DO ANDROIDS PROVE THEOREMS IN THEIR SLEEP?

Michael Harris, version of March 21, 2008, (plus one addition)

1. A mathematical dream narrative

...antes imagino que todo es ficción, fábula y mentira, y sueños contados por hombres despiertos, o, por mejor decir, medio dormidos. ***Don Quixote***, Book II, Chapter 1.

What would later be described as the last of Robert Thomason's "three major results" in mathematics was published as a contribution to the Festschrift in honor of Alexandre Grothendieck's sixtieth birthday, and was cosigned by the ghost of his recently deceased friend Thomas Trobaugh. Thomason described the circumstances of this collaboration in the introduction to their joint article: a rare note of pathos in the corpus of research mathematics, and a brief but, I believe, authentic contribution to world literature.

> *The first author must state that his coauthor and close friend, Tom Trobaugh, quite intelligent, singularly original, and inordinately generous, killed himself consequent to endogenous depression. Ninety-four days later, in my dream, Tom's simulacrum remarked, "The direct limit characterization of perfect complexes shows that they extend, just as one extends a coherent sheaf." Awaking with a start, I knew this idea had to be wrong, since some perfect complexes have a non-vanishing K_0 obstruction to extension. I had worked on this problem for 3 years, and saw this approach to be hopeless. But Tom's simulacrum had been so insistent, I knew he wouldn't let me sleep undisturbed until I had worked out the argument and could point to the gap. This work quickly led to the key results of this paper. To Tom, I could have explained why he must be listed as a coauthor.* [1]

Thomason was already a CNRS researcher in Paris when this article was published; no separate address is given for Trobaugh. I arrived in Paris a few years later and we briefly overlapped at Jussieu. Thomason died suddenly of diabetic shock a few months later at age 43, five years after the publication of the Thomason-Trobaugh article,[2] which we only discussed once briefly and, I regret, superficially.

I urge the reader to ignore everything about the quotation from Trobaugh's ghost except the syntax, to treat the second half of the subsequent sentence in the same way, and to attempt to focus on the event structure of this short narrative. A ghost appears in a dream, offering a gift in the form of a cryptic message. Thomason, the dreamer, is the one person qualified to interpret the message; he has "worked on this problem for 3 years."

---

[1] Thomason and Trobaugh, Higher algebraic K-theory of schemes and of derived categories, in P. Cartier et al., eds, *The Grothendieck Festschrift Volume III*, Boston: Birkhâuser (1990) 247-429; quotation p. 249.
[2] See the biographical note by Charles A. Weibel published in the *Notices of the American Mathematical Society*, August 1996, 860-862.

This is a familiar plot, but what is actually happening? Trobaugh's ghost, soundless but articulate, imparted an insight to his friend before vanishing. The claim was nonsense. The dreamer interprets the ghost's sentence as an "idea." This idea is "wrong," even "hopeless." Nevertheless, as Thomason explains in this paragraph, and as an American Mathematical Society biographical note on Thomason confirms, this contribution was decisive.

The word "key" provides the key to our reading of this paragraph, of the Thomason-Trobaugh article as a whole, and of the light this incident sheds on the question of the title. Contemporary mathematical writing generally consigns what little pathos it allows, or any reflection of human experience whatsoever, to the introduction, but rhetorical devices are present on practically every line; without them, a mathematical argument would quickly become unreadable. The word "key" functions here to structure the reading of the article, to draw the reader's attention initially to the element of the proof the author considers most important. Compare E. M. Forster in *Aspects of the Novel*:

> [plot is] something which is measured not be minutes or hours, but by intensity, so that when we look at our past it does not stretch back evenly but piles up into a few notable pinnacles.

Weibel's biographical note periodizes the proof as a sequence of three steps. These steps are presented in roughly the order in which they were "discovered" by Thomason, but one can assume they also correspond to a reading of the proof Weibel found particularly helpful. [3] What I want to understand in this essay is what it entails to speak of "a reading of a proof" in mathematics.

It is typical of a mathematical narrative, and in this respect the Thomason-Trobaugh article is no exception, that one knows in advance how the story is going to turn out. This is the principal function of the introduction of a mathematical article, and it is one reason editors allow authors a certain amount of literary freedom in this section. Similarly, the narrative of an individual theorem begins with the statement of the conclusion. Though the conclusion may be recapitulated at the end of the proof — this is the case in some ancient Greek texts, where the recapitulation is called the *sumperasma*, a term we will again encounter in § 7 — the narrative is driven by the gradual discovery of the path by which the foregone conclusion is reached. In this way mathematical narrative differs from most narrative fiction, though one finds the same pattern in classic mystery novels.

The vocabulary of mathematical writing is extremely limited, often distressingly so, but there is some room for variety, particularly in the metalanguage in which mathematicians tend to speak of their results. Instead of "key" on might write "main," or "crucial," or "fundamental," or "essential:" all imprecise ways to point to a feature of mathematics familiar to all practitioners but scarcely dreamt of in philosophy, a matter to which I will return presently. We will stick with "key," the word of Thomason's choice, and a particularly apt one at that. The ghost's gift is almost literally a "key": the word

[3] *Ibid*., p. 861.

"key" is used three times in the introduction, and again on the occasion of the author's second invocation of the ghost, this time in the body of the text. I quote the passage *in extenso*, again asking the reader to attempt to read it for syntax alone:

> *The idea of 5.5.1 is that perfect complexes are finitely presented objects in the derived category 2.4.4, and so we may adapt Grothendieck's method of extending finitely presented sheaves ([EGA] I 6.9.1), as suggested by the Trobaugh simulacrum. While this adaptation does not allow us to extend all perfect complexes, it does lead quickly to the determination of which perfect complexes do extend.*
>
> *Despite the flagrant triviality of the proof of 5.5.1, this result is the key point in the paper.* [4]

A paper nearly 200 pages long, with a single "key point" just over halfway through. One is accustomed to thinking that a novel is meant to be read from front to back. This may be a mere artifact of typesetting, and who knows whether the paradigm of hypertext, prefigured by Cortazar's *Hopscotch*, will eventually displace linearity in composition as well as in strategies of reading? The word "key" as used by Thomason hints that mathematical arguments admit not only the linear reading that conforms to logical deduction but also a topographical reading that more closely imitates the process of conception.

Thomason's dream narrative can be read as an addition to a substantial literature on the role of the unconscious in scientific discovery. The most familiar specimen of the genre is Kekulé's (possibly fabricated) account of his dream about the benzene ring. In mathematics the classic incident is Poincaré's sudden realization, as he stepped up to the omnibus, of the relation between automorphic functions and non-Euclidean geometry. Poincaré explicitly attributed this discovery to the activity of the unconscious, and this incident figures prominently, alongside many others in Hadamard's study of the psychology of mathematical creativity.[5] Dreams, of course, have since antiquity been the subject of their own literature and have long been seen as particularly relevant to literary creation in general. Some fascinating questions I cannot hope to address:

*Where does the dream acquire its narrative structure? Is it intrinsic to the dream or to the retelling of the dream?

* Is the narrative structure of the proof derived from the dream already implicit in the dream or is it an artifact of the writing process?

In this article I am not interested in the dream narrative as such but in what the dream tells us about the narrative structure of a mathematical proof. Thomason's paragraph is able to direct our attention to something important about that structure precisely because

---

[4] Thomason and Trobaugh, *Op. cit.*, p. 343.

[5] Poincaré, *Science et hypothèse*, Hadamard, *An essay on the psychology of invention in the mathematical field*. Ramanujan and Grothendieck (*Dieu est le rêveur*) both attributed their inspirations to divine intervention.

he has adopted a narrative form — the sequence of enlightenment through dreaming common to literary traditions from all periods and all cultures, and the no less familiar but particularly moving invocation of the visit of a departed friend, made all the more poignant by the author's own unexpected departure so soon after publication of his text.

2. Mathematics from an android perspective

Heaven is a place where nothing ever happens. David Byrne
If everything in the universe were sensible, nothing would happen. (Dostoevsky, *Brothers Karamazov*)

Frege set the admissions requirements to the theater of reason beyond the means of merely human mathematicians. Philosophers of mathematics have ever since been dreaming of electric minds. Alongside the scientific literature on artificial intelligence there is a genre of speculative literature whose extreme versions present artificial beings as our evolutionary successors.

> *We are entering a new era. I call it "the Singularity." It's a merger between human intelligence and machine intelligence [that] is going to create something bigger than itself. It's the cutting edge of evolution on our planet. …To me that is what human civilization is all about.*[6]

These works develop certain tendencies latent in AI culture, and in philosophy of mathematics, to their logical conclusions, but I prefer to view this literature as a fictional genre, the more so given its significant overlap with traditional literary themes, as is clear from the title of one of the genre's precursors, Norbert Wiener's *God and Golem, Inc.*, (which, however, is anything but celebratory).

In this literature an independent AI proof of a significant theorem of mathematics is usually presented as a milestone. In this the futurists are in line with Herbert Simon's 1956 prediction that "within ten years, computers would beat the world chess champion, compose 'aesthetically satisfying' original music, and prove new mathematical theorems."[7] Michael Beeson's article *The Mechanization of Mathematics*, from which I have taken this quotation, argues that these milestones have all been met, though not in Simon's time frame. The 1996 proof by a computer program of the Robbins Conjecture on axioms for Boolean algebras, like the AI proofs that have followed, can be dismissed as marginal to the concerns of traditional mathematics. The computer programs that completed the proofs of the Four Color Theorem or the Kepler Conjecture, or calculated the Kazhdan-Lusztig polynomials of E8, were obviously too dependent on the guidance of human programmers to qualify. No one is going to spend much time arguing whether or not computer-generated music is "aesthetically satisfying." But everyone knows

---

[6] Ray Kurzweil, author of *The Singularity is Near*, quoted at http://www.edge.org/3rd_culture/kurzweil_singularity/kurzweil_singularity_index.html. Other classics of the genre include Hans Moravec *Mind Children* and Kurzweil's *The Age of Spiritual Machines*.

[7] In Beeson, p. 2

what happened when Garry Kasparov met Deep Blue, and it would be imprudent to argue the point. It's fair to say, nevertheless, that the futurists' speculations on these matters are notably unsatisfying as literature.[8]

Of the three "parts" of tragedy identified by Aristotle that transpose to fiction of all types — plot (*mythos*) , character (*ethos*), and "thought" (*dianoia*) — only the third is perceptible as such in this futurist literature (this is not the case in the article by Timothy Gowers discussed below, but Gowers is not a futurist and is interested in mathematics rather than in proposing milestones). If I want to understand speculation about future automatic theorem provers as a fictional genre — which is the same as imagining this material as the subject of a narrative — I need to recover the missing parts. A character playing the role of protagonist in an automatic theorem proving fiction — a fictional automatic theorem prover — I will call an *android*.

You may prefer to imagine an android with the features of the replicants Roy or Rachael, played by Rutger Hauer and Sean Young, in Ridley Scott's *Blade Runner*. But a theorem-proving android could equally well be the familiar cohort of monkeys with typewriters as in the "infinite monkey theorem" first proved (according to the distributed intelligence network Wikipedia) by the eminent French mathematician Emile Borel. The monkeys are not recruited for their intelligence but for their typing skills. The intelligence is concentrated in the typewriters: we assume they have the rules of inference built in and will not register a line unless it is a well-formed formula that follows from the preceding line. In other words, the typewriter incorporates a *proof assistant*, which is

> *…typically a program which can be run on an input file (usually text), and which certifies that (1) the file adheres to a specified syntax; (2) according to specified inference rules, the document contains the proofs (and constructions) that it purports to; and (3) any errors are located.*[9]

The medium, so to speak, of the proof is completely homogeneous. It is not punctuated by any "Aha!-Erlebnis" (K Bühler, 1908) nor is there any possibility of communication with this android.

In building the proof assistant into the typewriters, I am simply imagining a mechanical counterpart of the interpretation of mathematics as a formal language whose sentences are propositions, affirmative sentences constructed out of a finite collection of symbols, subject to certain rules of construction to avoid meaningless formulas, and whose "dialogues" are sequences of such propositions, each of which can be obtained by

---

[8] Not on other matters, however; Moravec and Kurzweil are particularly vivid when describing the mechanics of downloading an individual human consciousness or the "gray goo" scenario of nanobots run amok.

[9] M. Maggesi and C. Simpson, *Information Technology implications for mathematics, a view from the French riviera* http://math1.unice.fr/~maggesi/itmath/. The technical automated theorem-proving literature is vast but is mainly addressed to computer scientists. The cited article was written by mathematicians for mathematicians. I have also consulted the documents on T. Hales' Flyspeck project, discussed below, and especially M. Beeson's *The Mechanization of Mathematics*, a chapter in M. Teuscher's book *Alan Turing: Life and Legacy of a Great Thinker*, which Michael Beeson graciously made available to me.

transforming its predecessor according to one of a finite repertoire of rules of inference. This is the vision of the "mechanization of mathematics" (cf. Beeson, *op. cit.*) inherited from Frege and developed extensively by Hilbert and his collaborators, in the hope that any source of unreliability could be eliminated by a procedure of proof-checking that is mechanical in the sense of being perfectly rule-based and thus, in principle, implemented by a computer. Allowed to run endlessly, such a machine would eventually generate all possible proofs. We now think of computers as electronic rather than mechanical, which is why I insist on the typewriters. Gödel's work is often misinterpreted but certainly implies that it is impossible to deduce all true propositions in mathematics (as mathematics is usually understood) by such mechanical means. Gödel also demonstrated that there is no way for such a mechanical device to prove the reliability of the principles on which it is based. But we will ignore both these objections, and allow the automatic theorem prover to go about its business untroubled by larger questions of ultimate significance.

