12

Understanding Proofs

JEREMY AVIGAD

'Now, in calm weather, to swim in the open ocean is as easy to the practised swimmer as to ride in a spring-carriage ashore. But the awful lonesomeness is intolerable. The intense concentration of self in the middle of such a heartless immensity, my God! who can tell it? Mark, how when sailors in a dead calm bathe in the open sea—mark how closely they hug their ship and only coast along her sides.' (Herman Melville, *Moby Dick*, Chapter 94)

12.1 Introduction

What does it mean to understand mathematics? How does mathematics help us understand?

These questions are not idle. We look to mathematics for understanding, we value theoretical developments for improving our understanding, and we design our pedagogy to convey understanding to students. Our mathematical practices are routinely evaluated in such terms. It is therefore reasonable to ask just what understanding amounts to.

The issue can be addressed at different levels of generality. Most broadly, we need to come to terms with the sort of thing that understanding is, and the sort of thing that mathematics is, in order to discuss them in an appropriate manner. We can narrow our focus by noticing that the term 'understanding' is used in different ways; we can ask, for example, what it

Early versions of parts of this essay were presented at a conference, *La Preuve en Mathématique: Logique, Philosophie, Histoire*, in Lille, May 2005, and at a workshop organized by Ken Manders at the University of Pittsburgh in July 2005. I am grateful to the participants and many others for comments, including Andrew Arana, Mic Detlefsen, Jeremy Heis, Jukka Keranen, Paolo Mancosu, Ken Manders, John Mumma, Marco Panza, and Stewart Shapiro. I am especially grateful for sharp criticism from Clark Glymour, who still feels that Part I is a fuzzy and unnecessary framing of the otherwise promising research program surveyed in Part II.

means to understand a mathematical definition, a theorem, or a proof; or to understand a theory, a method, an algorithm, a problem, or a solution. We can, alternatively, focus our task by restricting our attention to particular types of judgments, such as historical or mathematical evaluations of theoretical developments, or pedagogical evaluations of teaching practices and lesson plans. We can be even more specific by considering individual objects of understanding and particular evaluatory judgments. For example, we can ask what it means to understand algebraic number theory, the spectral theorem for bounded linear operators, the method of least squares, or Gauss's sixth proof of the law of quadratic reciprocity; or we can try to explain how the introduction of the group concept in the 19th century advanced our understanding, or why the 'new math' initiative of the 1960s did not deliver the desired understanding to students. The way we deal with the specific examples will necessarily presuppose at least some conception of the nature of mathematical understanding; but, conversely, the things we find to say in specific cases will help us establish a more general framework.

In this chapter, I will defend the fairly simple claim that ascriptions of understanding are best understood in terms of the possession of certain abilities, and that it is an important philosophical task to try to characterize the relevant abilities in sufficiently restricted contexts in which such ascriptions are made. I will illustrate this by focusing on one particular type of understanding, in relation to one particular field of scientific search. Specifically, I will explore what it means to understand a proof, and discuss specific efforts in formal verification and automated reasoning that model such understanding.

This chapter is divided in two parts. In Part I, I will argue that the general characterization of understanding mentioned above provides a coherent epistemological framework, one that accords well with our intuitions and is capable of supporting rational inquiry in practical domains where notions of mathematical understanding arise. In Part II, I will present four brief case studies in formal verification, indicating areas where philosophical reflection can inform and be informed by contemporary research in computer science.

Part I. The nature of understanding

12.1 Initial reflections

A central goal of the epistemology of mathematics has been to identify the appropriate support for a claim to mathematical knowledge. We show that a

theorem is true by exhibiting a proof; thus, it has been a primary task of the philosophy of mathematics to clarify the notion of a mathematical proof, and to explain how such proofs are capable of providing appropriate mathematical knowledge. Mathematical logic and the theory of formal axiomatic systems have done a remarkably good job of addressing the first task, providing idealized accounts of the standards by which a proof is judged to be correct. This, in and of itself, does not address more difficult questions as to what ultimately justifies a particular choice of axiomatic framework. But, as far as it goes, the modern theory of deductive proof accords well with both intuition and mathematical practice, and has had practical applications in both mathematics and computer science, to boot.

But in mathematics one finds many types of judgment that extend beyond evaluations of correctness. For example, we seem to feel that there is a difference between *knowing that* a mathematical claim is true, and *understanding why* such a claim is true. Similarly, we may be able to convince ourselves that a proof is correct by checking each inference carefully, and yet still feel as though we do not fully understand it. The words 'definition' and 'concept' seem to have different connotations: someone may know the definition of a group, without having fully understood the group concept. The fact that there is a gap between knowledge and understanding is made pointedly clear by the fact that one often finds dozens of published proofs of a theorem in the literature, all of which are deemed important contributions, even after the first one has been accepted as correct. Later proofs do not add to our knowledge that the resulting theorem is correct, but they somehow augment our understanding. Our task is to make sense of this type of contribution.

The observation that modern logic fails to account for important classes of judgments traces back to the early days of modern logic itself. Poincaré wrote in *Science et Méthode* (1908):

> Does understanding the demonstration of a theorem consist in examining each of the syllogisms of which it is composed in succession, and being convinced that it is correct and conforms to the rules of the game? In the same way, does understanding a definition consist simply in recognizing that the meaning of all the terms employed is already known, and being convinced that it involves no contradiction?
>
> … Almost all are more exacting; they want to know not only whether all the syllogisms of a demonstration are correct, but why they are linked together in one order rather than in another. As long as they appear to them engendered by caprice, and not by an intelligence constantly conscious of the end to be attained, they do not think they have understood. (Book II, Chapter II, p. 118)

In that respect, logic does not tell us the whole story:

> Logic teaches us that on such and such a road we are sure of not meeting an obstacle; it does not tell us which is the road that leads to the desired end. (*Ibid.*, pp. 129–130)

Philosophers of science commonly distinguish between the 'logic of justification' and the 'logic of discovery'. Factors that guide the process of discovery also fall under the general category of 'understanding', and, indeed, understanding and discovery are often linked. For example, understanding a proof may involve, in part, seeing how the proof could have been discovered; or, at least, seeing how the train of inferences could have been anticipated.

> It seems to me, then, as I repeat an argument I have learned, that I could have discovered it. This is often only an illusion; but even then, even if I am not clever enough to create for myself, I rediscover it myself as I repeat it. (*Ibid.*, Part I, Chapter III, p. 50)

While knowing the relevant definitions may be enough to determine that a proof is correct, understanding is needed to find the definitions that make it possible to discover a proof. Poincaré characterized the process of discovery, in turn, as follows:

> Discovery consists precisely in not constructing useless combinations, but in constructing those that are useful, which are an infinitely small minority. Discovery is discernment, selection. (*Ibid.*, p. 51)

These musings provide us with some helpful metaphors. Mathematics presents us with a complex network of roads; understanding helps us navigate them, and find the way to our destination. Mathematics presents us with a combinatorial explosion of options; understanding helps us sift through them, and pick out the ones that are worth pursuing. Without understanding, we are lost in confusion, wandering blindly, unable to cope. When we do mathematics, we are like Melville's sailors, swimming in a vast expanse. Just as the sailors cling to sides of their ship, we rely on our understanding to guide us and support us.

12.2 Understanding and ability

Let us see if we can work these metaphors into something more definite. One thing to notice is that there seems to be some sort of reciprocal relationship between mathematics and understanding. That is, we speak of understanding

theorems, proofs, problems, solutions, definitions, concepts, and methods; at the same time, we take all these things to contribute to our understanding. This duality is reflected in the two questions I posed at the outset.

