit into the margin of his copy of Diophantus’ *Arithmetica*; Wiles’ 108 pages of dense mathematics certainly fulfills this criterion. Here is the opening of Wiles’ proof: “An elliptic curve over $\mathbb{Q}$ is said to be modular if it has a finite covering by a modular curve of the form $X_0(N)$. Any such elliptic curve has the property that its Hasse-Weil zeta function has an analytic continuation and satisfies a functional equation of the standard type. If an elliptic curve over $\mathbb{Q}$ with a given $j$-invariant is modular then it is easy to see that all elliptic curves with the same $j$-invariant are modular… A well-known conjecture which grew out of the work of Shimura and Taniyama in the 1950s and 1960s asserts that every elliptic curve over $\mathbb{Q}$ is modular… In 1985 Frey made the remarkable observation that this conjecture should imply Fermat’s Last Theorem. (Frey 1986). The precise mechanism relating the two was formulated by Serre as the $\varepsilon$-conjecture and this was then proved by Ribet in the summer of 1986. Ribet’s result only requires one to prove the conjecture for semistable elliptic curves in order to deduce Fermat’s Last Theorem” (Wiles 1995: 443). For my brief exposition of the proof here, in Chap. 5 and in Appendix B, I relied upon the original article as well as didactic expositions and my own class notes (Darmon et al. 1997; Li 2001, 2012, 2013, 2014; Ribet 1995).

In number theory, as we have seen, the introduction of algebraic notation in the early seventeenth century precipitates the study of polynomials, algebraic equations, and infinite sums and series, and so too procedures for discovering roots and for determining relations among roots or between roots and coefficients, and ways
of calculating various invariants. The use of abstract algebra (groups, rings, fields, etc.) in the late nineteenth and early 20th centuries leads to the habit of studying the symmetries of algebraic systems as well as those of geometrical items, finitary figures and infinitary spaces. The habit of forming quotients or ‘modding out’ one substructure with respect to its parent structure often produces a finitary structure with elements that are equivalence classes, from two quite infinitary structures. This habit in turn suggests the use of two-dimensional diagrams characteristic of (for example) deformation theory, where the relations among the infinitary (or very high-dimensional) and the finitary are displayed in what might be called iconic fashion. Abstract algebra also produces the habit of seeking in the relation of structure to substructure other, analogous relations in different kinds of structure and substructure. For example, the Fundamental Theorem of Galois Theory tells us that, when $G$ is the Galois group for the root field $N$ of a separable polynomial $f(x)$ over a field $F$, then there is a one-one correspondence between the subgroups of $G$ and the subfields of $N$ that contain $F$. And Representation Theory instructs us to seek groups of matrices that will mimic in important ways the features of other infinitary and less well-understood groups of automorphisms. Abstract algebra also suggests the investigation of a given polynomial over various fields, just to see what happens, as modern logic (treated algebraically) suggests the investigation of non-standard models, just to see what they are like.

So what is an aspect of reference for one number theorist, like Barry Mazur who takes his orientation from algebraic topology and cohomology theory, may have played the role of analysis for other, more traditional number theorists like Eichler and Shimura, who begin from the arithmetic theory of Abelian varieties. What preoccupies one number theorist may remain tacit for another, and vice versa, so that the combination of their results (as in the case of Wiles’ proof) forces the articulation of ideas which had up till then remained out of focus, beyond the horizon of attention. Likewise, what remains tacit for the number theorist may be articulated by the logician, as we shall see in Chap. 5. For what remains tacit in one approach (given the strengths and limitations of a given mathematical idiom) must often be made explicit in another in order to bring the two approaches into productive relation, as novel strategies of integration are devised.