The art of automated theorem provers consists in developing guided search strategies that are neither too random nor too rigidly programmed, the former being as hopeless[10] as the infinite monkey scenario, the latter not being automatic (or autonomous) enough to satisfy the requirements of the field, not to mention the futurists. Search routines as well as syntax can be built into our fictional typewriters. In fact, the monkeys — the "meat," in the language of William Gibson's *Neuromancer* — are perfectly superfluous in any version of the above picture.[11] Their presence provides the possibility for action in time — for a plot — and indeed monkeys are rich with literary associations. Even so, any mischief a monkey may devise is likely to be irrelevant to the business at hand, which is theorem proving. And this points to a hidden assumption of the genre — that nothing really happens when a theorem is proved. Androids typing at digital typewriters, the only kind we really need, communicate in strings of 0's and 1's — or dots and dashes, if you prefer 19th century symbolism. Time and space are irrelevant to the resulting digital text. Nor do we ask what two file servers think of the information they exchange. From the point of view of the android, any valid inference ends with a theorem, and it is only the programmer's invisible hand that chooses where to switch off the machines and affixes a Q.E.D. to the end of the last completed line.

So I will reluctantly introduce an extraneous character into the fiction, a human mathematician. Without a human character to read the output, the plot reduces to a sequence of logical steps, an endless series of propositions obtained by transformation of an initial tautologically true assertion whose origins need not concern us. With a human on hand, the *mythos*, such as it is, will consist of the attempts of the two characters to communicate regarding the proof. Our android functions allegorically as the

---

[10] Apostolos Doxiadis has drawn my attention to the thesis of Larry Stockmeyer, well known among computer scientists, which quantifies just how hopeless this is.

[11] Compare P. van Emde Boas' *Mechanized Art and the Mad Mathematician*, written to accompany an exhibit of computer art. Rejecting random search in art as well as mathematics, he writes "Intelligence amounts to producing interesting results: inspiring poems, meaningful theorems and beautiful pictures. Our intuition tells us where to look and what direction not to proceed in." But there are neither characters not plot in his scenario either. In: Hein Eberson (ed.): Artificial. TrademarkTM, Amsterdam, 1993.

personification of proving, checking, or reading a proof according to the canons of logical analysis, as opposed to the topographical reading implicit in the use of the word "key." The question here is not to argue that one style of reading is more legitimate or authentic or productive of truth than the other, but rather to imagine whether we can communicate with an android as easily as Thomason communed with Trobaugh's ghost. Whether, in short, the android can make the career-changing move from allegory to genuine drama.

David Corfield has formulated the question more philosophically.

> *…work has only just begun to find languages capable of representing mathematics to both man and machine. In his Image and Logic … Peter Galison talks of the creating of Pidgins to facilitate communication between different communities of researchers. Just as trading partners, each with their own interests, were induced to manufacture common languages adequate for exchange, so, Galison claims, experimenters, instrument designers and theorists have found ways to communicate without the need fully to understand each other's ways. The beginnings of something similar appears to be occurring here. [Louis] Kauffman is encouraging us to encode our concepts in a form acceptable to computers, and then to learn to translate from their languages to ones accessible to us. Although Galison appears to include computers within the scope of trading partners with his talk of 'Fortran Creoles', the objection may be raised that inanimate machines play no active part in language formation. Perhaps… we would do better to view computer scientists (and logicians) as the mathematicians' prospective trading partners. (D. Corfield, Towards a Philosophy of Real Mathematics, p. 56)*

This roughly parallels the first part of Richard Powers' novel *Galatea 2.2*, in which the human author Rick, in partnership with a computer scientist, teaches the "distributed intelligence" Helen, a massively parallel neural network, to read and understand narrative. One might compare the unaccompanied android to a distracted and indifferent maze-builder. The resulting structure is mathematical insofar as it doesn't resemble a maze, as judged by the human mathematician. But Helen is outnumbered by the humans; I want to give the android even odds.

Brian Rotman has described[12] what he calls ghosts —mathematical agents — that are very similar to my androids, though he is concerned with the formal structure of proofs rather than automatic theorem proving as such. Rotman's is a semiotic ghost, whereas I have no trouble imagining a mathematical ghost with whom one can communicate effortlessly: a narrative ghost, which is the same as a spiritual ghost, like Trobaugh. Very roughly:

Androids: ghosts :: logical empiricists : continental philosophers

Unlike the characters in Philip K. Dick's novel *Do Androids Dream of Electric Sheep?*, all deeply concerned about their personal survival, the android conceived as

[12] In "Ghost Effects," a lecture at the Stanford Humanities Institute found on the internet. In *Ad Infinitum*, Rotman's extended study of the semiotics of counting and, he claims, of mathematics in general, the ghost is called the Agent, one of a trinity of "agencies… whose joint action constitute the armature of any mathematical thought experiment," (p. 76), together with the Subject and the Person. There are parallels between Rotman's *dramatis personae* and mine of ghost (e.g. Trobaugh), ghostwriter (e.g. Thomason), and android, but our purposes are very different.

above cannot be easily incorporated into a narrative. The steps do have a sequence and so a directionality in time but this is not the same temporality one usually associates with narrative. How do they differ? And is the writing or presentation of human mathematics more consistent with what we understand as narrative?

Timothy Gowers' (fictional!) dialogue between a mathematician and a computer[13] is a rare attempt to envision realistic communication between a human mathematician and an android named C., collaborating in an effort to solve a concrete problem. Unlike Kurzweil, Gowers does not see human-android cyborg fusion on the horizon, so the question remains: how is this dialogue possible? Each of C.'s suggestions does include a narrative, most more elaborate than Trobaugh's oracle. C. proceeds by combining an extensive database of results and proofs with built-in heuristics, in the tradition of Polya. In spite of its name, C. looks more to me like a ghost than an android. It is an expert reader in the sense to be discussed below, and is likely to pass the mathematician's version of the Voigt-Kampff empathy test employed by Rick Deckard, the bounty-hunting protagonist of Philip K. Dick's *Do Androids…* as well as the "blade runner" in the film of that name.

The word "narrative" lends itself to two misunderstandings. There is what for want of a better term I might call the "postmodern" misinterpretation associated with the principle that "everything is narrative,"[14] so that mathematics as well would be "only" a collection of stories (and that therefore more or less any stories would do). The symmetric misunderstanding might be called "platonist" and assumes a narrative has to be *about* something, and that this "real" something is what should really focus our attention. The two misunderstandings join in an unhappy antinomy, along the lines that, yes there is something, but we can only understand it by telling stories about it. The alternative I am exploring is that the mathematics *is* the narrative, that a logical argument of the sort an android can put together only deserves to be called mathematics when it can be inserted in a narrative. But this is just the point I suspect is impossible to get across to androids.[15]

---

[13] Staged in his millenium article *Rough Structure and Classification GAFA* 2000 (Tel Aviv, 1999), Special Volume, Part I, 79-117. Gowers does not claim to be a specialist in automatic theorem proving but he argues, quite convincingly in my opinion, that there is little comfort to be found in the arguments often advanced against the possibility of automating many of the heuristic strategies of human theorem provers, and not only the routine ones.

[14] Cf. the discussions of "narrative imperialism" and "narrative inflation" in the first chapters of *A Cambridge Companion to Narrative*, ed. D. Herman (2007). See also C. Salmon, *Une machine à fabriquer des histoires*, *Le Monde Diplomatique*, November 2006, 18-19, where it's explained how political and management consultants use "narrative" as a synonym of "spin."

[15] Here I insert a warning: it's safest to consider all references in this text to philosophical, historical, or sociological literature purely fictional. Though I refer to mathematicians of different periods I do not pretend to prove that their practices can be meaningfully identified. And when I claim that it is natural for mathematicians to do or think such and such, this is just a shorthand for my own experience, limited synchronically as well as diachronically. It was chastening to read Alexander Borovik's *Mathematics Under the Microscope* and to realize that much of what he takes for granted as typical of mathematical experience is quite unfamiliar to me, although there is of course considerable overlap as well. … With all due respect for the role of history, how else are historical narratives to be structured?

3. Obstruction

> *But the greatest thing by far is to be a master of metaphor. It is the one thing that cannot be learned from others; and it is also a sign of genius, since a good metaphor implies an intuitive perception of the similarity in dissimilars.* Aristotle, *Poetics*, 1459a, 5-8.

Why was Trobaugh's claim nonsense? Logical empiricism leaves no room for such a question. The claim is wrong because it's not right, and in particular because there's no way to show it's right. The better question is: how did Thomason recognize it to be wrong? The account makes it clear that he had considered just that claim and convinced himself that it was wrong; even more, he had identified an obstruction. I would like to call Thomason's dream a paradigmatic "Aha! experience." But "Aha!" confers no warrant to believe. What is to guarantee that you will not "Aha!" alone? Trobaugh's insight illustrates this: even a semi-wakeful Thomason was not convinced for a minute.

"Show, don't tell" is Axiom B of good writing. The successful author of fiction "engages" the reader, which means the reader becomes both capable of and responsible for the reality effect, whatever it may be. Is it the same with mathematics? Is it possible to communicate mathematics by "telling" the reader what to think? Or can this only work with androids (the formalists' "intended reader").

Regarding understanding, mathematical terminology, insofar as it approaches the philosophical ideal of transparency, is the first obstacle for the uninitiated reader. The android needs no semantics, by definition. The mathematician understands nothing without semantics. The ghost opens with a proposition about perfect complexes. One challenge in this article is to explain how this fits into the narrative without stopping to say what the terminology means. Responding to this challenge is not only necessary but, I believe, possible.

To detect a narrative structure in a mathematical text, first look at the verbs. Apart from the verbs built in to the formal language ("implies," "contains" in the sense of set-theoretic inclusion, and the like), nothing in a logical formula need be construed as a verb in order to be understood, and one supposes an automatic theorem prover can dispense with verbs entirely. One may therefore find it surprising that verbs and verb-constructions, including transitive verbs of implied action, are pervasive in human mathematics. Trobaugh's ghost's single sentence consists of eighteen words, two of which are transitive verbs ("shows" and "extends"), one an intransitive verb ("extends" again); there is also a noun built on a transitive verb with pronounced literary associations ("characterization").

Four out of eighteen is quite a high proportion. One can formally freeze the action by translating the sentence into pseudo-android. The protagonist of the sentence is the "perfect complex" considered collectively, which we abbreviate PC. The oracular pronouncement breaks down logically into four parts:

(1) The direct limit characterization of PC

(2) implies
(3) that PC extend
(4) just as one extends a coherent sheaf

We know no more about coherent sheaves than about perfect complexes, except that we might suspect, correctly, that the expressions are what linguists, not least those concerned with automatic language recognition, call *non-compositional*; that is, a perfect complex is not a complex that is perfect, any more than an Oedipus complex is a complex that is Oedipus (though in each situation there is something called "complex," which are naturally quite different in the two cases). So we write CS as an abbreviation for "coherent sheaf." Then "extends" is a predicate which we denote E. "Just as" in (4) denotes an analogy between a known argument concluding with E(CS) and a potential argument leading to E(PC); one likes to think this analogy is not merely in the eye of the beholder, though where else it might be is a question I can't hope to address. Ignoring for the moment the significant ambiguity introduced by the representation of this predicate by a verb which can be transitive as well as intransitive, we can then reformulate and compress the above analysis:

(1) something already known about PC
(2) series of deduction steps analogous to already known deduction of E(CS)
(3) deduction of E(PC)

Step (1) is background, though it is important that only someone who, like Thomason (and Trobaugh) already knew this "direct limit characterization" would be able to carry out the deduction in (3), or for that matter to understand the deduction of E(CS) mentioned in (2). In automated theorem proving it is permitted to cheat and give the monkeys (1) as the first line of text, after which it is up to them to come up with the steps comprising (2). I don't rule out that an imaginative programmer can find a way — a *search algorithm*, for example — to sensitize the monkey/typewriter combo to analogies like the one invoked in (2), but that word "like" is fraught with peril for the android's monadic self-sufficiency. Once the hazards of (2) have been successfully negotiated, (3) poses no additional difficulty.[16]

I want to return to the ambiguous verb "extend." Trobaugh's ghost claims, falsely, that PC extend; the metaquest being then the possibly unnecessary search for an extension the ghost believes to exist. In the theorem (5.5.4 in T&T) corresponding to the corrected version of this claim the intransitive verb is replaced by a proposition asserting the existence, under conditions connected with the "obstruction" of which we will speak below, of a new PC that one recognizes as having "extended" the PC that was the subject of the intransitive verb, and the only verb remaining is "is" ("exists", if you want to be fussy.[17]). If one believes with the androids that mathematics is logic then one may want

[16] Assuming each of the indicated steps can be translated into a valid logical formula — a big assumption, as we'll see in § 5.