The way the questions are posed seem to presuppose that understanding is a relationship between an agent, who understands, and mathematics, which is the object of that understanding. On the surface, talk of knowledge shares this feature; it is an agent that knows that a theorem is true. But the role of an agent is usually eliminated from the epistemological account; once knowledge of a theorem is analyzed in terms of possession of a proof, proofs become the central focus of the investigation. To adopt a similar strategy here would amount to saying that an agent understands X just in case the agent is in possession of Y. But what sort of thing should Y be?

Suppose I tell you that my friend Paolo understands group theory, and you ask me to explain what I mean. In response, I may note that Paolo can state the definition of a group and provide some examples; that he can recognize the additive group structure of the integers, and characterize all the subgroups; that he knows Lagrange's theorem, and can use it to show that the order of any element of a finite group divides the order of the group; that he knows what a normal subgroup is, and can form a quotient group and work with it appropriately; that he can list all the finite groups of order less than 12, up to isomorphism; that he can solve all the exercises in an elementary textbook; and so on.

What is salient in this example is that I am clarifying my initial ascription of understanding by specifying some of the abilities that I take such an understanding to encompass. On reflection, we see that this example is typical: when we talk informally about understanding, we are invariably talking about the *ability*, or a capacity, to do something. It may be the ability to solve a problem, or to choose an appropriate strategy; the ability to discover a proof; the ability to discern a fruitful definition from alternatives; the ability to apply a concept efficaciously; and so on. When we say that someone understands we simply mean that they possess the relevant abilities.

Ordinary language is sloppy, and it would be foolish to seek sharp accounts of notions that are used in vague and imprecise ways. But notions of understanding also play a role in scientific claims and policy decisions that should be subject to critical evaluation. In more focused contexts like these, philosophical clarification can help further inquiry.

What I am proposing here is that in such situations it is often fruitful to analyze understanding in terms of the possession of abilities. This is a straightforward extension of the traditional epistemological view: the ability to determine whether a proof is correct is fundamental to mathematics, and the

standard theory has a lot to say as to what that ability amounts to. But this way of framing things enables us to address a much wider range of epistemological issues; verifying correctness is only a small part of understanding a proof, and we commonly speak of understanding other sorts of things as well. The task of answering our two main questions is thereby reduced to the task of describing and analyzing the interrelated network of abilities which constitute the practice of mathematics, *vis-à-vis* fields of inquiry that rely, either implicitly or explicitly, on models of that practice.

It may be argued that this proposal runs counter to our intuitions. Understanding is clearly needed to carry out certain mathematical tasks, but although understanding can explain the ability to carry out a task successfully, isn't it a mistake to conflate the two? Suppose I am working through a difficult proof. The argument is confusing, and I struggle to make sense of it. All of a sudden, something clicks, and everything falls into place—now I *understand*. What has just happened, and what has it got to do with ability?

We have all shared such 'Aha!' moments and the deep sense of satisfaction that comes with them. But surely the philosophy of mathematics is not supposed to explain this sense of satisfaction, any more than economics is supposed to explain the feeling of elation that comes when we find a $20 bill lying on the sidewalk. Economic theories describe agents, preferences, utilities, and commodities in abstract terms; such theories, when combined with social, political, psychological, or biological considerations, perfectly well explain the subjective appeal of cold, hard cash. In a similar way, we should expect a philosophical theory to provide a characterization of mathematical understanding that is consistent with, but independent of, our subjective experience.

Returning to the example above, what lies behind my moment of insight? Perhaps, all of a sudden, I see how to fill in a gap in the argument that had me puzzled. I may realize that the third line of the proof appeals to a prior lemma, which simply has to be instantiated appropriately; or that the claim follows easily from a general fact about, say, Hilbert spaces. Or perhaps, with Poincaré, I feel as though I understand how the proof could have been discovered; that is, I see why, in this situation, it is natural to consider the objects that have been introduced, or to express a term in the form in which is has been presented. Perhaps I see why a certain hypothesis in the theorem is necessary, and what would go wrong if the hypothesis were omitted. Perhaps I have grasped a general method in the structure of the argument, one that can fruitfully be applied in other situations. Perhaps I have realized that the argument is just like one that I am fully familiar with, straightforwardly adapted to the case at hand. These insights are perfectly well explained in terms of the acquisition

of abilities to supply missing inferences, draw appropriate analogies, discover other theorems, and so on. And, in turn, these abilities are just the sort of thing that should explain our pleasure at having understood: it is simply the pleasure of having acquired a new skill, or of finding ourselves capable of doing something we could not do before.

According to this analysis, when we say that someone understands something—a theorem, a problem, a method, or whatever—what we mean is that they possess some general ability. Such uses of a particular to stand for something more general are familiar. Suppose I tell you that Rebecca, a bright girl in Ms Schwartz's second grade class, can multiply 34 by 51. What do I mean by that? Surely I am not just attributing to her the ability to utter '1734' in response to the corresponding query; even the dullest student in the class can be trained to do that. We are often tempted to say that what we mean is that Rebecca is able to arrive at the answer '1734' by the right process, but a metaphysical commitment to 'processes' is easily avoidable. What we really mean is that Rebecca is capable of carrying out a certain type of arithmetical operation, say, multiplying numbers of moderate size. Phrasing that in terms of the ability to multiply 34 by 51 is just a convenient manner of speaking.

This way of thinking about understanding is not novel. In the next section, I will show that the strategy I have outlined here accords well with Wittgenstein's views on language, and I will situate the framework I am describing here with respect to the general viewpoint of the *Philosophical Investigations* (Wittgenstein, 1953). In the section after that, I will address some of the concerns that typically attend this sort of approach.

12.3 Mathematics as a practice

An important segment of the *Philosophical Investigations* explores what it means to follow a rule, as well as related notions, like obeying a command or using a formula correctly.[1] The analysis shows that, from a certain philosophical perspective, it is fruitless to hope for a certain type of explanation of the 'meaning' of such judgments. Nor is it necessary: there are philosophical gains

[1] I have in mind, roughly, sections 143 to 242, though the topic is foreshadowed in sections 81 to 88. The literature on these portions of the *Investigations* is vast, much of it devoted to critiquing Kripke's interpretation (1982); see, for example, Goldfarb (1985) and Tait (1986). The present chapter is essentially devoted to showing that a 'straightforward' reading of the *Investigations* has concrete consequences for the development of a theory of mathematical understanding. I owe much of my interpretation of Wittgenstein on rule following, and its use in making sense of a mathematical 'practice', to discussions with Ken Manders.

to be had by exploring the relationships between such fundamental judgments, and unraveling problems that arise from confused or misguided applications of the terms.

From the perspective of the *Investigations*, language is a community practice. It determines what can meaningfully be said, while the meaning of a word or sentence is, reciprocally, determined by the way that word or sentence functions in the practice. On that view, meaning is closely linked with the possibilities for use. Philosophical difficulties arise, however, when we try to explain the relationship between the two.

> When someone says the word 'cube' to me, for example, I know what it means. But can the whole *use* of the word come before my mind, when I *understand* it in this way?
>
> Well, but on the other hand isn't the meaning of the word also determined by this use? And can these ways of determining meaning conflict? Can what we grasp *in a flash* accord with a use, fit or fail to fit it? And how can what is present to us in an instant, what comes before our mind in an instant, fit a *use*? (Wittgenstein, 1953, §139)

The problem is that understanding is only made manifest in an infinite array of uses. The common philosophical tendency is therefore to distinguish the two, and take understanding to be 'possession' of a meaning that somehow 'determines' the appropriate usage.

> Perhaps you will say here: to have got the system (or again, to understand it) can't consist in continuing the series up to *this* or *that* number: *that* is only applying one's understanding. The understanding itself is a state which is the *source* of correct use. (§146)

But Wittgenstein urges us against this way of thinking.