I will sketch the proof of Fermat’s Last Theorem in terms of two stages, briefly. The first stage concerns the result of Eichler-Shimura, which proves that given a certain kind of modular form, we can always find a corresponding elliptic curve. (This stage is explained at length in Chap. 5, with a glossary (and adumbrated in Appendix B), and its philosophical implications explored.) The second stage concerns Wiles’ result, proving the Taniyama-Shimura conjecture, that given a certain kind of elliptic curve, we can always find a certain kind of modular form. (To explain this stage, I would have to write another book; Appendix A offers some useful historical background.) Frey conjectured and Ribet proved that Fermat’s Last Theorem follows from this correspondence, carefully qualified. (Ribet shows that the correspondence rules out the possibility of counterexamples to Fermat’s Last Theorem; see Ribet 1990). The strategy that figures centrally in the Eichler-Shimura proof is the strategic use of $L$-functions (generalizations of the Riemann zeta
function, and Dirichlet series), where given a certain kind of modular form $f$ we have to construct a corresponding, suitably qualified, elliptic curve $E$. Another equally important strategy is to use representation theory in tandem with deformation theory, where $p$-adic families of Galois representations figure centrally in the proof of the Taniyama-Shimura conjecture. Given a certain kind of elliptic curve $E$, we investigate $p$-adic representations in order to construct a corresponding, suitably qualified, modular form $f$.

5 Wiles’ Analysis Considered as History

Andrew Wiles’ fascination with Fermat’s Last Theorem began when he was 10 years old, and culminated on the morning of September 19, 1994, when he finally put the last piece of the grand puzzle in place. In order to establish the isomorphism between $T_{\Sigma}$ and $R_{\Sigma}$, he had tried to use an approach involving ‘Iwasawa theory’, but that had been unsuccessful; then he tried an extension of the ‘Kolyvagin-Flach method’, but that attempt had stalled. While trying to explain to himself why this new approach didn’t seem to be working, he realized (inspired by a result of Barry Mazer’s) that he could use Iwasawa theory to fix just that part of the proof where the Kolyvagin-Flach approach failed; and then the problem would be solved. On that morning, something happened that was radically unforeseeable (even by Wiles, who was very discouraged and did not believe it would happen), and yet, once it actually took place, presented the kind of necessity that mathematical results present. It disrupted mathematics by changing its transversal relations, for now modular forms were proved to be correlated in a thoroughgoing way with elliptical curves, and at the same time established a new order. The unforeseeability was not merely psychological, subjective, and merely human; the disruption lay in the mathematical objects as well as in the mind of the mathematician.

What Wiles did on that morning can only be explained in terms of the mathematics. As Cavaillès argues, “I want to say that each mathematical procedure is defined in relation to an anterior mathematical situation upon which it partially depends, with respect to which it also maintains a certain independence, such that the result of the act [geste] can only be assessed upon its completion” (Cavaillès and Lautmann 1946: 9). A mathematical act like Wiles’ is related to both the situation from which it issues and the situation it produces, extending and modifying the pre-existing one. It is both a rupture and a continuation, an innovation and a reasoning. To invent a new method, to establish a new correlation, even to extend old methods in novel ways, is to go beyond the boundaries of previous applications; and at the same time in a proof the sufficient conditions for the solution of the problem are revealed. What Cavaillès calls the fundamental dialectic of mathematics is an alliance between the necessary and the unforeseeable: the unforeseeability of the mathematical result is not appearance or accident, but essential and originary; and the connections it uncovers are not therefore contingent, but truly necessary.
Another way of describing Cavaillès’ insight is to say that he is trying to uncouple the connection between necessity and the Kantian a priori, which offers only Kantian analysis, the unpacking of what is already contained in a concept, or synthesis, which must be referred to the mind of the knower. What happened when Wiles finally proved Fermat’s Last Theorem? Just at that point, the unsolved problem was solved, the unforeseeable flipped over and was seen at last, the indeterminately possible became the determinately necessary. It was at once an event in the biography of Andrew Wiles: his alone was the consciousness in which this amazing peripety or reversal took place, this discovery, a change from ignorance to knowledge. No one else could have shared that discovery as it happened for the first time, that singular event, for no one can inhabit the mind of another: as Leibniz said, we are windowless monads. Yet both the dramatic structure the act already possessed, and the argumentative structure inherent in the proof underlay the story Wiles recounted over and over the next day, to himself (checking the proof), then to his wife, then to his colleagues, then to the world, in different fashions. And the story-argument didn’t change thereafter, as it was reenacted in the thoughts of those mathematicians who knew enough number theory to check the proof, all of whom found it successful.