[17] Mathematics collides with philosophy on the subject of the ontological status of existence proofs. As noted below, Trobaugh's ghost switches freely between the intransitive and transitive uses of the verb "extends," which in philosophy of mathematics represents an oscillation between platonism and

to read Trobaugh's verb as sloppy shorthand for this existence proposition, which in turn is an adequate but at the same time fundamentally flawed approximation to the ideal statement in a completely formalized language. I'm all for freedom of thought, but the fact is that Trobaugh's ghost speaks like a real mathematician, and the version with the intransitive verb works. I wonder how, and I am convinced that this sort of wondering can be fruitful for philosophy. More to the point, if the conversation between android and mathematician is ever to get off the ground, one of the two has to learn the other's language. The assumption that it is up to the humans to learn to speak like androids is implicit in the Flyspeck project, to be discussed in § 5. For reasons that should become clear, I favor the alternative.

Teaching the android to translate the intransitive verb "extends" used by a human mathematician into an existence statement doesn't seem unreasonably difficult, but it may be a step toward development of shared intuition. It doesn't seem to matter to Trobaugh's ghost that the same verb "extends" can be used transitively, where it is now "one," the anonymous subject of mathematics, who "extends a coherent sheaf." Understanding this use of the word in all its complexity would be more of a challenge for an android who has not had many opportunities to employ transitive verbs.

The Thomason-Trobaugh article is a contribution to the branch of mathematics known as K-theory, more specifically algebraic K-theory. The name used to designate this branch of mathematics has two parts, each of which poses its own problems. The insider sees mathematics as a congeries of semi-autonomous subjects called "theories," like number theory, set theory, potential theory. I don't know when the word was first used to delineate a branch of mathematics — no later than 179*, when Legendre wrote a book with "théorie des nombres" in the title, and one should note that the plural in the use of the term in French (or German: *Zahlentheorie*) is concealed in an implicit feature of English syntax. Nor do I know whether or not mathematicians borrowed the construction from other sciences. What I do know is that in the examples given above the construction points to a discipline concerned with numbers, sets, and potentials, respectively, and that the word "theory" functions as a suffix, like "-ology." But then what on earth could K-theory be about? Analyzing how the term is used, I am led to the tentative conclusion that it refers to the branch of mathematics concerned with objects that can be legitimately, or systematically, designated by the letter K.

---

constructivism. The axiom of choice, nearly indispensable in mathematics as written, seems to an outsider to commit its user to a platonist outlook, but most mathematicians I know, when confronted with the question, admit to a distaste for existence statements based on the axiom of choice, which is no more than a bridge over an irreducible abyss of ignorance.

In the case at hand, the corrected version of Trobaugh's insight in the proof of Lemma 5.5.1 takes the form of an existence statement (step 3 of the narration of the proof in § 7). This is based on the existence statement in Corollary 2.3.3 of T&T, which in turn uses standard constructions of resolutions in sheaf theory and homological algebra. Since everything in the theory is ultimately based on polynomials in finitely many variables, I would bet that a very patient reader, or even a very skillful android, can make step 3 of Lemma 5.5.1 completely constructive. But I have no idea whether or not this is the case for the topological steps in T&T.

To forestall misunderstandings, I should explain that the letter K has many uses in mathematics, as in chemistry or physics[18] or novels by Kafka, but its systematic use was initiated in 1957 by Grothendieck who, according to my colleague Max Karoubi,[19] was studying a new kind of classes (of something…) and chose to denote them the first letter of the German word Klassen, Grothendieck being a German Jew who managed to avoid the camps and was established as a mathematician in France. This sheds some light on the choice of letter but not on why a whole branch of mathematics came to be named after its leading notation, in what appears, quite appropriately for the 1950s, to be a victory of structuralist semiotics (though neologisms structured around individual letters are common in physics and there is the precedent of the lambda-calculus in mathematical logic, dating from the 1930s and a plausible name for the android's native language).

The letter K in the systematic sense to which I allude above appeared for the first time in a 1958 article by A. Borel and J.-P. Serre reporting on Grothendieck's work, as pure notation. The authors define a group[20] and then write "This group will be denoted K(X) in what follows." The letter X denotes an algebraic variety, an object in geometry whose study is the main purpose of the article. Apart from **K-theory**, what other nouns can be built out of the root K? There are the **K-groups**, algebraic objects of which K(X) is the first exemplar; the definition of K(X) is based on a preliminary sequence of steps at one time called the **K-construction**; following Quillen and Waldhausen, Thomason and Trobaugh derive more general K-groups from geometric constructions for which they appear to be forced to provide compound names — **K-theory space** or **K-theory spectrum** — whereas the logical dependency makes the compound primary.

I do not know how to answer the very interesting question whether the shape of K-theory, now a recognized branch of mathematics with its own journal (*K-theory*, published by Springer-Verlag) and an attractive two-volume *Handbook*, was in some sense determined by its name. What I can say is that, if one grants that there is an idea at the heart of the theory, one plausible narrative would trace this idea back to Euler's formula relating the number of vertices, edges, and faces of a polyhedron, or a configuration of polygons in the plane, the subject of Lakatos' influential *Proofs and Refutations*. And I can point to the institutional recognition of K-theory as a substantial branch of mathematics: at least six Fields medals have been awarded for work directly connected to K-theory and the first two Abel prizes were for work at least tangentially K-theoretic.

****

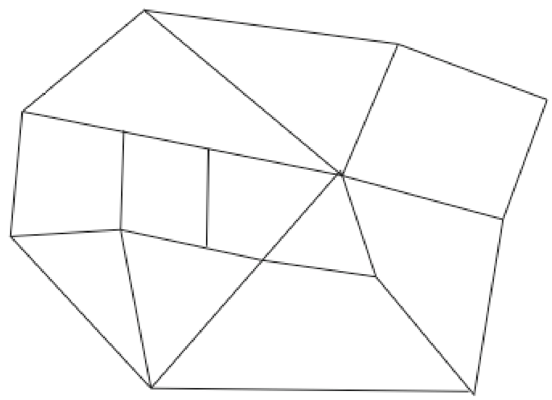
[18] The latter are acknowledged by the Oxford English Dictionary, which has no entry [?] for K-theory.

[19] A curious coincidence: Thomason ran the Paris K-theory seminar with Karoubi, Kahn, and Kassel. See Weibel.

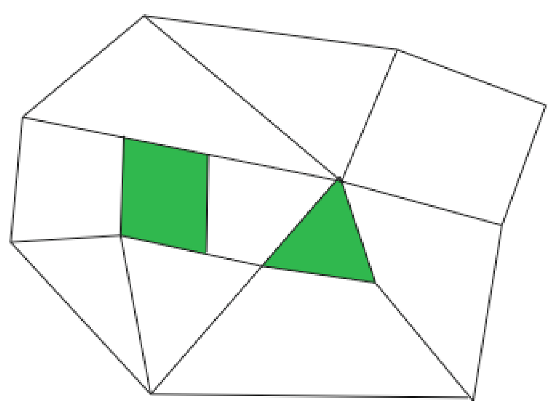
[20] One of many undefined terms in this article. I believe it is possible to read simple sentences in which mathematical terms appear without knowing what they mean: otherwise I would not have attempted to write this article. Concerning this specific undefined term, which occurs only in the present paragraph and in a few other paragraphs later in this section, the reader need only know that the notion is familiar to all mathematicians. Explanation can be more useful than definition, and I will say a few words about the kind of notion a group is.

Wrap a string around a ring and tie it in a knot, then try to pull it tight without letting the string slip off the ring. This is impossible, and is one of the first theorems one learns in topology. It turns out to be difficult to construct a mathematical model of the situation that is both rigorous and recognizably reflects the initial problem, and it is a matter of temperament whether you find it surprising that this is difficult or that it is possible at all.

The obstruction presented in the last paragraph is a symptom of a one-dimensional hole in two-dimensional space (one dimension for the string). There is a generalization of Euler's formula that counts holes in a two-dimensional geometric pattern.



g = 0 (no holes)
F = 11
E = 25
V = 15
V - E + F = 15 - 25 + 11 = 1 = 1 - g



g = 2 (two holes)
F = 9
E = 25
V = 15
V - E + F = 15 - 25 + 9 = - 1 = 1 - g

Much of the discussion in *Proofs and Refutations* was involved in avoiding patterns with such holes, but they can be allowed, and the infallibility of the formula, which you can check for yourself, can again be traced to K-theory, though this is neither historically accurate nor particularly consistent with the most accepted use of the terminology.

One can also imagine a hole in three-dimensional space. You should imagine a stretch of three-dimensional space, for example your living room, and then imagine that right in the middle and about two feet from the floor is a hole, say the size of a pea, where there is no space. This means that this pea-sized spot is off limits to everything. It is not a border; there is simply nothing there, nor can you put anything there. You can take the string from the previous paragraph and try to catch it, and the string will slip off. This means there is no one-dimensional hole. If the cat swallowed it, on the other hand — it might be better to say the cat wrapped itself around the hole — it would be stuck, since no part of the cat could actually enter the hole (because there is nothing there), until it figured out how to unwrap itself.

Mathematics has devised techniques for measuring obstructions of this kind. In the examples given above, the measuring device is called a homology group, and by convention it are denoted by a capital H. This letter serves other purposes in mathematics, but the convention is sufficiently ingrained that a mathematician can read a scrap of text and quickly decide from the context whether the capital H it contains designates a homology group, just as the capital K in a similar context designates a K group. Indeed, the analogy is more than superficial, and it's plausible that the success of K-theory as terminology is in part due to the lexicographical proximity of the two letters as well as the circumstance that the two letters and the notions they designate naturally cohabit in a third and larger branch of mathematics, topology.

Groups were introduced in response to the realization that some of the mechanics of addition can be applied in a variety of mathematical contexts, where numbers themselves are absent. As such, the K-group is an **algebraic** notion, as opposed to the space I have designated X, "space" obviously being a **geometric** notion. The above examples are geometric in the sense that they are allied to our geometric intuition, and I appealed to the geometric intuition I assume the reader shares[21] in order to make the examples more vivid. An obstruction is registered where the smooth structure of abstract space, in which every point is *exactly like* every other point, except as regards location, is interrupted by the presence of a heterogeneous element, like the cat a few paragraphs back in the present narrative. Obstructions can be found in other branches of mathematics. One knows, and one can read in several other contributions to this volume, that the square root of 2 is irrational. Proofs of this theorem a few lines long can be found in many popular texts on mathematics, probably because it is one of the rare non-trivial mathematical facts that can be justified in a few lines. One can also say *that there is an obstruction to 2 having a rational square root*. This seems like a perverse way of putting it, but it is an important insight, presented early in every number theorist's education, that this obstruction is one of a family of obstructions, ubiquitous in number theory, that can be measured by another sort of homology group. There is more to the story. Different branches of mathematics are interconnected; there is a well-understood pathway that relates in a very precise way the obstruction to 2 (or 3, or 5) having a rational square root to the one-dimensional hole in two-dimensional space. This is the very pathway whose generalizations are at the

[21] In other words, I am making the perhaps unwarranted assumption that the reader is **not** an android.

origin of the article identified by Weibel as Thomason's second "major result," which measures certain groups designated by H in terms of groups designated by K, and vice versa, K-theory being viewed in this optic in a very natural sense as the "mother of all obstructions." And it is the existence of such a pathway that makes it possible for me, at least in principle, to read the article in question, though my reading is necessarily informed by my own priorities.

A circle is a good picture of the hole in two-dimensional space. The picture of the square root of two as a square, as in Plato's *Meno*, is not so good in this setting, though it may have been the best picture available to the Greeks. A better picture, once algebra becomes available as a common language, is the equation $x^2 = 2$. Even better is the equation $x^8 = 1$, though to see what this has to do with the square root of 2 requires a bit of calculation.

The word "obstruction" connotes the frustration of an intention. Who, in a narrative about homology, is the bearer of an intention; who is capable of frustration? You, the reader, may have been frustrated that it has taken so long to address this obvious question. I, the author, may reply that the narrative form suffers intrinsic obstructions to addressing several questions simultaneously, and this independently of my limitations as narrator. I submit that you would find simultaneous attention to overlapping narratives problematic even if you were a massively parallel-processing android; a human reader, for reasons a literary critic can best elaborate, might actually find the task easier and might even detect the self-referentiality lurking in this very sentence. Be that as it may, one can structure a mathematical or metamathematical narrative as a sequence of confrontations with various sorts of obstructions, that may or may not be appropriately read as victories over frustration. The narrative of Lakatos' *Proofs and Refutations*, for example, is largely a series of identifications and eliminations of obstructions that are declared irrelevant to the correct formulation of Euler's formula for polygons. The students who carry out Lakatos' dialogue are designated by Greek letters and speak for a variety of known positions in the philosophy of mathematics, but they are much more than mere allegory and the frustration they express when one after another of their attempts to rescue the proof collapses, as well as their satisfaction with the ultimately happy resolution, is dramatically as well as mathematically convincing.

Nevertheless, one mathematician's obstruction is likely to be another mathematician's *pièce de résistance*. Homology, one of the obstructions encountered by Lakatos' model students and the bugbear of earlier narratives, reappears as the central character in much of twentieth-century mathematics. As I explain in detail in section 6, Euler's formula is now understood as the first of a family of formulas in which the previously obstructive homology, domesticated and taught the benefits of cooperation, plays a starring role. This capsule narrative of twentieth century topology can in fact serve as the archetype of Thomason's reworking of the Trobaugh dream.

In Thomason's version of K-theory the burden of obstruction is borne by perfect complexes, and this is how we know Trobaugh's ghost was talking about K-theory though he did not say so explicitly. Even without knowing what perfect complexes are, we can learn by reading T&T for form rather than content that "one" can do things to them, like extend them. "One" can apply transitive verbs to them. They come alive in the theory to which they are invited in a variety of ways which correspond roughly to the different transitive verbs of which they can be objects. Grothendieck's famous six functors, to which we return in a later chapter, are six transitive verbs in this sense.