> If one says that knowing the ABC is a state of the mind, one is thinking of a state of a mental apparatus (perhaps of the brain) by means of which we explain the *manifestations* of that knowledge. Such a state is called a disposition. But there are objections to speaking of a state of mind here, inasmuch as there ought to be two different criteria for such a state: a knowledge of the construction of the apparatus, quite apart from what it does … (§149)

The discussion in the *Investigations* aims to convince us that this last way of framing matters is problematic. For example, we may attribute someone's ability to continue a sequence of numbers correctly to the fact that he has grasped the right pattern. But substituting the phrase 'grasping the correct pattern' for 'understanding' is little more than word play, unless we say more about what has been 'grasped'. The appropriate pattern may, perhaps, be described by an algebraic formula, and so, at least in some cases, we may

explain the ability to continue the sequence in terms of knowing the correct formula. But then we are left with the task of explaining how he is able to apply the algebraic formula correctly. It is not enough to say that the formula simply 'occurs to him' as he produces the desired behavior; perhaps he will continue to think of the formula and do something entirely unexpected at the next step. So we have simply replaced the problem of explaining what it means to understand how to continue the sequence with the problem of explaining what it means to understand how to apply a formula correctly. To make matters worse, there may be other ways in which we can account for the person's ability to continue the sequence according to the pattern we have in mind; or we may find that the person is simply able to do it, without being able to explain how.

> We are trying to get hold of the mental process of understanding which seems to be hidden behind those coarser and therefore more readily visible accompaniments. But we do not succeed; or, rather, it does not get as far as a real attempt. For even supposing I had found something that happened in all those cases of understanding,—why should *it* be the understanding? … (§153)

The solution is, in a sense, just to give up. In other words, we simply need to resist the temptation to find a suitable 'source' for the behavior.

> If there has to be anything 'behind the utterance of the formula' it is *particular circumstances,* which justify me in saying that I can go on—when the formula occurs to me.
>
> Try not to think of understanding as a 'mental process' at all.—For *that* is the expression which confuses you. But ask yourself: in what sort of case, in what kind of circumstances, do we say, 'Now I know how to go on,' when, that is, the formula *has* occurred to me?—
>
> In the sense in which there are processes (including mental processes) which are characteristic of understanding, understanding is not a mental process. (§154)

If our goal is to explain what it means to say that someone has understood a particular word, formula, or command, we simply need to describe the circumstances under which we are willing to make this assertion. In doing so, we may find that there is a good deal that we can say that will clarify our meaning. Giving up the attempt to identify understanding as some sort of *thing* doesn't mean that we cannot be successful, by and large, in explaining what understanding amounts to.

Thus, from a Wittgensteinian perspective, the philosopher's task is not to explain the feeling of having understood, or any underlying mental or physical processes. The challenge, rather, is to clarify the circumstances under we which we make our ascriptions.

> … when he suddenly knew how to go on, when he understood the principle, then possibly he had a special experience … but for us it is *the circumstances* under which he had such an experience that justify him in saying in such a case that he understands, that he knows how to go on. (§155)

We should not be distressed by the fact that our ascriptions of understanding are fallible. I may reasonably come to believe that a student understands the fundamental theorem of calculus, but then subsequently change my mind after grading his or her exam. This does not in any way preclude the utility of trying to explain what it means to 'understand the fundamental theorem of calculus', in terms of what we take such an understanding to entail.

The aim of the *Investigations* is to shape the way we think about language and thought. Here, I have proposed that this world view is relevant to the way we think about mathematics. The nature of our more specific goals does, however, impose some important differences in emphasis. In the excerpts I have quoted, Wittgenstein is primarily concerned with exploring what it means to follow a rule, obey a command, or use a word *correctly*. When it comes to the philosophy of mathematics, I believe it is also fruitful to explore what we take to constitute *appropriate* behavior, even in situations where we take a goal or a standard of correctness to be fixed and unproblematic. For example, we may agree, for the moment, to take provability in some formal axiomatic theory to provide an appropriate standard of correctness, and then ask what types of abilities are appropriate to finding the proofs we seek. Or we may be in a situation where we have a clear notion as to what counts as the solution to a particular problem, and then wonder what type of understanding is needed to guide a student to a solution. Of course, some goals of the practice, like finding 'natural' definitions or 'fruitful' generalizations, are harder to characterize. And the distinction between goals and the methods we use to achieve them blur; the goals of finding natural definitions and fruitful generalizations can also be interpreted as means to further goals, like solving problems and proving theorems. The network of goals is complex, but we need not chart the entire territory at once; by focusing on particular phenomena of interest we can start by mapping out small regions. The claim I am making here is simply that the terrain we are describing is best viewed as a network of abilities, or mechanisms and capacities for thought.

Indeed, the *Investigations* is not only concerned with questions of correctness. The work is, more broadly, concerned with the effective use of language with respect to our various goals and ends.

> Language is an instrument. Its concepts are instruments. Now perhaps one thinks that it can make no *great* difference *which* concepts we employ. As, after all, it is

> possible to do physics in feet and inches as well as in metres and centimetres; the difference is merely one of convenience. But even this is not true if, for instance, calculations in some system of measurement demand more time and trouble than is possible for us to give them. (§569)
>
> Concepts lead us to make investigations; are the expressions of our interest, and direct our interest. (§570)

One finds similar views on the role of specifically mathematical concepts in Wittgenstein's other works. For example, we find the following in the *Remarks on the Foundations of Mathematics*:

> The mathematical Must is only another expression of the fact that mathematics forms concepts.
>
> And concepts help us to comprehend things. They correspond to a particular way of dealing with situations.
>
> Mathematics forms a network of norms. (Wittgenstein, 1956, VI, §67)

This stands in contrast to the traditional view of mathematics as a collection of definitions and theorems. For Wittgenstein, a proposition is not just an object of knowledge, but, rather, something that shapes our behavior:

> The mathematical proposition says to me: Proceed like this! (§72)

With respect to propositions in general, we find in *On Certainty*:

> 204. Giving grounds, however, justifying the evidence, comes to an end;—but the end is not certain propositions' striking us immediately as true, i.e. it is not a kind of *seeing* on our part, it is our *acting*, which lies at the bottom of the language game. (Wittgenstein, 1969)

This way of thinking challenges us to view mathematics in dynamic terms, not as a body of knowledge, but, rather, as a complex system that guides our thoughts and actions. We will see in Part II of this essay that this provides a powerful and fundamentally useful way of thinking about the subject.

12.4 A functionalist epistemology

I have proposed that a theory of mathematical understanding should be a theory of mathematical abilities. In ordinary circumstances, when we say, for example, that someone understands a particular proof, we may take them to possess any of the following:

- the ability to respond to challenges as to the correctness of the proof, and fill in details and justify inferences at a skeptic's request;

- the ability to give a high-level outline, or overview of the proof;
- the ability to cast the proof in different terms, say, eliminating or adding abstract terminology;
- the ability to indicate 'key' or novel points in the argument, and separate them from the steps that are 'straightforward';
- the ability to 'motivate' the proof, that is, to explain why certain steps are natural, or to be expected;
- the ability to give natural examples of the various phenomena described in the proof;
- the ability to indicate where in the proof certain of the theorem's hypotheses are needed, and, perhaps, to provide counterexamples that show what goes wrong when various hypotheses are omitted;
- the ability to view the proof in terms of a parallel development, for example, as a generalization or adaptation of a well-known proof of a simpler theorem;
- the ability to offer generalizations, or to suggest an interesting weakening of the conclusion that can be obtained with a corresponding weakening of the hypotheses;
- the ability to calculate a particular quantity, or to provide an explicit description of an object, whose existence is guaranteed by the theorem;
- the ability to provide a diagram representing some of the data in the proof, or to relate the proof to a particular diagram;

and so on. The philosophical challenge is to characterize these abilities with clarity and precision, and fit them into a structured and informative theory. Thanks to our Wittgensteinian therapy, we will not let the phrase 'possess an ability' fool us into thinking that there is anything mysterious or metaphysically dubious about this task. We have serious work to do, and worrying about what sort of thing is being 'possessed' is an unnecessary distraction.