The retelling, which includes his narrative of the proof-search, and the published 108 page proof, is marked by the idiosyncracy and irrevocability of that historical moment in one obvious way: Wiles could only make use of results that had been discovered up to that point in history when he finished devising his proof. The Taniyama-Shimura conjecture requires the availability of modular forms, which rests on the work of Poincaré in exhibiting their infinite symmetry, which requires the work of Klein and Riemann in formulating hyperbolic space as devised by Lobachevsky. What is requisite for formulating and solving a problem lies only in the past, made available by instruction or textbook. Fermat could formulate the problem of the Last Theorem, but despite his boast he could not have solved it.

6 The Math Genie

The incoherence of the notion of an Ideal Chronicle bears not only on the reality of human beings and their acts, but on the possibility of giving a complete speech about the totality of human action because the very description of an event is interpretive, because one cannot eliminate from the description of events terms that link it to both its past and its future, and because the description of an event changes with time as the event comes to stand in different relations to events that come after it. Likewise, the notion, suggested by Kitcher, of some Ultimate System in mathematics is also incoherent, not because the reality or intelligible unity of mathematical things is doubtful, but because we cannot give a complete speech or formal theory about them. The very description of a mathematical object is interpretive, because it is given against a background of antecedent knowledge and by means of a certain notation; one cannot eliminate from its description terms that link it to past
problems and problems still to be solved; and the relations and correlations in which it stands to other objects change over time, as new objects are discovered and older objects are forgotten.

Like perceived things, mathematical things are problematic. Just as perceived things call for analysis in order to uncover the conditions of their intelligibility, so mathematical things call for analysis in order to uncover the conditions of solvability of the problems in which they are always embedded. But to be problematic is an historical feature. Objects are problematic when we understand enough about them to see with some precision what we don’t yet know about them. And as soon as we learn something new about them, in virtue of that very discovery they typically come to stand in novel, unforeseen relations with other objects that make them problematic again. As I showed in Chap. 1, when the classical problems concerning the circle were solved during the 17th century by the new analytic geometry and the infinitesimal calculus, those same discoveries relocated the circle in relation to transcendental curves (especially sine and cosine), the definition of curvature and the generalized notion of a surface, etc., and re-embedded it in a host of new problems. We know what we know about the circle up to this point in history, with the means at our disposal as those means have been deployed; and we can’t yet know other things about it, though we can question, postulate, conjecture, hypothesize, acts that project us towards that future though we are not quite yet there. Asking questions and making conjectures is a way of approaching knowledge that we do not yet have: so analysis is not just a pathway into the history of mathematics (though it is that), but also an unfinished bridge to the future. The relation between a problem and its conditions of solvability may be lifted, as it were, out of history: this is what we do when we turn the search for a proof into a proof. Yet the fact that the problem was a problem and is now a solved problem, is a fact that belongs as much to history as to mathematics. Insofar as every theorem may be said to be a solved problem, the same holds true for theorems (see Hersh 1999).

There are no problems without problematic things; problems exist in mathematics because we encounter things that trouble us. What would a problem be about if it weren’t about some thing? What would a problem be without its aboutness? Versions of Fermat’s Last Theorem existed before Fermat, as a range of problems about positive whole numbers (or rather, triples of them); Fermat turned it into a problem about a set of polynomial equations; Andrew Wiles turned it into a problem about a correlation between elliptic curves and modular forms. The evolution of the problem depends on rational relations among different kinds of objects: numbers serve as conditions of intelligibility for polynomial equations, and the latter for the Taniyama-Shimura correlation. Wiles’ result can be read backwards, to hold for the equations and the relevant number-triples; the aboutness of a problem may change, but the aboutness-apropos-earlier-things remains as a condition of the intelligibility of the problem. And there are no things without problems. Things, even they serve as conditions of intelligibility of other things, don’t wear the conditions of their own intelligibility on their faces; it is always a problem for reflection to find the conditions of intelligibility of a thing.
Indeed, if a mathematician could magically reach into the future and bring back future results, then there would be no activity called solving problems. The appearance of mathematical problems as problems requires history, and history demands our patience as we wait to see how things will turn out. Let us suppose that there is an Ultimate System, a complete system of mathematics independent of the accidents of history, presided over by the Math Genie. This genie comes to the aid of mathematicians and maybe even philosophers who rub the relevant lamp: he brings the solution to any problem that troubles you in the form of information about the objects involved. He can violate history because he has access to the Ultimate System. Imagine what this genie might have done for Fermat in the mid-seventeenth century, when in fact the mathematical resources for proving his Last Theorem were lacking! He would have set out Andrew Wiles’ proof for him, but of course to ensure that Fermat understood the proof, the genie would have had to teach him, perhaps in a series of seminars, all the 18th, 19th, and 20th c. mathematics linking what he knew to what Wiles knew, perhaps keeping him alive by philtres until the process could be completed. In this case, the genie would have to offer information, not just about numbers and equations, but about the correlation between modular forms and elliptic curves. He would have to bring numbers into rational relation with objects that, given the constraints of 17th c. mathematics, were not even thinkable.