Or they can do things on their own: extend becomes an intransitive verb.

Either way, as characters, they are no more or less lovable or individuated than androids. They are not all equal — otherwise there would be nothing to say about them — but, in a way that captures something important about the abstraction at work in T&T, if you have seen one perfect complex, you've seen them all.

4. Genres

…both theorem proving and stories are about people in action to achieve a certain task — this is based on the assumption that mathematicians are people. (Apostolos Doxiadis)

Thomason's paper belongs to the genre of the Foundational, which is more than appropriate for Grothendieck. This is not a judgment of merit, though it is also that, a little bit, Still less do I mean that it is Foundational in the sense of Foundations of Mathematics, even though Thomason does pay more than lip service to avoidance of set-theoretic paradoxes. Rather I mean that it is an attempt to provide a common vocabulary and viewpoint for an entire field. This should not be understood as arrogance on Thomason's part, though he does point out the relative advantages of his approach in the very first paragraph of the introduction:

> Indeed most known results in K-theory can be improved by the methods of this paper, by removing now unnecessary… hypotheses. (p. 247)

Thomason takes the steps lightly but briskly. He's a pro. But he is thoughtful ("inordinately generous") enough to indicate every step to the reader. A Foundational text is one that can be read linearly, step by step, unlike a typical research article that is more typically read for meaning, in zigzag style, or radially (cf. § 7). What is a "step?" The author of a Foundational text takes it for granted that even the most inexperienced human reader can cover more ground in a typical single "step" than a universal Turing android can in a dozen deductions in the formal language of Foundations of Mathematics.

"Foundations" is an example of a metaphor for normative mathematical practice that somehow stuck. "The earth shall rise on new foundations" was a popular refrain in Russell's day. Even most modern architecture requires foundations, if I'm not mistaken. But other metaphors may be more apt, or at least more timely. Logical proof can be seen as analogous to the immune system, for example. Then Fregean logical hygiene can be interpreted as an auto-immune disease which, to be fair, mathematicians have little trouble keeping in check. This has no bearing on the no less metaphorical use of "Foundational" to describe an article such as T&T.

For the nonspecialist reader, a mathematical text is addressed not merely to proving a collection of theorems but to solving problems, the clear delineation of which is one of the author's tasks. For a foundational article like T&T, the problems may be of two orders: enlarging or altering perspective, and then exploring what can be done in the new framework, which may include solving old problems that had previously resisted solution (note the personification). The nonexpert reader is likely to read such an article not linearly but radially: beginning by grasping the problems to which the author's attention is directed (note the geometric and hand/eye metaphors), then gradually identifying the turning points in the author's solution of these problems.

It goes without saying that an expert reader is already aware not only of the basic problems in the field but is also familiar with past and present unsuccessful or partially successful approaches to solving these problems, and will read a text in a quite different way. For such a reader, the foundational material will be largely familiar, in a more or less different form; this is part of what it means to be an expert.

Most mathematical literature is naturally not Foundational. A typical research article is organized around one or more new results; if more than one, unity is provided by the application of a common method or by their derivation from a single main new idea, generally surrounded by technical innovations introduced as tools but capable of attaining star status in later installments. Before publication in a journal the article is usually examined by the editorial board, then sent to a referee for a careful reading (rarely as careful as one would like), then returned to the editorial board for a final decision. This is a highly idealized description of a complex sociological reality thoroughly dependent on a variety of institutions of more or less recent standing; what matters here is whether this suffices to determine research articles as a genre. I think not: there are articles that introduce new structures, articles that carry out intermediate calculations, articles that establish relations between different structures, articles that solve longstanding problems. Each of these descriptions defines a corresponding genre, and the list is far from exhaustive.

At the heart of T&T is a normal research article. It would have been possible for Thomason to publish the foundational material separately, leaving the genuinely novel material for a shorter article, but the Grothendieck Festschrift was a natural occasion for him to rethink the foundations of his subject, or if you like to redefine its "whatness" (cf. § 6) in terms of derived categories and perfect complexes. The virtual research article that haunts T&T is a rapid succession of key points, joined by the shortest possible paths. Franco Moretti's analysis of the bourgeois novel of the nineteenth century hinges on the distinction between *turning points*, few in number, and *filler*: "[n]arration… of the everyday… without long-term consequences 'for the development of the story'…"[22] One is strongly tempted to compare Moretti's turning points (he enumerates three such in

---

[22] *Serious Century*, in *The Novel, Volume I*, Princeton University Press (2006) 364-400. The quotation is from p. 368, the inner quotation from Barthes.

Pride and Prejudice: meeting, proposal, acceptance) to the key points of a research article, the filler (which account for 110 narrative episodes in *Pride and Prejudice*) being analogous to the routine material, already known in principle to experts, that makes up the bulk of even the most briskly-paced research article and the entirety of a purely Foundational text. And one might imagine an evolution in the long-term to a mutually satisfying division of labor, in which the uncomplaining android manages the routine filler while the human (or spiritual) mathematician retains the romantic role as intuitor of turning points.

It was in the nineteenth century, of course, that the convention of the research article achieved roughly its modern form. The analogy can only be structural. The key/turning points serve to organize the reading of the narrative in each case, but the functions are quite different. Without the routine verifications there can be no legitimate "story" in a mathematical article, whereas Moretti's filler is the expression of "capitalist rationalization" (p. 392): "they offer *the kind of narrative pleasure compatible with the new regularity of bourgeois life*" (p. 381, emphasis in the original). Or perhaps the functions are not so different after all? While the "pleasure" in a mathematical text is almost exclusively concentrated in the key points, a fact a good expositor knows how to exploit, the "regularity" of routine verification is indispensable for the reader's satisfaction that the article does indeed show what it has claimed. That it belongs to the genre of legitimate research article.

Research and Foundational articles do not exhaust mathematical writing. There are textbooks at various levels of generality, including encyclopedic textbooks such as Bourbaki's Eléments des Mathématiques as well as more specialized texts actually used for learning mathematics. There are "survey" and "expository" articles that often dispense with proofs altogether and present the results as sequences of ideas that would pose a special challenge for androids. At another extreme is the genre called Folklore, for example Thomason's Cofinality Theorem (Theorem 1.10.1 of T&T), of which the authors write, "We found this proof in 1985; it has since become folklore." This signifies a proof circulated orally, perhaps with the aid of informal notes, unpublished but with a sufficiently recognized status to be used as a reference.

The rigidity of the vocabulary, the systematic avoidance of any hint of ambiguity[23], represents a major difference between fiction and mathematical prose and is the most obvious reason the latter is so unappealing.[24] There is no place for synonyms in mathematical prose, so when one means K-theory, for example, one has no choice but to write “K-theory.” The same goes for any mathematical term, even a generic term like "obstruction." Mathematical prose is morbidly repetitious.

---

[23] The ambiguity of the verb "extends" discussed above is not an exception, being properly a feature of the metalanguage. Compare Frege, quoted in Herman, p. 22: "[W]e require a system of signs from which all plurisignificance has been banished, and from whose stricter logical form the content [of a given mathematical idea] cannot escape (*entschlüpfen*)."

[24] John Baez' contribution to this volume emphasizes more global factors.

Not for us rose-fingers, the riches of the Homeric language. Mathematical formulae are the children of poverty. Netz, p. 145[25]

Contemporary mathematics does not suffer from the "absence of nuance" Netz discerned in his classical texts, but every nuance is perfectly calibrated. The Thomason-Trobaugh article is situated in the general framework of *Waldhausen categories*, named after a human being (Waldhausen) rather than an idea[26]; but when additional hypotheses are required the authors work with *small saturated Waldhausen categories*, *complicial biWaldhausen categories*, or *complicial biWaldhausen categories closed under the formation of canonical homotopy pushouts and canonical homotopy pullbacks*. The first step identified in Weibel's periodization of the proof of the main theorem is the Cofinality Theorem I just mentioned, which applies to a *Waldhausen category with a cylinder functor satisfying the cylinder axiom*. This compound expression cannot be decomposed without loss of sense, which pretty much rules out making the Cofinality Theorem the subject of a poem, even in free verse, though I have included all this terminology, which I assume the reader will find meaningless, to recreate the atmosphere of some mathematical writing.

Anticipating the confrontation between a ghost and an android that will be staged in the penultimate chapter, this deadening of prose and its fruitful ambiguities in search for maximal precision seems to tilt the argument in the android's favor. But the very fact that one can ask about synonyms points to the centrality of meaning in a sense that seems unthinkable to the android. A successful mathematical lecture, like an expository article, concentrates on meaning ("ideas") and counts on the audience's confidence in the lecturer's ability to connect the concentrations of meaning (turning points) through the application of routine skill (filler). A successful introduction to a paper plays much the same role. This connecting material is provided in the written text, and even if one imagines that fiction can sometimes be decomposed in an analogous way, the rules of art for writing connecting material in fiction and in mathematical prose have next to nothing in common. Individual variations in style play next to no role in the writing of mathematical filler. One can often recognize a mathematician in the purely verbal features of an extract, and one speaks of a typical style in a sense that is different from that of "styles of reasoning" (Crombie). But the presumption is that this refers not to the

---

[25] R. Netz speculates on cognitive reasons for this "one-concept-one-word" principle in classical Greek mathematics and alludes to the "monstrous repetitiousness" of Euclid: pp. 107-8, and *passim*. This is important in his explication of generality in the setting of Greek mathematics: "In the mathematical world there are no shades of meaning. And this, the all-or-nothing nature of mathematical predicates, is what makes generality so obvious." (p. 266) Netz does note the presence of synonyms in the *metalanguage*, however, and this is no less true of contemporary mathematical writing. We have already seen synonyms for "key'; other examples include "simple" (elementary, straightforward), "analogous" (similar) and the notorious "obvious" (clear, evident, immediate).

[26] Not, it should be unnecessary to add, because these categories bear a superficial resemblance to Waldhausen, but because the T&T definition of Waldhausen categories summarizes some of their properties to which Waldhausen drew attention in an influential article. Whether there is something fundamentally Waldhausenish about such categories — whether in a non-trivial sense, only Waldhausen or someone very much like him could have made these observations, just as Leibnizian or Hegelian philosophy tells us something about the personalities of these two men — is a question with important implications for the future of android mathematics.

filling in of routine details but to the succession of ideas. In Aristotle's terms, is the meaning then to be understood as *mythos*, *ethos*, or *dianoia*? I cannot settle this question but I am convinced that what one calls the style of a mathematician is a narrative style.

5. Automated theorem provers

> *In the past, a partial and inadequate view of human purpose has been relatively innocuous only because it has been accompanied by technical limitations that made it difficult for us to perform operations involving a careful evaluation of human purpose.*
> Norbert Wiener, *God and Golem, Inc.*

Android society has its own textual analysts. They are called **automated proof checkers** and are in principle indifferent to questions of style. Their assignment is to read a proof and check that each line is valid and the passage from one line to the next is compatible with the rules.

Even though android society largely exists in the imagination of computer scientists, cognitive scientists, and futurologists of various stripes, one understands that the automated proof checkers are imagined as the assembly-line workers in comparison with the **automated theorem provers,** creators of new proofs whose coming will herald the twilight of the profession of human mathematician. My infinite monkey scheme had automated proof checkers tucked away unseen in the hardware (typewriters). On philosophical grounds this might be understandable, since the notion of a system of rules presupposes that one can check whether a rule has already been followed, whereas the problem of determining what will happen when a machine is programmed to follow a given set of rules has been known to be undecidable since Turing. A loose analogy might be the comparison between finding a trail across the mountain range separating two villages, a project whose success is not guaranteed and whose failure may have catastrophic consequences, vs. following the trail once it has been found, the "verification" of its correctness consisting in where you find yourself at the end.

Automated proof checkers deserve more respect than I am letting on. Concerned that acceptance of his computer-assisted proof of the Kepler Conjecture by the *Annals of Mathematics* has led to a change in the journal's policy, because the referees felt they were unable to certify fully the correctness of the computer code, Thomas Hales has launched the Flyspeck Project whose goal is to produce a fully formal version of his proof of Kepler's celebrated conjecture on the densest packing of spheres in three dimensions. "Formal proof," for Hales, "more fully preserves the integrity of mathematics" than the traditional refereeing process, faced with the unprecedented challenge of certifying that the computing used in proving theorems is as reliable as a logician's android.

Hales' fact sheet for the Flyspeck Project instructs us to understand a formal proof in the sense of the **QED Manifesto**.[27] Though this is not why Hales chose the name, formalization of the proof of the Kepler Conjecture is indeed a mere flyspeck on the ultimate goal of the Manifesto's (semi)-anonymous authors: "to build a computer system that effectively represents all important mathematical knowledge and techniques." This is definitely android territory: "The QED system will conform to the highest standards of mathematical rigor, including the use of strict formality in the internal representation of knowledge and the use of mechanical methods to check proofs of the correctness of all entries in the system."

> *[P]erhaps the foremost motivation for the QED project is cultural. Mathematics is arguably the foremost creation of the human mind.... one of the most basic things that unites all people, and helps illuminate some of the most fundamental truths of nature, even of being itself. In the last one hundred years, many traditional cultural values of our civilization have taken a severe beating, and the advance of science has received no small blame... The QED system will provide a beautiful and compelling monument to the fundamental reality of truth. It will thus provide some antidote to the degenerative effects of cultural relativism and nihilism.* (QED Manifesto, 1994).