And yet we may still be plagued by qualms. Our analysis entails that understanding only becomes manifest in an agent's behavior across a range of contexts, and we seem to have come dangerously close to identifying understanding with the class of relevant behaviors. Such a 'dispositional' or 'behavioral' account of understanding has famously been put forth by Gilbert Ryle (1949) as part of a more general philosophy of mind. Since Ryle's approach is commonly viewed as having failed, it is worth reviewing some of the usual criticisms, to see what bearing they have on the more specific issues addressed here.[2]

[2] These criticisms are enumerated, for example, in Carr (1979).

Ryle intended his dispositional theory to account for ascriptions of a variety of mental states, including things like belief, desire, intent, and so on. He begins his account, however, with a discussion of ascriptions of 'knowing how'. He has been criticized, in this respect, for failing to distinguish between knowing how to perform an action, and the ability to do so. For example, it seems to reasonable to say that an arthritic piano player knows how to play the *Moonlight Sonata,* and that an injured gymnast knows how to perform a back flip, even if they are temporarily or permanently unable to do so. Here, we may have similar concerns that there may be situations under which it makes sense to say that someone understands a proof, but is unable to exhibit the expected behaviors. But the examples that come to mind are contrived, and it does not seem unreasonable to declare these outside the scope of a suitably focused theory of mathematical understanding. If we view at least the outward signs of mathematical activity as essentially linguistic, it seems reasonable to take verbal and written communication as reliable correlates of understanding.

A further critique of Ryle's analysis is a pragmatic one. Ryle imagines, for example, characterizing mental states, like hunger, in terms of dispositions to behave in certain ways, like opening the refrigerator door when one is in the kitchen. But one would not expect an agent to open the refrigerator door if he or she held the belief that the door was wired to an explosive device, and so circumstances like that need to be excluded. To start with, it is unsettling that one may have to characterize such contexts in terms of other mental states, like the agent's beliefs, when that is what a dispositional account is designed to avoid. But even more significantly, it seems unlikely that one can characterize the rogue circumstances that need to be excluded from the class of contexts in which we expect to observe the characteristic behavior. To be sure, mitigating factors like the one I described above are patently irrelevant to hunger, and one would like to exclude them with a suitable *ceteris paribus* clause. But this is exactly the point: it is hard to see how one can gauge relevance prior to having some kind of understanding of what it means to be hungry.

With respect to the topic at hand, these concerns translate to doubts that one can adequately characterize the behaviors that warrant attributions of understanding. But, once again, the problem can be mitigated by limitations of scope. Our theory need not account for the full range of human behaviors, as a theory of mind ought to do. We would like our theory to help explain why certain proofs are preferred in contemporary mathematics, why certain historical developments are viewed as advances, or why certain expository practices yield desired results. Moving from a general theory of mind to a more specific theory of mathematical understanding gives us great latitude in bracketing issues that we take to fall outside our scope. We should be able

to screen off extraneous beliefs and desires, and assume nothing about an agent's intent beyond the intent to perform mathematically. I am certainly not claiming that it is obvious that one can provide adequate characterizations of the circumstances under which we tend to ascribe mathematical understanding; only that it is not obvious that attempts to do so are doomed to failure.

Perhaps the most compelling criticism of a dispositional account is that even if we could characterize the behaviors that are correlated with the mental states under consideration, identifying the mental states with the associated behaviors simply tells the wrong kind of story. We expect a philosophical theory to provide some sort of causal explanation that tells us how intelligent and intentional behavior is brought about; it seems unsatisfying to identify 'knowing how to play the piano' with successful performance, when what one really wants of a theory is an account of the mental activity that makes such performance possible. In the case at hand, we would like a theory that explains how a proper understanding enables one to function mathematically. This insistence has not only an intuitive appeal, but also a pragmatic one. For example, in so far as our theory is to be relevant to mathematical exposition and pedagogy, we would expect it not only to characterize the outward signs of mathematical understanding, but also provide some hints as to how they can be encouraged and taught. Similarly, I take it that a theory of mathematical understanding should be of service to computer scientists trying to write software that exhibits various types of competent mathematical behavior. Even if we set aside the question as to whether it is appropriate to attribute 'understanding' to a computer, we might expect a good philosophical theory not just to clarify and characterize the desired behaviors, but also to provide some guidance in bringing them about.

We are therefore tempted to renounce our therapy and try, again, to figure out just what understanding *really* is. What saves us, however, is the observation that our theory of mathematical abilities need not degenerate to a laundry list of behavioral cues. The abilities we describe will interact in complex ways, and will not always be cast in terms of behavioral manifestations. Consider our explanation of what it means for Paolo to understand group theory. Some of the relevant abilities may be cast in terms of behaviors, for example, the ability to state a theorem or answer a question appropriately. But others may be cast in more abstract terms, such as the ability to 'recognize' a group structure, 'determine' subgroups and cosets, 'apply' a lemma, or 'recall' a fundamental fact. In fact, we often take these abstract abilities to provide the 'mechanisms' that explain the observable behaviors. We may be relieved to learn that our Wittgensteinian training does not preclude talk of mechanisms, provided that we keep in mind that 'these mechanisms are only hypotheses, models designed

to explain, to sum up, what you observe' (Wittgenstein, 1953, §156). What gives our theoretical terms meaning is the role they play in explaining the desired or observed behaviors; 'An "inner process" stands in need of outward criteria' (Wittgenstein, 1953, §580). While these outward criteria are necessary, they are also sufficient to give our terms meaning. So, we can once again set our metaphysical qualms aside.

What we are left with is essentially a functionalist approach to explaining various aspects of mathematical understanding. The fundamental philosophical challenge is to develop a language and conceptual framework that is appropriate to our goals. If you want an explanation of how a car works, a description of the subsystems and their components, situated against general understanding as to how these interact, may be just what you need to keep your car running smoothly, and to diagnose problems when they arise. A more fine-grained description is more appropriate if you are studying to be a mechanic or engineer. What we are seeking here are similar explanations of how mathematical understanding works. In this case, however, our intuitions as to how to talk about the relevant subsystems and components is significantly poorer. I have argued that a theory of mathematical abilities and their relationships should do the trick, but, at this point, the proposal is vague. The only way to make progress is to pay closer attention to the data that we are trying to explain, and to the particular aims that our explanations are to serve.

Part II. Formal verification

12.5 The nature of proof search

In Part I, I described a general way of thinking about mathematical understanding. My goal in Part II is to show that this way of thinking is fruitful in at least one scientific context where informal notions of understanding are used. In doing so, I will consider only one small aspect of mathematical understanding, with respect to one particular scientific practice. While I expect that the general perspective will be useful in other domains as well, and that the problems that arise share enough common structure that they can be supported by a unified conceptual framework, I cannot make this broader case here. So I ask you to keep in mind that, in what follows, we are considering only one restricted example.

We have seen that understanding an ordinary textbook proof involves, in part, being able to spell out details that are left implicit in the presentation. I have argued elsewhere (Avigad, 2006) that it is hard to make sense of

this aspect of understanding in terms of traditional logical analyses. Formal axiomatic deduction provides a model of proof in which *every* detail is spelled out precisely, in such a way that correctness boils down to pattern matching against a manageable list of precisely specified rules. In contrast, ordinary textbook proofs proceed at a higher level, relying on the reader's ability to 'see' that each successive step is warranted. In order to analyze this capacity, we need a model of proof on which such 'seeing' is a nontrivial affair.