However, we have just told the story from the point of view of the Math Genie (and tacitly assumed that he is located in our era, cleverly disguised as eternity). We must tell it from Fermat’s point of view; but then we see that Fermat could not have made his request successfully in the first place. Suppose that Fermat had asked the Math Genie to bring back the solutions to his problem, as he himself enunciated it, for the genie might have required that all requests be precisely specified. In that case, he could hardly have brought back Wiles’ result, for though Wiles showed that it entailed a proof of Fermat’s Last Theorem, Fermat could never have asked the genie for a proof of the Taniyama-Shimura conjecture and its reduction to his problem. Alternatively, the genie might have acceded to his general request to bring back problems related to the natural numbers: but then he would have had to go into the future and bring back all such problems involving all the new objects that include the natural numbers in their genealogy. This is a limitless prospect: by now we have seen enough of analysis to know that an analysis typically uncovers new objects and problems: there is always tacit knowledge at the metalevel to enunciate, new generalizing abstractions to create (and thereby to lose or forget other things), new correlations to explore, and so forth. Fermat would have been swamped.

So either Fermat can’t ask for what he wants; or to the extent that he can ask for it, the genie can only offer him an unsurveyable infinity, without any kind of closure, of solved problems. The incoherence of supposing that mathematics can be liberated from the accidents of history, that all problems might be solved and an Ultimate System projected in which “the set of unanswered problems would be empty” shows that the historical location of a problem and the way in which the objects involved in it are problematic is not accidental but essential to the problem as such. Problems can only appear in a situation where some things are known and some things are not yet known; the enunciation of a problem is just saying what
precisely is not known against the background of what has been discovered so far, and suggestions about how to proceed to solve the problem require even greater precision. Moreover, problems and their solutions are the articulation of mathematics: they provide it with the intelligible structure that may be written afterwards as theorems and axioms that organize theorems. The Math Genie is a useful fiction, like Descartes’ Evil Demon, to show the incoherence of an idea, in this case, that of mathematics without problems. The thought experiment just entertained shows at least two reasons why there cannot be a complete speech about mathematical things, any more than there can be a complete speech about human actions. (And this is no more a reproach to the reality of mathematical things than it is to the reality of who we human beings are and what we do.) One is that there are many different kinds of mathematical things, which give rise to different kinds of problems, methods, and systems. The other is that mathematical things are investigated by analysis, which is a process at once logical and historical in which some things, or features of things, that were not yet foreseen are discovered, and others are forgotten. Indeed, these two aspects of mathematics are closely related. For when we solve problems, we often do so by relating mathematical things to other things that are different from them, and yet structurally related in certain ways, as when we generalize to arrive at a method, or exploit new correlations. We make use of the internal articulation or differentiation of mathematics to investigate the intelligible unities of mathematics. To put it another way, just as there is a certain discontinuity between the conditions of solvability of a problem in mathematics and its solution (as Cavaillé noted), so there is a discontinuity between a thing and its conditions of intelligibility (as Plato noted). An analysis results in a speech that both expresses, and fails to be the final word about, the thing it considers.

References


Starry Reckoning: Reference and Analysis in Mathematics and Cosmology
Grosholz, E.R.
2016, XVIII, 187 p. 15 illus., Hardcover
ISBN: 978-3-319-46689-7