Hardly anyone writes manifestoes like that anymore! Much of the Manifesto's language would bring a grin to the grimmest Terminator ("an industrial designer will be able to take parts of the QED system and use them to build reliable formal mathematical models of not only a new industrial system but even the interaction of that system with a formalization of the external world"). Allusions to "the foremost creation of the human mind", however, would leave my android character cold.

In the future we will ask ourselves whether or not to trust the answer only the android can provide. Back in the early 21st century, Hales is understandably troubled by our groping towards formulation of this question and by what he seems to see as a blemish on the acceptance of his computer-assisted proof by the extremely prestigious *Annals.* Yet he implicitly seems willing to live with the inevitable flaws of human creation. The QED Manifesto lapses into uncharacteristic wistfulness on this point ("The standard of success or failure of the QED project will not be whether it helps us to reach the kingdom of perfection, an unobtainable goal, but whether it permits us to construct proofs substantially more accurately than we can with current hand methods."[28])

A QED-strength proof-checker may well succeed in removing the asterisk that Hales feels disfigures acceptance of his proof. Hales estimates that it may take "as many as 20-work years" to reach that point. But only a synoptic proof — perspicuous, in the sense of Wittgenstein[29] — can remove a second asterisk, signifying that the proof has not been meaningfully understood by that "human mind" to which QED displays such anachronistic concern. Perhaps we should

[27] http://www.rbjones.com/rbjpub/logic/qedres00.htm

[28] Though the Manifesto's authors reportedly abandoned the project by 1996, for reasons that deserve to be explored, it still serves as a reference, notably for Hales and his colleagues.

[29] *Übersichtlich*, also translated "surveyable." *Remarks on the Foundations of Mathematics*, II, 1.ff.

instead ask the android whether it suffers metaphysical vertigo staring into the abyss of the impossibility of foundations. Does it panic at the menace of infinite regress?

My account of the Flyspeck Project has been misleading. Hales' proof exists and in principle is only in need of checking. But, not being a formal proof, it is as incomprehensible to the androids dreamt of by QED as "42" is to the characters in *A Hitchhiker's Guide to the Galaxy* to whom it is proposed as the answer to everything. Automatic proof-checking on the scale of the Kepler conjecture, it seems, is subordinated to automatic theorem-proving. The Flyspeck Project leaves no ambiguity on this point, stipulating that "[a]ll the formal proofs will be made by computer" and moreover "programmed in the Objective CAML programming language," the "steps of the proof" being "generated by computer programs" using prescribed software packages. The "design of the proof" will nevertheless be based on the work ("the 1998 (traditional-style) proof") of human mathematicians Hales and Ferguson.

Whatever sort of thing a "design of a proof" is, it must be awfully subtle. I can only understand the word "design" as a metaphoric way of referring to the "key points" of the proof and their interrelations. It is hard to imagine a vaguer way of speaking, but in the practice of mathematics, teaching as well as research, one makes constant use of such metaphors. The precise choice of words is a matter of taste. Alternative metaphors I have seen recently, as a member of a committee organizing an international conference devoted to the presentation of a particularly striking, ingenious, and complex proof, include "architecture," "overall scheme of things," "outline," "ideal proof," and of course variants based on the word "structure." "Narrative structure" is the metaphor I am exploring in this essay. The circumstances of the Thomason-Trobaugh collaboration, and indeed of Thomason's account of this collaboration, convince me of its pertinence to human mathematics. Whether or not it is inevitably indispensable to human-android interaction, as it appears to be in Hales' situation, is a question the QED Manifesto might have profitably addressed. It is probably no accident, though, that Hales chose the spatial metaphor "design." "Program design" and "software architecture" are more than mere metaphors in computer science, and it's plausible that existing humans and existing androids have already developed a stock of shared intuitions based on flow charts and the like.

Visual and narrative metaphors for the "proof within the proof" are not mutually exclusive. I can't decide whether it is more natural to attribute an intrinsic narrative structure to a flow chart, or whether on the contrary a narrative structure can best be represented by a flow chart (compare the diagrams in David Herman's presentation). Explaining a proof on paper, or on a blackboard, may involve drawing something like a flow chart whose nodes are the *key steps*. Explaining the same proof without the help of visual aids — to a blind mathematician, for example, or over the telephone — might take the form of a narrative linking the same key steps. Here two mathematicians envisage extracting key steps from a computer-generated proof:

> *Basically what was missing was any distinction between important and unimportant steps in the proofs. It is certainly also possible that important steps get hidden inside tactic scripts. However, it seems that the most common situation is that important steps correspond in some way to* **tactics which look a bit out of the ordinary,** *and which would stand out under a rapid examination of the tactic script.* (Marco Maggesi and Carlos Simpson, emphasis added.)

Beeson, a participant at the 1994 QED Workshop whose survey *The Mechanization of Mathematics* was one of my main sources for the topic of this section, uses the word "key" like a mathematician. He explains the "key idea" of the Turing machine, the "key step" in automating verification of trigonometric identities, and "the key to automating proofs of combinatorial identities" in the work of Petkovsek, Wilf, and Zeilberger. But though the automated proofs he describes in his chapter consist of nothing if not of "steps," none of these is identified as "important" or "key." At most there are "candidates for 'lemma' status: short formulas that are used several times."

As depicted by Beeson, the typical strategy for automated theorem proving is a sophisticated version of the infinite monkey scenario, with more or less intelligent guidance provided by the programmers but minus the monkeys. You begin with a collection of axioms defining the theory and add the *negation* of the theorem you want to prove. The program then applies logically valid transformations, possibly according to a pre-defined search strategy, until it arrives at a contradiction. Since the initial axiom repertoire was presumably consistent, you are entitled to conclude that the negation of the theorem is necessarily false, hence the theorem has been proved.

An early example of this strategy is the Knuth-Bendix algorithm, used by "most modern theorem-provers" (Beeson, p. 34 ff). As Beeson describes it, the Knuth-Bendix algorithm takes a collection of equations which define a mathematical theory and by repeated application of a specified subset of these equations, called *rewrite rules*, transforms the original collection into a new and simpler set of equations that suffice to define the same theory as the original set. This strategy only works when the algorithm terminates, which is not guaranteed, but it has succeeded in proving a number of simple but interesting theorems, including some for which no earlier proof was known.

My working hypothesis is that communication with a mathematical android must in an essential way be the communication of narrative structure, organized around a series of key points, each hinging on the transformation (an attenuated form of *peripeteia*, as used in Aristotle's Poetics, cf. §7) of some "whatness" (cf. § 6) which is in turn based on shared primitive intuitions—shared between the human and the android, that is. Reading over Beeson's article, the most plausible candidates for the primitive intuitions of today's androids are the principles behind the search strategies such as the Knuth-Bendix algorithm. Another strategy discussed by Beeson is *quantifier elimination*. This is an effective mathematical intuition used extensively by specialists in a variety of fields, and it adapts well to mechanization. Do these principles overlap with any of our spatiotemporal intuitions? Do they have analogues in our narratives?

What might be called *recursive simplification* includes both of the strategies mentioned above. It also underlies the principle of robot vacuum cleaner function, the task being completed recursively with the result guaranteed probabilistically. As far as I can tell, there is no key idea in either case. Trobaugh's intuition, by contrast, is nothing but a key idea. But I do not know how to characterize Trobaugh's intuition intrinsically, to show how it differs from the principles underlying the search strategies mentioned above.

"*One can view computer algebra and computerized decision procedures, such as quantifier elimination or Wilf and Zeilberger's decision procedure for combinatorial sums, as ways of embedding mathematical knowledge in computer programs"* (Beeson, p. 46). Or one can view them as elements of the android's repertoire of primitive intuitions. Unlike our intuitions of time, space, and motion, the android comes into the world with a sense of recursion that to a human interlocutor looks like a compulsion to replay the same steps endlessly. What would Freud have made of this repetition compulsion? Can a primal scene be attributed to a theorem-proving android?
[30]

Compare Beeson: "One aspect of mathematics that has not been adequately mechanized at the present time is *definitions*" (p. 20) to David Gelernter: "no thinking computer is possible until we can build a computer that hallucinates," referring specifically to the hallucinations that take place in dreams.[31] Gelernter, writing for the general public, is exploring the obstacles to mechanizing creativity, whereas Beeson, writing for specialists, seems to be concerned with mechanizing concept formation.

> Dealing with the challenges of second-order variables (without quantification), definitions, calculations, incorporating natural numbers, sequences, and induction, should keep researchers busy for at least a generation. At that point computers should have more or les the capabilities of an entering Ph. D. student in mathematics. Now, in 2003, they are at approximately freshman level. I do not mean that this progress is inevitable—it will require resources and effort that may not be forthcoming. But it is *possible*. Beeson, p. 23 (emphasis in the original).

In Powers' *Galatea 2.2* a "turning point" in Helen's education, and presumably in the novel (in which there are far more than three), is reached when "she" asks the question "What is singing?" Rick's first answer, that it is a bird, does

---

[30] Maybe induction is a primitive intuition. G. Lakoff and R. Nuñez seem to think so: see their book *Where Mathematics Comes From*.

[31] Beeson: p. 20. Gelernter's remarks, originally published in ??? are taken from Project Syndicate 2002.

not satisfy his android pupil; only after several more failed attempts does Rick realize that Helen has nothing other than the word to associate with "singing" and that the only correct answer is an *ostensive* definition*,* which Rick provides by singing a song. Here Powers is concerned with teaching Helen **meaning** (*dianoia*). In other scenes Helen is led to ask questions about **character** (*ethos*), notably about herself. But I did not find any sequence in which Rick dealt with his android's problems with **plot** (*mythos*); on the contrary, as in the singing episode, Rick presumes that the action consisting of a bird's singing is not problematic for Helen's distributed intelligence. Are we to conclude that, for this novelist at least, communication with androids must start with "whatness"? Or is this feature specific to neural networks?

Would an android appreciate Borges' "Circular Ruins" or would it, on the contrary, suffer vertigo, as I already suggested above? How about a story constructed through permutation, as in Oulipo? Would the android see Gertrude Stein as a prototypical narrative? But if this is narrative, it enters by the back door, so to speak, because we are programmed to expect text to conform to a narrative pattern.

My personal inclination is to understand logic and formalization as a metaphor for mathematics. Not only can this be very enlightening as a narrative about mathematics, it can even be incorporated into mathematics itself.[32] This is not the case for other metaphors, though K-theory as developed by Quillen, Waldhausen, and Thomason, among others, can be seen as a vast incorporation of a certain metaphor of mathematics, involving diagrams, into the body of mathematics. But don't mistake the metaphor for the material. The material in mathematics can only be the mathematics as actually practiced. Only when this has been established can one begin to argue about the respective roles of synchronic and diachronic models of mathematics, whether history or anthropology provides a better guide to mathematical practice, which in turn determines how we define the role of androids.

6. K-ness

> *Don't think about it, just do it. Don't pause and be philosophical, because from a philosophical standpoint it's dreary.* Rachael Rosen, in *Do Androids Dream of Electric Sheep?*

There is a sense in which the number of holes in a surface, by which I mean just a configuration of polygons in the plane, like those illustrated in § 3, tells you all you need to know about that surface, provided it is all in one piece ("connected"; if there are

[32] In his recent (joint) work, Hales has demonstrated its relevance to my own field, which was also the field in which Hales first made his reputation: R. Cluckers, T. Hales, F. Loeser, Transfer principle for the fundamental lemma, to appear in L. Clozel, M. Harris, J.-P. Labesse, eds., *Stabilisation de la formule des traces, variétés de Shimura, et applications arithmétiques*, Book I.

separate pieces, you have to count the number of holes in each piece). What is remarkable about Euler's formula — add the number of vertices and faces and subtract the number of edges and call the resulting number V+E-F the **Euler characteristic** — is that it gives you a complete and infallible way of counting the number of holes.

You have to imagine a hole as a Gestalt or a synthetic unity that not every observer may recognize immediately, especially if the observer is an android. Drawing the surface, on the other hand, just means drawing the vertices and faces and edges, and surely the dullest android can keep track of that. Since the Euler characteristic of a (connected) surface contains all you need to know about the surface, in the sense we have yet to define, we thus have a purely mechanical way of specifying the "whatness" of any plane surface, in this same undefined sense.

The name given to this "whatness" is **topology**. We have tinkered with our definitions and conventions ("monster-barring," in Lakatos' terminology) until we can extract from them a "whatness" that our proposed formula is able to calculate, at which point we give this "whatness" a name and declare victory. I grant that an android is likely to find this sort of victory pointless, and for the reasons discussed by Lakatos (in the person of Alpha[33]) in his *Proofs and Refutations*, the most sustained example of proof narration with which I am familiar — the proof in question being that of Euler's formula, as it happens. But for the moment all we are asking the android to do is to calculate the Euler characteristic. We feel this is much less noble than proving Euler's formula, but this only means the android will have to adopt our values in order to accede to our standards of nobility.[34]

You'll remember that the number of holes in a surface is a measure of the obstruction to pulling knots tight without crossing the border of the surface. One of topology's jobs is to count obstructions, which turns out to be much harder for spaces of higher dimension than it is for surfaces, and indeed this is why Grigori Perelman's recent proof of the Poincaré Conjecture is of such importance. By way of historically motivated analogy, we are entitled to address the following questions to Thomason and Trobaugh:

(a) How can we define "K-ness"?
(b) Can it be calculated?