Coming to terms with the ability to understand higher-level proofs is a central task in the field of formal verification and automated reasoning. Since early in the 20th century, it has been understood that, at least in principle, mathematical proof can be modeled by formal axiomatic deduction. In recent years, a number of computational 'proof assistants' have been developed to make such formalization feasible in practice (see Wiedijk, 2006). In addition to verifying mathematical assertions, such systems are sometimes developed with the goal of verifying that (mathematical descriptions of) hardware and software systems meet their specifications, or are free from dangerous bugs. Since, however, such systems and specifications are modeled in mathematical terms, the two efforts overlap, to a large extent.

A user's interaction with such a system can be seen as an attempt to provide the computer with enough information to see that there is a formal axiomatic proof of the purported theorem. Alternatively, the formal 'proof scripts' that are given to the computer can be viewed (like informal proofs) as providing instructions as to how to find a formal axiomatic derivation. The task of the computational proof assistant is to use these scripts to construct such a derivation. When it has done so, the system indicates that it has 'understood' the user's proof by certifying the theorem as correct. In Avigad (2006), I consider a range of informal epistemological judgments that are not easily explicated on the standard logical models, and argue that 'higher-level' notions of proof, akin to the scripts just described, are better equipped to support the relevant judgments.

Proof assistants like Coq, Isabelle, and HOL-light provide a style of proof development that allows the user to view the task of theorem proving in a goal-driven manner. Stating a theorem can be seen as a way of announcing the goal of proving it. Each step in a proof then serves to reduce the currently open goals to ones that are (hopefully) simpler. These goals are often represented in terms of *sequents*. For example, if A, B, C, and D are formulas, the sequent $A, B, C \Rightarrow D$ represents the goal of showing that D follows from A, B, and C. The most basic steps correspond to logical rules. For example, the 'and introduction' rule reduces the goal $A, B, C \Rightarrow D \wedge E$ to the pair of goals $A, B, C \Rightarrow D$ and $A, B, C \Rightarrow E$. This corresponds to the situation where, in

an ordinary proof, we have assumed or established A, B, and C, and need to prove an assertion of the form 'D and E'; we can do this by noting that 'it suffices to establish D, and then E, in turn' and then accomplishing each of these two tasks. We can also work forwards from hypotheses: for example, from $A \wedge B, C \Rightarrow D$ we can conclude $A, B, C \Rightarrow D$. In ordinary terms, if we have established or assumed 'A and B', we may use both A and B to derive our conclusion. A branch of the tree is closed off when the conclusion of a sequent matches one of the hypotheses, in which case the goal is clearly satisfied.

Things become more interesting when we try to take larger inferential steps, where the validity of the inference is not as transparent. Suppose, in an ordinary proof, we are trying to prove D, having established A, B, and C. In sequent form, this corresponds to the goal of verifying $A, B, C \Rightarrow D$. We write 'Clearly, from A and B we have E', thus reducing our task of verifying $A, B, C, E \Rightarrow D$. But what is clear to us may not be clear to the computer; the assertion that E follows from A and B corresponds to the sequent $A, B \Rightarrow E$, and we would like the computer to fill in the details automatically. 'Understanding' this step of the proof, in this context, means being able to justify the corresponding inference.

In a sense, verifying such an inference is no different from proving a theorem; a sequent of the form $A, B \Rightarrow E$ can express anything from a trivial logical implication to a major conjecture like the Riemann hypothesis. Sometimes, brute-force calculation can be used to verify inferences that require a good deal of human effort. But what is more striking is that there is a large class of inferences that require very little effort on our part, but are beyond the means of current verification technology. In fact, most textbook inferences have this character: it can take hours of painstaking work to get a proof assistant to verify a short proof that is routinely read and understood by any competent mathematician.

The flip explanation as to why competent mathematicians succeed where computers fail is simply that mathematicians understand, while computers don't. But we have set ourselves precisely the task of explaining how this understanding works. Computers can search exhaustively for an axiomatic derivation of the desired inference, but a blind search does not get very far. The problem is that even when the inferences we are interested in can be justified in a few steps, the space of possibilities grows exponentially.

There are a number of ways this can happen. First, excessive case distinctions can cause problems. Proving a sequent of the form $A \vee B, C, D \Rightarrow E$ reduces to showing that the conclusion follows from each disjunct; this results in two subgoals, $A, C, D \Rightarrow E$ and $B, C, D \Rightarrow E$. Each successive case distinction

again doubles the number of subgoals, so that, for example, 10 successive case splits yield 1024 subgoals.

Second, proving an existential conclusion, or using a universal hypothesis, can require finding appropriate instances. For example, one can prove a sequent of the form $A, B, C \Rightarrow \exists x\, D(x)$ by proving $A, B, C \Rightarrow D(t)$ for some term t; or possibly $A, B, C \Rightarrow D(t_1) \vee \ldots \vee D(t_k)$ for a sequence of terms $t_1, \ldots, t_k$. Dually, proving a sequent of the form $\forall x\, A(x), B, C \Rightarrow D$ can involve finding appropriate instances $t_1, \ldots, t_k$ for the universal quantifier, and then proving $A(t_1), \ldots, A(t_k), B, C \Rightarrow D$. The problem is that there may be infinitely many terms to consider, and no *a priori* bound on k. In situations like this, search strategies tend to defer the need to choose terms as long as possible, using Skolem functions and a procedure known as 'unification' to choose useful instantiations. But even with such methods, there are choices to be made, again resulting in combinatorial explosion.

Finally, common mathematical inferences typically require background facts and theorems that are left implicit. In other words, justifying an inference usually requires a background theory in addition to purely logical manipulations. In that case, the challenge is to determine which external facts and theorems need to be imported to the 'local' list of hypotheses. Once again, the range of options can render the problem intractable.

All three factors—case disjunctions, term instantiation, and the use of a background theory—are constantly in play. For example, suppose we are reasoning about the real numbers. The trichotomy law states that for any x and y, one of the expressions $x < y$, or $x = y$, or $y < x$ must hold. At any point, a particular instantiation of this law may be just the thing needed to verify the inference at hand. But most of the time, choosing terms blindly and splitting across the disjunction will be nothing more than an infernal waste of time.

Despite all this, when we read a proof and try to fill in the details, we are somehow able to go on. We don't guess terms blindly, or pull facts at random from our shelves. Instead, we rely on our understanding to guide us. At a suitable level of abstraction, an account of how this understanding works should not only explain the efficacy of our own mathematical practices, but also help us fashion computational systems that share in the success. This last claim, however, is often a sticking point. One may object that the way that we, humans, understand proofs is different from the ways in which computers should search for proofs; determining the former is the task of cognitive psychologists, determining the latter is the task of computer scientists, and philosophy has nothing to do with it. Ordinary arithmetic calculation provides an analogy that seems to support this distinction. Cognitive psychologists can determine how many digits we can hold in our short term memory, and use

that to explain why certain procedures for long multiplication are effective. Computer scientists have developed a clever method called 'carry save addition' that makes it possible for a microprocessor to perform operations on binary representations in parallel, and therefore multiply numbers more efficiently. Each theory is fruitfully applied in its domain of application; no overarching philosophy is needed.

But this objection misses the point. Both the experimental psychologist and the computer scientist presuppose a normative account of what it means to multiply correctly, and our foundational accounts of arithmetic apply equally well to humans and machines. In a similar way, a theory of mathematical understanding should clarify the structural aspects of mathematics that characterize successful performance for agents of both sorts. To be sure, when it comes to verifying an inference, different sorts of agents will have different strengths and weaknesses. We humans use diagrams because we are good at recognizing symmetries and relationships in information so represented. Machines, in contrast, can keep track of gigabytes of information and carry out exhaustive computation where we are forced to rely on little tricks. If we are interested in differentiating good teaching practice from good programming practice, we have no choice but to relativize our theories of understanding to the particularities of the relevant class of agents. But this does not preclude the possibility that there is a substantial theory of mathematical understanding that can address issues that are common to both types of agent. Lightning fast calculation provides relatively little help when it comes to verifying common mathematical inferences; blind search does not work much better for computers than for humans. Nothing, *a priori*, rules out that a general philosophical theory can help explain what makes it possible for either type of agent to understand a mathematical proof.

In the next four sections, I will consider four types of inference that are commonly carried out in mathematical proofs, and which, on closer inspection, are not as straightforward as they seem. In each case, I will indicate some of the methods that have been developed to verify such inferences automatically, and explore what these methods tell us about the mechanisms by which such proofs are understood.

12.6 Understanding inequalities

I will start by considering inferences that are used to establish inequalities in ordered number domains, like the integers or the real numbers. For centuries,

mathematics was viewed as the science of magnitude, and judgments as to the relative magnitude of various types of quantities still play a key role in the subject. In a sense, methods of deriving *equalities* are better developed, and computer algebra systems carry out symbolic calculations quite effectively. This is not to say that issues regarding equality are trivial; the task of determining whether two terms are equal, in an axiomatic theory or in an intended interpretation, is often difficult or even algorithmically undecidable. One strategy for determining whether an equality holds is to find ways of putting terms into canonical 'normal forms', so that an assertion $s = t$ is then valid if and only if s and t have the same normal form. There is an elaborate theory of 'rewrite systems' for simplifying terms and verifying equalities in this way, but we will not consider this here.

Let Γ be a set of equalities and inequalities between, say, real or integer-valued expressions, involving variables $x_1, \ldots, x_n$. Asking whether an inequality of the form $s < t$ is a consequence of Γ is the same as asking whether Γ together with the hypothesis $t \leq s$ is *un*satisfiable. Here, too, there is a well-developed body of research, which provides methods of determining whether such a system of equations is satisfiable, and finding a solution if it is. This research falls under the heading 'constraint programming', or, more specifically, 'linear programming', 'nonlinear programming', 'integer programming', and so on. In automated reasoning, such tasks typically arise in connection to scheduling and planning problems, and heuristic methods have been developed to deal with complex systems involving hundreds of constraints.

But the task of verifying the entailments that arise in ordinary mathematical reasoning has a different character. To start with, the emphasis is on finding a proof that an entailment is valid, rather than finding a counterexample. Often the inequality is tight, which means that conservative methods of approximation will not work; and the structure of the terms is often more elaborate than those that arise in industrial applications. On the other hand, the inference may involve only a few hypotheses, in the presence of suitable background knowledge. So the problems are generally smaller, if structurally more complex.

Let us consider two examples of proofs involving inequalities. The first comes from a branch of combinatorics known as *Ramsey theory*. An (undirected) *graph* consists of a set of vertices and a set of edges between them, barring 'loops', i.e. edges from a vertex to itself. The *complete graph on n vertices* is the graph in which between any two vertices there is an edge. Imagine coloring every edge of a complete graph either red or blue. A collection of k points is said to be *homogeneous* for the coloring if either all the edges between the points are red, or all of them are blue. A remarkable result due to F. P. Ramsey is that for any

value of k, there is an n large enough, so that no matter how one colors the the complete graph of n vertices edges red and blue, there is a homogeneous subset of size k.

This raises the difficult problem of determining, for a given value of k, how large n has to be. To show that, for a given k, a value of n is not large enough means showing that there is a graph of size n with no homogeneous subset of size k. Paul Erdös pioneered a method, called *the probabilistic method*, for providing such lower bounds: one imagines that a coloring is chosen at random from among all colorings of the complete graph on n vertices, and then one shows that with nonzero probability, the graph will have no homogeneous subset of size k.

Theorem 1. *For all $k \geq 2$, if $n < 2^{k/2}$, there is a coloring of the complete graph on n vertices with no homogeneous subset of size k.*

Proof. For $k = 2$ this is trivial, and for $k = 3$ this is easily verified by hand. So we can assume $k \geq 4$.

Suppose $n < 2^{k/2}$, and consider all red–blue colorings, where we color each edge independently red or blue with probability 1/2. Thus all colorings are equally likely with probability $2^{-\binom{n}{2}}$. Let A be a set of vertices of size k. The probability of the event A_R that the edges in A are all colored red is then $2^{-\binom{k}{2}}$. Hence it follows that the probability p_R for some k-set to be colored all red is bounded by

$$p_R = Prob \bigcup_{|A|=k} A_R \leq \sum_{|A|=k} Prob(A_R) = \binom{n}{k} 2^{-\binom{k}{2}}.$$

Now for $k \geq 2$, we have $\binom{n}{k} = \frac{n(n-1)(n-2)\cdots(n-k+1)}{k(k-1)\cdots 1} \leq \frac{n^k}{2^{k-1}}$. So for $n < 2^{k/2}$ and $k \geq 4$, we have

$$\binom{n}{k} 2^{-\binom{k}{2}} \leq \frac{n^k}{2^{k-1}} 2^{-\binom{k}{2}} < 2^{\frac{k^2}{2} - \binom{k}{2} - k + 1} = 2^{-\frac{k}{2}+1} \leq 1/2.$$

Since $p_R < 1/2$, and by symmetry $p_B < 1/2$ for the probability of some k vertices with all edges between them colored blue, we conclude that $p_R + p_B < 1$ for $n < 2^{\frac{k}{2}}$, so there *must* be a coloring with no red or blue homogeneous subset of size k. □

The text of this proof has been reproduced with only minor modifications from Aigner and Ziegler's *Proofs from the Book* (2001). (For sharper bounds and more information see Graham *et al.*, 1994.) To make sense of the proof, remember that $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ is the number of ways of choosing a subset of k elements from a set of n objects. The details of the argument are not so important; I am specifically interested in the chain of inequalities. You may

wish to pause here to reflect on what it would take to verify these inferences axiomatically.

Before discussing that issue, let us consider a second example. The following fact comes up in a discussion of the Γ function in Whittaker and Watson (1996).

Lemma 1. *For every complex number z, the series $\sum_{n=1}^{\infty} |\log(1 + \frac{z}{n}) - \frac{z}{n}|$ converges.*[3]

Proof. It suffices to show that for some N, the sum $\sum_{n=N+1}^{\infty} |\log(1 + \frac{z}{n}) - \frac{z}{n}|$ is bounded by a convergent series. But when N is an integer such that $|z| \leq \frac{1}{2}N$, we have, if $n > N$,

$$\begin{aligned}
\left|\log\left(1 + \frac{z}{n}\right) - \frac{z}{n}\right| &= \left|-\frac{1}{2}\frac{z^2}{n^2} + \frac{1}{3}\frac{z^3}{n^3} - \ldots\right| \\
&\leq \frac{|z|^2}{n^2}\left\{1 + |\frac{z}{n}| + |\frac{z^2}{n^2}| + \ldots\right\} \\
&\leq \frac{1}{4}\frac{N^2}{n^2}\left\{1 + \frac{1}{2} + \frac{1}{2^2} + \ldots\right\} \\
&\leq \frac{1}{2}\frac{N^2}{n^2}.
\end{aligned}$$

Since the series $\sum_{n=N+1}^{\infty}\{N^2/(2n^2)\}$ converges, we have the desired conclusion. □

Once again, the text is only slightly modified from that of Whittaker and Watson (1996), and the chain of inequalities is exactly the same. The equality involves a Taylor series expansion of the logarithm. The first inequality follows from properties of the absolute value, such as $|xy| = |x| \cdot |y|$ and $|x + y| \leq |x| + |y|$, while the second inequality makes use of the assumption $|z| < \frac{1}{2}N$ and the fact that $n > N$.