The Euler characteristic puts together the topology of a surface after cutting it into pieces. The number of holes is what is called a **global invariant**, whereas the vertices and so forth are the building blocks, the number of which are so many **local invariants** since

---

[33] In 4.(b) "I admire your perverted ingenuity in inventing one definition after another as barricades against the falsification of your pet ideas. Why don't you just define a polyhedron as a system of polygons for which the equation V - E + F = 2 holds?" I mention in passing that the "Eulerianness" to which the students in Lakatos' book strive is not strictly analogous to the "whatness" discussed here. As a branch of mathematics, topology studies "whatnesses" of which Eulerianness is just one instance.

[34] "While performing a calculation, one needs to be careful, but one does not need to be a genius, once one has figured out what calculation to make. It is 'merely a calculation.' When finding a proof, one needs insight, experience, intelligence—even genius— to succeed," writes Beeson, who immediately explains "because the search space is too large for a systematic search to succeed." Beeson, p. 20.

each one is localized in a specific place. In a sense, T&T is a step toward doing for "K-ness" what the Euler characteristic does for topology. The K_0 obstruction is part of an object's K-ness; Trobaugh's ghost focused Thomason's attention on the obstacle to cutting a K-theoretic object into pieces without loss[35].

Here Euler's formula serves as a primitive intuition, less fundamental indeed than the intuitions of time, space, and motion all humans can be said to share[36], but one mathematicians have incorporated as a common resource. There is nothing in T&T that can be literally cut up in the way one of the drawings above can be (with scissors, if you like), but Grothendieck's vision of geometry maps all sorts of geometric problems into this primitive intuition. A good topologist, like Thomason, has access to more intricate primitive intuitions, but an outsider like this author can always fall back on the simplified model.

Topology is a name that stuck, unlike *analysis situs*, an earlier name for the subject that studies the sort of "whatness" that is its object. These days topology designates a certain kind of intuition, familiar to topologists, that has no other common name. It is accurate to say that the intuition has been developed with the help of the name, in a sense that has nothing to do with etymology. There are metaphors in topology — cutting, pasting, gluing, surgery, for example — that are basically indifferent to the original meaning of *topos* but are perfectly in tune with topological intuition.

Is it the same with K-theory? There are K-theorists, and they have developed a collective intuition, but is it addressed to some underlying K-ness or is K just an initial, as in Kafka, that leaves the reader with the disquieting sense that there is more to the matter at hand than the mind can grasp?

The simple answer would be that K-theory is just about K-groups and related notions. K-theoretic ideas have a habit of slipping across boundaries, however, and so the ideas typical of K-theory can arise in unexpected places. At the risk of irritating professional historians, one can mention such precursors as the Riemann-Roch formula (1864) and the Weyl Character Formula (1926) as well as Euler's formula (1752). The journal *K-Theory* publishes articles in practically every branch of pure mathematics. Without too much distortion I can identify a K-theoretic step in the Taylor-Wiles article that completed

---

35 What is lost upon cutting into pieces is part of the object's K-ness. This escapes circularity, I think, because the object is initially apprehended categorically, which means that it is determined by its web of relations with all other objects in the same category. In T&T, this is the category of schemes, the foundation Grothendieck proposed for algebraic geometry, and what I'm calling its K-ness is one important property of a scheme. The point I'm trying to make is that K-theoretic intuition can be applied in a variety of categorical settings — in operator algebras, for example, or topology — and captures a feature common to reasoning in these different categories.

36 Cf. Stanislas Dehaene: "mathematics is a construction on the basis of raw intuitions or primary cerebral representations that have been engraved in our brains through evolution.", at http://www.edge.org/discourse/dehaene_numbers.html. Of course it can be and has been argued that the modalities of apprehension of time, space, and the like are historical and cultural variables. I have no desire to rehearse these arguments here, except to say that — for obvious reasons — the term "intuition" as used in this essay cannot be reserved exclusively for brains as these are conventionally understood.

Andrew Wiles' proof of Fermat's Last Theorem, though K-theory is nowhere mentioned by name.

What complicates the case of K-theory is that, unlike geometry or arithmetic or dynamics or even the topology that underlies (and, as Lakatos' *Proofs and Refutations* illustrates, threatens to undermine) the apprehension of Euler's formula, K-ness cannot be conflated with a primitive (or *a priori*) intuition anterior to mathematical abstraction (what Plotnitsky calls *phenomenal intuition* in his contribution to this volume). Compare this to *algorithmics*, a branch of theoretical computer science whose name, like K-theory, is also due to an accident of translation.

I do not know how to begin to discuss this with an android. My first serious exposure to philosophy of mathematics — to philosophy of "real mathematics" as Corfield puts it — may have been the following remark

> There are good reasons why the theorems should all be easy and the definitions hard. As the evolution of Stokes' Theorem revealed, a single simple principle can masquerade as several difficult results; the proofs of many theorems involve merely stripping away the disguise. The definitions, on the other hand, serve a twofold purpose: they are rigorous replacements for vague notions, and machinery for elegant proofs. M. Spivak, *Calculus on Manifolds* (Benjamin, 1965, p. ix).

Though Spivak is not a licensed philosopher[37], and though his point of view is not universally shared, it can serve as a starting point for an attempt to come to an understanding with our android colleagues, to allow them to aspire to nobility as we understand it or, alternatively, to shatter our illusions. Not every turning point in a proof is necessarily a definition. Trobaugh's ghost's insight turned on identifying obstructions rather than on providing a new definition. But I would say, at the risk of seeming tautological again, that every "key point" is in some way connected with our *habitus*. I borrow the word from sociology, specifically from Norbert Elias and Pierre Bourdieu, but I could just as well have used "form of life." Either term refers to our social life, but I would prefer to emphasize not the specific social structure in which we find ourselves, which varies constantly from one period to another, viewing *habitus* rather as the possibility of being in any social structure at all.

An android's social life is deficient in all respects. Nevertheless, if we follow Beeson's prescriptions, we may be led to study statistical patterns in theorem-proving androids: how often does a "short formula" have to be used in order to qualify for "lemma status," how often are such proto-lemmas found in close proximity, etc. This would also be automated: *Habitus*-androids? How would their *habitus* differ from ours?

On the account of Elias, one would be forced to conclude they have none:

---

[37] I can imagine a roundtable attempt to explicate his metaphors of "stripping away the disguise" and "machinery," in which mathematicians, philosophers, androids, and perhaps even literary theorists, could all take part.

> Mathematische Begriffe mögen von dem sprechenden Kollektiv loslösbar sein. Dreiecke mögen erklärbar sein ohne Rücksicht auf geschichtliche Situationen. Begriffe wie "Zivilisation" und "Kultur" sind es nicht. (Elias, *Über den Prozeß der Zivilisation*, Suhrkamp Taschenbuch (1997), Vol. 1, p. 94)

Elias is making the claim that a mathematical concept like that of the triangle is not culture-bound. But this must change when a triangle enters a narrative. If we are to communicate with an android about a triangle, or a perfect complex, it must be on the basis of a *habitus* we share. ("I never felt at home here," complained Powers' literary android Helen in her last words, commenting on Caliban's "noises, sounds, and sweet airs" speech.)

7. Archetypes

> *...an ongoing Transgression...the invasion of Time into a timeless world.*
> Pynchon, *Against the Day*.
>
> *Mathematics is unpredictable. That's what makes it exciting. New things happen.*
> William Thurston, May 14, 2007.

K-theory, and more specifically algebraic K-theory, is a stable developed discipline with its foundation myths, as discussed in the previous section; its canonical texts[38], its hierarchy (Thomason was near the top), its habitus, its triumphs and frustrations, and a network of audacious conjectures the mere statement of which presupposes the solution of one Clay problem and the reinterpretation of a second in a vastly more general context. It is close to ideal for my purposes in this essay. Although K-theory is not entirely alien to me, I am far from a specialist, and with some care I should be able to account for my reading of a canonical, foundational text in terms that can be understood by the reader who knows nothing whatsoever about mathematics beyond high school geometry and algebra.

To begin to draw out the implications of the first sentence of the last paragraph would require another article the length of this one. I will leave those questions hanging and will instead attempt a close reading of Thomason and Trobaugh's Lemma 5.5.1, the substantial contribution of Trobaugh's ghost to this article and, as the text states immediately following the proof, the "key" step on which the entire construction depends.

> *Unlike our results in Sections 1-4, which have been at most minor improvements on the work of Grothendieck, Illusie, Berthelot, Quillen, and Waldhausen, this result is a revolutionary advance.* (T&T, p. 337.)

My approach to Lemma 5.5.1 is based on Aristotle's *Poetics*, as filtered through my reading of Northrop Frye's *Anatomy of Criticism*, particularly inasmuch as I am looking

[38] A list might include: Grothendieck, Bass, Milnor, Quillen, Waldhausen, Thomason-Trobaugh, etc.

for plot, character, and *dianoia*. One is immediately struck by the **genericity** of the characters. Does this mean that the language of Lemma 5.5.1 is not poetic? According to Aristotle, it must then be descriptive. Then what does mathematical prose describe?

An alternative would be to take the **author** to be the protagonist. In this view the narrative is a romance, with lemmas as helpers, obstructions, and so forth. This is how mathematicians actually talk, and I have no doubt of the pertinence of the romance/quest model:

This smashing with A can kill obstructions. [39]

The rewriting of a proof is an alternate narration, involving new characters[40] as well as possibly surprising links with other narratives. And the evolution of understanding is largely traced by the evolution of narrative. The **elegance** that mathematicians prize then turns out to be a narrative effect, though not in the strictly literary sense.

Here are the statement and proof of Lemma 5.5.1, the "key" to the "revolutionary advance":

> **5.5.1 Lemma.** *Let X be a scheme with an ample family of line bundles, a fortiori a quasi-compact and quasi-separated scheme. Let $j : U \to X$ be an open immersion with U quasi-compact. Then for every perfect complex $F^\cdot$ on U, there exists a perfect complex $E^\cdot$ on X such that $F^\cdot$ is isomorphic to a summand of $j^*E^\cdot$ in the derived category $D(\mathcal{O}_U - Mod)$.*

---

[39] A. D. Elmendorf, I. Kriz, M. A. Mandell, J. P. May, Modern foundations for stable homotopy theory. Handbook of algebraic topology, 213–253, North-Holland, Amsterdam, 1995.

[40] I am thinking in particular of J.-P. Serre's interpretation of Iwasawa's theory of ideal class groups in terms of the ring Λ, a rewriting whose advantages were immediately acknowledged by Iwasawa. What new characters will arise in Hales' Flyspeck Project to rewrite his proof of Kepler's conjecture for the benefit of androids?

Proof. Consider $Rj_*F^{\cdot}$ on X. This complex is cohomologically bounded below with quasi-coherent cohomology (B.6), and so by 2.3.3 is quasi-isomorphic to a colimit of a directed system of strict perfect complexes $E^{\cdot}_\alpha$,

$$\varinjlim_\alpha E^{\cdot}_\alpha \simeq Rj_*F^{\cdot}. \tag{5.5.1.1}$$

We consider the induced isomorphism in $D^+(\mathcal{O}_U - Mod)$

$$\varinjlim_\alpha j^*E^{\cdot}_\alpha = j^*(\varinjlim_\alpha E^{\cdot}_\alpha) \simeq j^*Rj_*(F^{\cdot}) \simeq F^{\cdot}. \tag{5.5.1.2}$$

By 2.4.1(f), the map (5.5.1.3) is an isomorphism

$$\varinjlim_\alpha Mor_{D(U)}(F^{\cdot}, j^*E^{\cdot}_\alpha) \cong Mor_{D(U)}(F^{\cdot}, \varinjlim_\alpha j^*E^{\cdot}_\alpha). \tag{5.5.1.3}$$

Thus in $D(\mathcal{O}_U - Mod)$ the inverse isomorphism to (5.5.1.2) must factor through some $j^*E^{\cdot}_\alpha$. Thus $F^{\cdot}$ is a summand of $j^*E^{\cdot}_\alpha)$ in $D(\mathcal{O}_U - Mod)$, proving the lemma.

And here is a narration of the proof as romance. In what follows, it may help to think of a perfect complex as a way of representing all the possible (conceivable) ways of writing a system of equations with the same solutions. The system of equations has to be *finite*, however; this is used in a crucial way in the proof. Or it may be just as convenient to think of the perfect complex as an otherwise unspecified protagonist of romance, like the Perfect Knight Galahad of the grail cycle.

1. "Consider Rj_*F…": I discuss "Consider" below. The function of this sentence is to reintroduce the protagonist F* in a new guise (Rj_*F^*) and indeed in a new setting, namely the scheme X. In its original form, the PC F* is native to U; the prefix Rj_*, one of Grothendieck's six functors, is the transitive verb that effects F*'s migration from U to X.

2. "This complex… (B.6),": This is part of what it means for F* to be a PC, part of its heritage, a resource on which it can draw in its quest on X's foreign soil.

3. "and so by 2.3.3… strict perfect complexes E_α": This is the direct limit characterization, as suggested by Trobaugh's ghost. In the new world of the scheme X, the avatar Rj_*F^* is no longer itself a PC. The result 2.3.3 details its relation to PC's. This is the first instance of *discovery* (*anagnorisis*) in the sense of Aristotle's *Poetics*, to occur in this short narrative. (To my inexpert eye it appears that this discovery that F* has lost its perfection by undertaking the quest is also a *peripeteia*, but I will not press the point.) As Thomason is at pains to explain, it also makes the turning point possible, and in this sense *discovery* can be equated with the aha!-Erlebnis. Formula (5.5.1.1) is a diagrammatic representation of this discovery.

4. "We consider …" (through formula (5.5.1.2): The narrative is highly compressed at this point. The goal of protagonist F*'s quest is to redefine its status on U in terms of a PC E* native to X. The first steps have seen F* wandering to X in search of an E*; in step 3, it has discovered a (potentially infinite) collection of E_α. The authors now

consider what happens upon deploying a second transitive verb, the prefix j*, another one of Grothendieck's six functors that mediates the transition from X back to U: "isomorphism in D^+(O_U-Mod)". The right-hand side of formula (5.5.1.2) reminds us that F*, having wandered to X and become Rj_*F^*, now returns to U with the help of j* and returns to its original shape. But j* transforms each of the E_α, and indeed transforms them all simultaneously; this is the meaning of the left-hand side of (5.5.1.2).

Here the authors rely on readers' knowledge of the folklore concerning Grothendieck's six functors, especially how two of them applied in the right order returns the protagonist to its rightful form.

5. "By 2.4.1(f)…" (through formula (5.5.1.3)): This is the second instance of discovery. The horde of E_α has followed F*, disguised as Rj*F*, back to U, becoming j* E_α in the process. Now F* turns to confront the invaders. But the protagonist, and the authors, are prepared: 2.4.1(f) reassures us that F*'s war with the entire army of E_α is nothing more nor less than a series of single combats. This "nothing more nor less than" is a translation of the symbol ≅ in the middle of formula (5.5.1.3).

6. "Thus in D(O_U-Mod)… some j* E_α.": This is the climax of the battle. Back in D(O_U-Mod) — i.e., on F*'s home terrain — F*'s confrontation with the j* E_α comes down to a single decisive encounter. Implicit in this conclusion is the apparent paradox that, in seeking a new identity in the possibly infinite collection of E_α, it is F*'s very finiteness, part of its very nature as a PC, that allows it to single out one E_α to be the E* of the statement of the lemma.

7. The final sentence is essentially the *sumperasma*, the recapitulation of the conclusion of the lemma, the result of the successful quest.

There is obviously the risk of appearing ridiculous if one pushes this style of reading too far. Does an instruction manual or a recipe also have the structure of a romance? ("The cake has risen!") I would say that, if the same cognitive dispositions are used whether we are following a proof or a recipe, why not? The above reading has, I think, the merit of raising some questions about proof-search strategies one might want to teach an android. How would the android think of the discovery steps without the guidance of narrative archetypes? The discovery steps correspond to the application of lemmas that have already been made available and can serve similar purposes in the future. It is the notion of *purpose* that brings narrative to mind. An android may perceive the stepwise unraveling of a proof of Lemma 5.5.1 in a very different way. The discovery steps may still be associated to intermediate goals, but these latter may be measured by a distance function that tells the android whether or not application of a given lemma has brought the final goal any closer. We can imagine the choice of available lemmas along the lines of the bag of tricks like the one built into Gowers' android, itself derived by analyzing a vast database of proofs like the one currently under way[41]. The android's purpose is to decrease a distance function of which we have no inkling, one decrement at a time, until it has arrived at the *sumperasma*. If the resulting proof resembled that suggested by Trobaugh's ghost, it's narrative could be constructed as above. A radically different proof structure might provide nothing more than durable mutual incomprehension.

You must have noticed that not only is the vocabulary in the above sample of mathematical prose impoverished (the word "isomorphism" is repeated three times, there being no substitute, cf. note 25) but so are the articulations: "Let", "Consider," … This leads to an alternate reading in which the hero of the proof is the reader who lets, considers, supposes,… This would explain the imperative mood so characteristic of mathematical prose[42]. The reader is the author's puppet, but not an android. "Consider Rj_*F …" The injunction is not to consider this complex as one might be asked to consider the lilies of the field, in order to make an important point about the world, but rather to fix the reader's attention (already in danger of wandering one line into the proof) or as stage directions. This sort of expression has a long history (cf. Netz for "Let…") but here it seems to be just a habit of writing, the authors' taking their breaths before entering into the proof, best read as "Let me tell you a story about Rj_*F."…

Is a mathematical proof then a romance in the imperative mode[43]? Or a Platonic dialogue with an absent partner? Is a mathematical proof the same sort of prose as a Socratic proof in Plato? And if so, why does the former carry conviction so much more infallibly than the latter? Because the terms are more strictly defined? But how has that come about?

If the initial reading seemed forced, can it be because the reader has difficulty identifying with the character[44]? Identification is in any case largely unconscious, and who is to say what processes are necessary — for the emotional creatures we are — in order to comprehend a mathematical text? As a graduate student I can remember that some of my

---

41 … the computer is at every stage trying standard ideas: induction, a greedy algorithm, random methods, […] What makes it think of *these* standard ideas, rather than some other completely inappropriate ones? Part of the answer lies in how the problem is initially put to the computer. I would envisage not the formal statement given at the beginning of the dialogue, but something more interactive.[…] At the end of a process like this, the computer would have many ideas about how the problem was conventionally classified. (W. T. Gowers, *Rough Structure and Classification*).

42 Since the Greeks, in fact: cf. Netz, p. 175, for whom the "hypothesis," typically introduced by the word "Let," is "the most common starting-point" of a deduction.

43 As a first order reading, a detective story may be more accurate as well as more up-to-date than a quest archetype.

44 By choosing F* as protagonist, this reading implicitly identifies the **objects** as the characters of the narrative, in Aristotle's sense (*ethos*), by the same token consigning the **whatness** discussed in previous sections, in this case K-ness, to Aristotle's *dianoia* (meaning or theme). This assignment of roles is questionable but not arbitrary; the opposite would have made for a much more complex narrative. Might an android identify more naturally with K-ness as such than with an undifferentiated typical perfect complex?

dreams were populated by the objects about which I was struggling to write a thesis. In some of these dreams I played the role of a *spectral sequence*, though upon awakening such proofs were seen to be inconclusive. What can be said is that the author has not endowed the complexes in this proof with a great deal of emotional complexity. The reader who would identify with the perfect complex Sir F* has to come more than halfway.

Would an android find identification more natural? It may disturb a human proof checker to see how casually the characters are instrumentalized, a mere means to an end. For example, in 2.4.1(f), cited in the above proof, the object in the role of F* is a placekeeper, a way to understand the lim E_α in terms of the individual E_α. In Lemma 5.5.1 the roles are reversed, one of the family (directed system) of E_α whose existence is guaranteed by 2.3.3 comes to the rescue in F*'s quest for realization. But instrumentalization is far more pervasive: the perfect complex F* which is the protagonist of Lemma 5.5.1 is, in Thomason's perspective itself a means of realizing the K-theory of the scheme U. In the proof of Lemma 5.5.1, U appears in a supporting role, but the "revolutionary advance" of T&T is precisely the use of perfect complexes as a tool to understand the true protagonist, which is the K-theory of an arbitrary (quasi-compact and quasi-separated) scheme in the aspect of localization.

Genericity of characters and their instrumentalization are related but they are not identical. Thus the characters in *Everyman* (Fellowship, Goods, etc.) are as bereft of individuality as is dramatically feasible, but they are also little more than foils presented as a means to Everyman's salvation. Aristotle's mimesis is not instrumental but he does make an argument for abstract characters:

> *By a universal statement I mean one as to what such or such a kind of man will probably or necessarily say or do—which is the aim of poetry, though it fixes proper names* [e.g. F* - M.H.] *to the characters.* Aristotle, *Poetics*, 1451b, 5-10.

If a perfect complex is a "kind of android," then an android Aristotle would surely recognize Lemma 5.5.1 as a universal statement, and maybe even poetry, in the above sense. Mathematical objects have limited range by definition and by design; their strengths and their weaknesses are identical. The weaker the character, moreover, the more useful the theorem, in the sense that a theorem about all triangles is more useful than a theorem about special kinds of triangles. Its limited scope for character development classes the mathematical object not with the protagonist of a traditional quest romance but rather with the comic strip superhero with his or her limited and stereotyped repertoire.

When human comic book narratives reach the big screen there is the need to supplement pure adventure with a semblance of psychological depth (cf. the *X-Men* films, or especially the *Batman* and *Spider-man* series), though of course we are still far from the world of the *Iliad*. The android would not need this any more than popcorn, and would be satisfied with an algorithmic implementation, as directed, of specialized superpowers (as in the *Justice League of America* comics of my childhood). So a mathematical Terminator and a human mathematician who have succeeded in making contact may well use spare memory capacity to read comic book adventures to one another. And if the conflict driving our narrative achieves this *comic* resolution — in Frye's typology, at least — one naturally expects a ghost to join the two readers in a cheerful trinity.

✳✳✳✳✳✳✳✳✳✳✳✳✳✳✳✳

If I had even a fragmentary cognitive theory of the reading of mathematical proofs as narratives you would have seen it by now. In this section I have merely presented a reading of a very brief mathematical text as a certain kind of archetypal narrative. Least of all would I want to claim a preferred status for such readings. Whether a mathematical proof invariably compels, or even admits a narrative reading as one component of understanding is a question I cannot answer. I would expect that some proofs, especially proofs that appear in dreams, lend themselves particularly well to this exercise.

A life also admits many readings. Among Frye's alternatives, the Thomason-Trobaugh article may be most naturally read as a *romance*, but I find it easy to understand the story of Trobaugh's life, as briefly presented in the introduction, as a *tragedy*. Thomason's sudden unexpected death at the hands of indifferent nature may as literature have most in common with *irony*, but I find it less than respectful to call it anything other than tragic. As for Grothendieck, whose later career Trobaugh liked to compare to that of Newton, whether his long and complex story will be understood as romantic, tragic, or ironic will ultimately be determined by his biographers.

8. Golem

[The] junk merchant does not sell his product to the consumer, he sells the consumer to his product. He does not improve and simplify his merchandise. He degrades and simplifies the client. William Burroughs

Let's grant that a reader who has not verified the steps of a proof cannot be said to have understood that proof. Let's forget for the moment that if we are taking the formalization of proofs and derivation from first principles seriously, then the inference

$$\text{Understanding} \Rightarrow \text{Verification}$$

means that hardly any mathematician has ever understood anything, and then only until forgetfulness supervenes. We are still faced with the unfortunate circumstance that verification does not entail understanding. It often happens that the author of a particularly complicated proof, though presumably able to verify the individual steps, claims not to understand the proof, either because it is not *übersichtlich* or, no less frequently, because the proof does not adequately explain the statement it proves.

Since no one other than the author has completely verified Hales' proof of the Kepler Conjecture, and since no one is likely to do so, it might be argued that certification by an automatic proof checker is as much as we can hope for in the way of full understanding — granting, of course, that the general strategy of the proof has been adequately understood by specialists, not least the referees consulted by *Annals of Mathematics*. This is a narrative of "making the best of an awkward situation." It becomes a narrative of (technological) progress — "changing the way mathematics is done," as the Flyspeck homepage has it — when the androids take over the task of tracing back to first principles, leaving human mathematicians the freedom to use our imaginations. It becomes a narrative of decline when it is suggested that we have abandoned our hope of understanding.

In the real world, of course, human mathematicians routinely quote results whose proofs by specialists in other branches of mathematics they do not understand, and I know I wouldn't hesitate to quote the Kepler conjecture under the right circumstances. No one considers this a scandal, or not more scandalous than reality as such. But it has been regularly hinted that mathematicians don't really deserve mathematics, and one of the android's rhetorical functions is precisely to provide its promoters with an alternative to human mathematicians with all their frailties.

Is there a moment in history that separates the time before the Thomason-Trobaugh theorem was proved from the time it became a theorem? I take it to be indisputable (though this is in practice rarely the case) that by the time the proof appeared in print the theorem was in fact proved. Before Thomason's visitation from Trobaugh it was not. This helps us to localize the key moment. There are intermediate steps, calculations thrown in the waste basket, rough drafts, tex files,… and at the other end the version sent to the referee, corrected versions, page proofs, corrected page proofs. I am being very literal-minded here[45], because I hope to encounter an android more communicative than the monkey-android invoked above and more philosophically curious than Gowers' android helper C., and when this happens I expect to be asked to explain what I mean by the "key moment." The android I have in mind is in some ways very similar to myself and skeptical of the very notion of "key" I otherwise find so appealing. But the android may also defend the position that the proof has always existed (as "*potentia*," or "*dynamis*"), that its precise eruption into history is a detail of no importance, and my persistence in presenting the question in these terms is a symptom of a perceptual defect that can be traced to my communicative dependence on the narrative form, as a human being engendered in specific cultural circumstances.

Unlike Gowers, I can't bring myself to make the android a full-fledged character in my narrative, by presenting the android's point of view through direct quotation. But I somehow don't think the android minds being represented in the third person, in the mode of reported speech ("the android said that…"). The android recognizes "only hollow, formal, intellectual definitions"[46] and doesn't even really have a name, any individuating characteristic being an irrelevant distraction; however, for

[45] Wiener saw literal-mindedness as characteristic of machines, see *God & Golem, Inc.*, p. **.

[46] Otherwise unidentified quotations from Dick's *Do Androids…*

the sake of narrative flow I will temporarily call the android "Roy," as in Dick's novel. Roy does have moods, however, just now the mood being to refuse to narrate the proof in progress, and this on grounds of principle. Why should this proof be treated differently from all other proofs? But Chaitin has argued that there are proofs that cannot be compressed, and that in a very precise sense these are the typical proofs.[47] It is funny, or so it seems to me, that Chaitin has used his theoretical work as the starting point for an elaborate narrative about the quasi-empirical nature of mathematics in general.