Inequalities like these arise in all branches of mathematics, and each discipline has its own bag of tricks for bounding the expressions that arise (see, for example Hardy *et al.*, 1988; Steele, 2004). I have chosen the two examples above because they invoke only basic arithmetic reasoning, against some general background knowledge. At present, providing enough detail for a computer to verify even straightforward inferences like these is a burdensome chore (see Avigad *et al.*, 2007).

In sufficiently restricted contexts, in fact, decision procedures are often available. For example, in 1929 Presburger showed that the theory of the

[3] Here we are taking the principal value of $\log(1 + \frac{z}{n})$.

integers in a language with 0, 1, +, and < is decidable, and in the early 1930s, Tarski showed that the theory of the real numbers in a language with 0, 1, +, ×, and < is decidable (the result was not published, though, until 1948, and was soon reprinted as Tarski, 1951). These decision procedures work for the full first-order language, not just the quantifier-free fragments, and have been implemented in a number of systems. But decision procedures do not tell the whole story. Thanks to Gödel, we know that decidability is the exception rather than the rule; for example, the theory of the integers becomes undecidable in the presence of multiplication, and the theory of the reals becomes undecidable in the presence of, say, the sine function. Even in restricted settings where problems are decidable, full decision procedures tend to be inefficient or even infeasible. In fact, for the types of inferences that arise in practice, short justifications are often possible where general decision procedures follow a more circuitous route.

There has therefore been a fair amount of interest in 'heuristic procedures', which search for proofs by applying a battery of natural inferences in a systematic way (for some examples in the case of real arithmetic, see Beeson, 1998; Hunt *et al.*, 2003; Tiwari, 2003). One strategy is simply to work backwards from the goal inequality. For example, one can prove an inequality of the form $1/(1+s) < 1/(1+t)$ by proving $1+s > 1+t > 0$, and one can do that, in turn, by proving $s > t > -1$. This amounts to using basic rules like

$$x > y, y > 0 \Rightarrow 1/x < 1/y$$

to work backwards from the goal, a technique known as 'backchaining'.

Backchaining has its problems, however. For example, one can also prove $1/x < 1/y$ by proving that x and y are negative and $x > y$, or that x is negative and y is positive. Trying all the possibilities can result in dreaded case splits. Search procedures can be nondeterministic in more dramatic ways; for example, one can prove $q + r < s + t + u$, say, by proving $q < t$ and $r \leq s + u$, or by proving $q < s + t + u$ and $r \leq 0$. Similarly, one can prove $s + t < 5 + u$ by proving, say, $s < 3$ and $t \leq 2 + u$.

So, simply working backwards is insufficient on its own. In the proof of Theorem 1, we used bounds on $\binom{N}{k}$ and N to bound composite expressions using these terms, and in the proof of Lemma 1, we used the bounds $|z| \leq N/2$ and $N < n$ to bound the terms $|z^i/n^i|$. Working forwards in this way to amass a store of potentially useful inequalities is also a good idea, especially when the facts may be used more than once. For example, when all the terms in sight are positive, noting this once and for all can cut down the possibilities for backwards search dramatically. But, of course, deriving inequalities blindly

won't help in general. The choice of facts we derive must somehow be driven by the context and the goal.

A final observation is that often inferences involving inequalities are verified by combining methods from more restricted domains in which it is clear how one should proceed. For example, the additive theory of the reals is well understood, and the multiplicative fragment is not much more complicated; more complex inferences are often obtained simply by combining these modes of reasoning. There is a body of methods stemming from a seminal paper by Nelson and Oppen (1979) that work by combining such 'local' decision procedures in a principled way.

These three strategies—backward-driven search, forward-driven search, and combining methods for handling more tractable problems—are fundamental to automated reasoning. When it comes to real-valued inequalities, an analysis of the problem in these terms can be found in Avigad and Friedman (2006). But what we really need is a general theory of how these strategies can be combined successfully to capture common mathematical inferences. To some extent, it will be impossible to avoid the objection that the methods we devise are merely '*ad hoc* and heuristic'; in real life, we use a considerable amount of mucking around to get by. But in so far as structural features of the mathematics that we care about make it possible to proceed in principled and effective ways, we should identify these structural features. In particular, they are an important part of how we understand proofs involving inequalities.

12.7 Understanding algebraic reasoning

The second type of inference I would like to consider is that which makes use of algebraic concepts. The use of such concepts is a hallmark of modern mathematics, and the following pattern of development is typical. Initially, systems of objects arising in various mathematical domains of interest are seen to share a common structure. Abstracting from the particular examples, one then focuses on this common structure, and determines the properties that hold of all systems that instantiate it. This infrastructure is then applied in new situations. First, one 'recognizes' an algebraic structure in a domain of interest, and then one instantiates facts, procedures, and methods that have been developed in the general setting to the case at hand. Thus algebraic reasoning involves complementary processes of abstraction and instantiation. We will consider an early and important example of such reasoning, namely

the use of the concept of a group to establish a proposition in elementary number theory.

A group consists of a set, G, an associative binary operation, $\cdot$, on G, and an identity element, 1, satisfying $1 \cdot a = a \cdot 1 = a$ for every a in G. In a group, every element a is assumed to have an *inverse*, a^{-1}, satisfying $a \cdot a^{-1} = a^{-1} \cdot a = 1$. It is common to use the letter G to refer to both the group and the underlying set of elements, even though this notation is ambiguous. We will also write ab instead of $a \cdot b$. Exponentiation a^n is defined to be the the product $a \cdot a \cdots a$ of n copies of a, and if S is any finite set, $|S|$ denotes the number of elements in S.

Proposition 1. *Let G be a finite group and let a be any element of G. Then $a^{|G|} = 1$.*

This proposition is an easy consequence of a theorem known as *Lagrange's theorem*, but there is an even shorter and more direct proof when the group is *abelian*, that is, when the operation satisfies $ab = ba$ for every a and b.

Proof of Proposition 1 when G is abelian. Given G and a, consider the operation $f_a(b) = ab$, that is, multiplication by a on the left. This operation is injective, since $ab = ab'$ implies $b = a^{-1}ab = a^{-1}ab' = b'$, and surjective, since for any b we can write $b = a(a^{-1}b)$. So f_a is a bijection from G to G.

Suppose the elements of G are $b_1, \ldots, b_n$. By the preceding paragraph, multiplication by a simply permutes these elements, so $ab_1, \ldots, ab_n$ also enumerates the elements of G. The product of these elements is therefore equal to both

$$(ab_1)(ab_2)\cdots(ab_n) = a^n(b_1b_2\cdots b_n),$$

and $b_1b_2\cdots b_n$. This implies $a^n = 1$, as required. □

To apply Proposition 1 to number theory, we need only find a suitable group. Write $a \equiv b(m)$, and say that a and b are *congruent modulo m*, if a and b leave the same remainder on division by m, that is, if m divides $a - b$. The relationship of being congruent modulo m is symmetric, reflexive, and transitive. It also respects addition and multiplication: if $a \equiv a'(m)$ and $b \equiv b'(m)$, then $(a + a') \equiv (b + b')(m)$ and $aa' \equiv bb'(m)$. Any integer, a, is congruent to a number between 0 and $m - 1$ modulo m, namely, its remainder, or 'residue', modulo m. Thus congruence modulo m divides all the integers into m 'equivalence classes', each represented by the corresponding residue. We can think of addition and multiplication as operations on these equivalence classes, or, alternatively, as operations on residues, where after each operation we take the remainder modulo m. (For example, under clock arithmetic, we only care about the values $0, \ldots, 11$, and we are adding modulo 12 when we say that '5 hours after 9 o'clock, it will be 2 o'clock'.)