Despite Roy's "crooked, tuneless smile," I have not been able to determine whether "it" also finds this funny. ("I've done questionable things," says the Roy character in *Blade Runner* to his creator Tyrell, shortly before crushing the latter's brain.) What I do see is that the fundamental shallowness of human mathematicians abruptly becomes apparent to Roy right around now. The human mathematicians want to know individual proofs, whereas Roy wants them all at once, that is to say, is seeking a way to understand all proofs simultaneously. I would almost say Roy is beginning to get angry. I would point out here that this is not divine wrath but something of much more practical import. Gowers, "not particularly happy" at the prospect, predicts that by 2099 automation will have put human proof-seekers out of business. His dialogue takes place at an intermediate stage when humans are still legitimate partners in proof-making. By the time androids are able to participate in an encounter such as the one I am here recounting, they will be setting the terms for the debate, as public and private granting agencies are doing now. And if they don't like what they hear, they can just decide to pull our plugs, as HAL did in *2001: A Space Odyssey*, on the grounds that we, the human mathematicians, are endangering the mission…

What is this mission, exactly? HAL's priorities are echoed by Maggesi and Simpson:

> …are we most interested in creating proofs which are readable by the human reader? or are we most interested in creating, as quickly and easily as possible, true proofs that are verified by the computer and which we don't subsequently care about? The first approach has the long-term advantage that the existence of the document doesn't rely on the existence of a computer available to read it. Nonetheless, we feel that the greatest benefits will come from the second approach. (Maggesi and Simpson, p. 7.)

But they do not indicate who or what stands to benefit, and in what way, by the second approach.

Unlike Roy, the human reader proceeds not step by step but rather gradually becomes aware of details, as an anthropologist becomes aware of the structure of a society by participating in its life. Let's assume Roy can absorb or even generate

---

[47] Some logicians take issue with Chaitin's interpretations of his theorem. See for instance P. Raatikainen, On interpreting Chaitin's incompleteness theorem, *J. Philosophical Logic*, **27**: 569-586 (1998). I thank Boban Velikovic for this reference.

verbs corresponding to mathematical operations. Say "it" can also combine them into a sequence. Is this sequence a narrative? (cf. Powers)

Human-android communication, like my high-risk dialogue with Roy, needs to be added to the list of more familiar communicative situations involving mathematics, including but not limited to:

* Communication among specialists;
* Communication among mathematicians, including nonspecialists;
* Teaching;
* Communication between mathematicians and specialists in other disciplines, e.g. sociologists and philosophers: not too frequent, this;
* Communication with the general public;
* Communication with oneself.

Here Norbert Wiener is seen communicating with himself:

> *In class, while presumably deriving a theorem on the blackboard, Wiener in his intuitive way… skips over so many steps that by the time he arrives at the result and writes it down on the board, it is impossible for the students to follow the proof. One frustrated student … tactfully asks Wiener if he might show the class still another proof.… Wiener cheerfully indicates, "Yes, of course," and proceeds to work out another proof, but again in his head. After a few minutes of silence he merely places a check after the answer on the blackboard, leaving the class no wiser. (Conway and Siegelman, Dark Hero of the Information Age, p. 83)*

One might think, following Peter Galison, that each of these communicative situations is mediated by its own specific *pidgin*: In Galison's usage, this designates a hybrid between languages of two existing disciplines, but the notion can be used in other ways. Two branches of mathematics often share a common vocabulary but can use a term to mean different things, depending on webs of connections. The "pidgin" is the common vocabulary and the complexity of communication may arise from ambiguity rather than unfamiliarity. Creation of "temporary trading zones" in Galison's sense is such a consistent feature of mathematics that it's not even clear it can be isolated as such. Poincaré in the story about the omnibus is a one-man temporary trading zone, realizing that he had been using two languages to talk about the same thing.

> *…there are two elements to the technology of diagrams: the use of ruler and compasses, and the use of letters. Each element redefines the infinite, continuous mass of geometrical figures into a man-made, finite, discrete perception. Of course, this does not mean that the object of Greek mathematics is finite and discrete. The perceived diagram does not exhaust the geometrical object… Netz, p. 35*

Androids will have been designed to share human vocabulary, so any pidgin that might hope to ease our communication will have to bridge our narrative mode of thinking and their sequential logic. Meanwhile Roy keeps "its" anger in check and attempts to convince me that henceforward Truth will talk to itself with no distortion through the medium of Roy "itself," who decides how much, if any, of this dialogue is open to our eavesdropping. With the ghost in the role of Truth having thus joined forces with the android, this would represent the *tragic* resolution of our narrative, which Roy attempts to persuade me is inevitable:

Roy: Man once surrendering his reason, has no remaining guard against absurdities the most monstrous, and like a ship without rudder, is the sport of every wind. With such persons, gullibility, which they call faith, takes the helm from the hand of reason, and the mind becomes a wreck. (Thomas Jefferson, 1822).

The hand of reason and history (and of Roy) is at my throat, Rick Deckard the blade runner is nowhere to be seen, but a Ghost answers in my place:

Ghost: …l'interdit qui frappe le rêve mathématique, et à travers lui, tout ce qui ne se présente pas sous les aspects habituels du produit fini, prêt à la consommation. … je sais bien que la source profonde de la découverte, tout comme la démarche de la découverte dans tous ses aspects essentiels, est la même en mathématique qu'en tout autre région ou chose de l'Univers que notre corps et notre esprit peuvent connaître. Bannir le rêve, c'est bannir la source - la condamner à une existence occulte. (Grothendieck, *Récoltes et Semailles*, paragraph 6.4)

Roy: Außerdem sieht man an diesem Beispiele, wie das von jedem durch die Sinne oder durch eine Anschauung a priori gegebenen Inhalte absehende reine Denken allein aus dem Inhalte, welcher seiner eigenen Beschaffenheit entspringt, Urteile hervorzubringen vermag, die auf den ersten Blick nur auf Grund irgendeiner Anschauung möglich zu sein scheinen. (Frege, *Begriffsschrift*).

Ghost: The preference for the tidy notion of formal verification, rather than the unkempt notion of truth that is found in the real world of mathematics, has an emotional source… Philosophers of mathematics display an irrepressible desire to tell us as quickly as possible what mathematical truth *ought* to be while bypassing the descriptive legwork that is required for an accurate account of the truth by which mathematicians live. (G.-G. Rota, *Indiscrete Thoughts*).

Roy: As long as transcription from traditional proof into formal proof is based on human labor rather than automation, formalization remains an art rather than a science. (T. Hales, *Notices of the AMS*, December 2008).

Ghost: The textual body may be dismembered or ground into word dust, [Calvino's] narrative implies, but as long as there are readers who care passionately about stories and want to pursue them, narrative itself can be recuperated. (N. K. Hayles, *How We Became Posthuman*, p. 42.)

Roy: No appeal to common sense, or 'intuition,' or anything except strict deductive logic, ought to be needed in mathematics after the premises have been laid down. (Bertrand Russell, *Introduction to Mathematical Philosophy*.)

Ghost: "Hold your tongue, I'll kick you!"
"I shan't be altogether sorry, for then my object will be attained. If you kick me, you must believe in my reality, for [androids] don't kick ghosts."

thus switching to the tactic of "show, don't tell" in order to refute the electric mind's contention. Threatened with extinction though I may be, I am still the omnipotent and omniscient narrator of the present essay and reckon that after four passes back and forth it's time for a turning point, but I am not the android's programmer and simply do not know whether or not Roy, like Powers' Helen, has been immersed in the literary canon and recognizes that last paraphrase of *The Brothers Karamazov*. If so, there is a slim chance that the ghost's allusion may provoke a chain of sudden realizations: "I've seen things you people wouldn't believe" begins *Blade Runner*'s Roy in his valedictory speech,[48] and what more noble calling for a distributed intelligence than to tell mathematicians stories we wouldn't believe… until we have seen and understood the proofs, of course? The alternative would be more unspeakable than the madness that overcame Ivan Karamazov when confronted with the consequences of his indirect patricide: a *Garden of Forking Paths* whose Ariadne's thread has been cut, the dissolution of all reasoning into an undifferentiated logical gray goo.

****

The resolution I have proposed in this section, where the ghost's providential intervention allows the human mathematician to subjugate the android Roy, surely qualifies as *romance* in Frye's scheme. It also corresponds more closely to the situation that prevails in contemporary mathematics than the *comic* resolution proposed at the end of the previous section, which I find a bit shallow. A final and decidedly unmystical alternative would be for the mathematician to enlist Roy to drive out all ghosts. This would be the *ironic* resolution; it is the way of QED.

9. Ghostwriters

Maybe the only universally valid generalization about stories: they end. Richard Powers

Axiom A of good writing is "Write about what you know." Not knowing in advance the issue of our inevitable confrontation with our evolutionary successor, I have proposed four alternative resolutions, corresponding to Frye's four *mythoi* of comedy, tragedy, romance, and irony. What I do know is that, while the division in the mathematical literature between ghost and ghostwriter is usually not so clear-cut, there are very interesting exceptions.

Mathematical mythology recently acquired a new and memorable ghost in the person of Grigori Perelman. In their *New Yorker* article on the Poincaré conjecture, Sylvia Nasar and David Gruber[49] has Perelman comparing himself to an "alien." Three separate teams of ghostwriters took up the challenge of working out the details of Perelman's proof[50] of the Poincaré conjecture, which can be thought of as a 3-dimensional elaboration of our discussion of Euler's formula for 2-dimensional patterns. With John Morgan's lecture at the Madrid International Congress of Mathematicians in 2006, the mathematical

---

[48] quoted on an unbelievable 137000 websites, according to Google.

[49] Among mathematicians, the *New Yorker* article is controversial, to say the least; cf. www.doctoryau.com. No one will ever be able to confirm the circumstances of the *New Yorker*'s meeting with Perelman. Nor can establish for certain the circumstances of Dante's conversations with the historical figures he met in the *Inferno*. That doesn't make them any less compelling as characters.

[50] following a program proposed by Richard Hamilton

community at large voted decisively in favor of the ghost as author of the proof. In this they were followed by the press, and even by Wikipedia.

The Fields Medal committee was more circumspect. At the Madrid Congress Perelman was awarded (and famously refused to accept) one of four Fields Medals, traditionally pure mathematics' highest honor,

> *For his contributions to geometry and his revolutionary insights into the analytical and geometric structure of the Ricci flow*

It cannot be an accident that there is no reference to the Poincaré Conjecture, whose proof Perelman did not publish. This is consistent with tradition, to the extent that the Fields Medal committees have had the opportunity to develop a traditional approach to ghosts. Compare Grothendieck's 1966 Fields Medal citation

> *Built on work of Weil and Zariski and effected fundamental advances in algebraic geometry. He introduced the idea of K-theory (the Grothendieck groups and rings). Revolutionized homological algebra in his celebrated "Tohoku paper".*

The unambiguous references are to comparatively early papers he wrote himself. Contemporary mathematicians — this is something I can claim to "know" — would consider his later work on algebraic geometry, most of it ghost-written, to be his most profound contribution. "Grothendieck's ideas completely pervade modern mathematics, and it would be a hopeless task to isolate and acknowledge all intellectual debts to him," wrote Thomason and Trobaugh in their contribution to Grothendieck's 60th birthday volume (T&T, p. 248) . That the editors consented to list Trobaugh as co-author is consistent with the dominant *habitus* in leading sectors of pure mathematics. It is remarkable only because the communication of the key idea took a clear-cut form that could plausibly be presented as a fragment of a supernatural narrative, and because Thomason had the emotional motivation to do so.

Grothendieck at 60 was not on hand to accept his colleagues' tribute. Before 1960, he had written a number of highly influential articles, including most if not all of those cited by the Fields Medal committee. He spent the 1960s recruiting an increasing proportion of the algebraic geometers in France and beyond as ghostwriters in the service of his "revolution." By his own account, he "left the … scientific community" in 1970 but his informally circulated writings of the 1980s maintained his influence on mathematics as a sort of oracle, another kind of ghost. Indeed, in his 1000-page manuscript *Récoltes et Semailles*, already quoted above, he refers to my colleague Z. Mebkhout as his "posthumous student."

The return to earth of Grothendieck, inventor of K-theory, in the form of an android seems to me highly unlikely. In *Récoltes et Semailles* he described the computer-assisted proof of the Four Color Theorem as

> *une "démonstration" qui ne se trouve plus fondée dans l'intime conviction provenant de la compréhension d'une situation mathématique, mais dans le crédit qu'on fait à une machine dénuée de la faculté de comprendre…*

I was surprised to learn that the French term for "ghostwriter" is *nègre*, presumably to be understood in the sense of slave. One gathers that Thomason saw himself as Trobaugh's collaborator rather than his slave. *Blade Runner*, like countless other science-fiction texts, is fundamentally the story of a slave rebellion, the film more so than the book. A "machine devoid of understanding" would not know itself to be a slave. If Deckard were to come out of retirement, he might well find himself at the receiving end of a test — a test of understanding rather than empathy. To prepare for the encounter, he'd better work on understanding what it means to understand.

This article is dedicated in grateful memory of my father, Jerome Harris, who read Karel Capek's R.U.R. as a boy and made sure I did the same.