Two integers a and m are said to be *relatively prime* if they have no factors in common other than ± 1. If two numbers a and b are relatively prime to m, then so is their product, ab. Also, if a and m are relatively prime, the Euclidean algorithm tells us that there are integers x and y such that $ax + my = 1$. In the language of congruences, this says that if a and m are relatively prime, there is an x such that $ax \equiv 1(m)$.

What this means is that the residues that are relatively prime to m form a group under the operation of multiplication modulo m. For $m \geq 2$, *Euler's φ function* is defined by setting $\varphi(m)$ equal to the number of these equivalence classes, that is, the number of integers between 0 and $m - 1$ that are relatively prime to m. The following theorem, known as *Euler's theorem*, is then an immediate consequence of Proposition 1.

Theorem 2. *For any $m \geq 2$ and any number, a, relatively prime to m, $a^{\varphi(m)} \equiv 1(m)$.*

In particular, if p is prime, then $1, \ldots, p - 1$ are all relatively prime to p, and $\varphi(p) = p - 1$. This special case of Euler's theorem is known as Fermat's little theorem:

Corollary 1. *If p is prime and does not divide a, then $a^{p-1} \equiv 1(p)$.*

Neither Euler's theorem nor Fermat's little theorem has anything to do with groups *per se*. Nonetheless, these theorems are usually understood as reflections of the underlying algebraic structure on residues. In fact, Euler published the first proof of Fermat's little theorem in 1736, and a proof of the more general Theorem 2 in 1760, well before the first axiomatic definition of a group was given by Cayley in 1854. Chapter III of Dickson's three-volume *History of the Theory of Numbers* (1966) enumerates dozens of proofs of Fermat's and Euler's theorems; see also Wussing (1984) for a discussion of early algebraic proofs.

A comparison of some of the various proofs can be used to illustrate the advantages of the algebraic method. The most commonly cited advantage is, of course, generality: a single abstract theorem about groups has consequences everywhere a group can be identified, and it is more efficient to carry out a general argument once than to have to repeat it in each specific setting. But this does not tell the whole story; for example, algebraic methods are often deemed useful in their *initial* application, before they are applied in other settings. Sometimes the benefits are terminological and notational: group theory gives us convenient methods of calculating and manipulating expressions where other authors have to resort to cumbersome locutions and manners of expression. But algebraic abstraction also has a way of focusing our efforts by suppressing distracting information that is irrelevant to the solution of a problem. For example, in the last step of the argument above, it suffices to

know that $b_1 b_2 \cdots b_n$ has a multiplicative inverse modulo m; in other words, that there is number c such that $cb_1 b_2 \cdots b_n$ is equal to 1 modulo m. Recognizing that fact eliminates the need to clutter the proof with calculations that produce the particular c. Finally, an important feature of algebraic methodology is that it enables us to discover notions that are likely to be fruitful elsewhere. It provides a uniform way of 'seeing' analogies in otherwise disparate settings. Echoing Wittgenstein, algebraic concepts 'lead us to make investigations; are the expressions of our interest, and direct our interest'.

From a traditional logical perspective, algebraic reasoning is easily explained. Proposition 1 makes a universal assertion about groups, giving it the logical form $\forall G\ (Group(G) \rightarrow \ldots)$. Later, when we have defined a particular object G and shown that it is a group, applying the proposition requires nothing more than the logical rules of universal instantiation and modus ponens.

But somehow, when we read a proof, we are not conscious of this continual specialization. Once we recognize that we are dealing with a group, facts about groups are suddenly ready to hand. We know how to simplify terms, and what properties are potentially relevant. We are suddenly able to think about the objects in terms of subgroups, orbits, and cosets; our group-theoretic understanding enables us to 'see' particular consequences of the abstract theory. The logical story does not have much to say about how this works. Nor does it have much to say about how we are able to reason in the abstract setting, and how this reasoning differs from that of the domain of application.

In contrast, developers of mechanized proof assistants have invested a good deal of effort in understanding how these inferences work. In formal verification, the notion of a collection of facts and procedures that are 'ready to hand' is sometimes called a 'context'. Proof assistants provide various methods of reasoning within such a context; Isabelle implements the notion of a 'locale' (Ballarin, 2006), while Coq supports a system of 'modules' (Bertot and Castéran, 2004). Here is a localic proof of Proposition 1, which is understood by Isabelle:

```
lemma (in comm-group) power-order-eq-one:
  assumes finite (carrier G) and a:carrier G
  shows a (^) card(carrier G) = one G
proof–
  have (⨂ x:carrier G. x) = (⨂ x:carrier G. a ⊗ x)
    by (subst (2) finprod-reindex [symmetric],
        auto simp add: Pi-def inj-on-const-mult surj-const-mult prems)
  also have ... =(⨂ x:carrier G. a) ⊗ (⨂ x: carrier G. x)
    by (auto simp add: finprod-multf Pi-def prems)
  also have (⨂ x:carrier G. a) = a (^) card(carrier G)
```

```
    by (auto simp add: finprod-const prems)
    finally show ?thesis
      by (auto simp add: prems)
qed
```

The notation (**in** *comm-group*) indicates that this is a lemma in the commutative group locale. The notation (^) denotes exponentiation, the expression *carrier G* denotes the set underlying the group, and the notation $\bigotimes x{:}carrier\ G$ denotes the product over the elements of that set. With this lemma in place, here is a proof of Euler's theorem:

```
Theorem euler-theorem:
  assumes m > 1 and zgcd(a, m) = 1
  shows [a^phi m = 1] (mod m)
proof–
  interpret comm-group [residue-mult-group m]
    by (rule residue-group-comm-group)
  have (a mod m) {^}m (phi m) = {1}m
    by (auto simp add: phi-def power-order-eq-one prems)
  also have (a mod m) {^}m (phi m) = (a^phi m) mod m
    by (rule res-pow-cong)
  finally show ?thesis
    by (subst zcong-zmod-eq, insert prems auto simp add: res-one2-eq)
qed
```

The proof begins by noting that the residues modulo *m* from an abelian group. The notation {^}*m* then denotes the operation of exponentiation modulo *m* on residues, and {1}*m* denotes the residue 1, viewed as the identity in the group. The proof then simply appeals to the preceding lemma, noting that exponentiating in the group is equivalent to applying ordinary integer exponentiation and then taking the remainder modulo *m*.

The relative simplicity of these proof scripts belies the work that has gone into making the locales work effectively. The system has to provide mechanisms for identifying certain theorems as theorems about groups; for specializing facts about groups, automatically, to the specific group at hand; and for keeping track of the calculational rules and basic facts that enable the automated tools to recognize straightforward inferences and calculations in both the general and particular settings. Furthermore, one needs mechanisms to manage locales and combine them effectively. For example, every abelian group is a group; the multiplicative units of any ring form a group; the collection of subgroups of a group has a lattice structure; the real numbers as an ordered field provide instances of an additive group, a multiplicative group, a field, a linear ordering, and so on. Thus complex bookkeeping is needed to keep track of the facts and

procedures that are immediately available to us, in ordinary mathematics, when we recognize one algebraic structure as present in another.

The work that has been done is a start, but proof assistants are still a long way from being able to support algebraic reasoning as it is carried out in introductory textbooks. Thus, a good deal more effort is needed to determine what lies behind even the most straightforward algebraic inferences, and how we understand simple algebraic proofs.

12.8 Understanding Euclidean geometry

Our third example will take us from modern mathematics to the mathematics of the ancients. Over the centuries, the style of diagram-based argumentation of Euclid's *Elements* was held to be the paradigm of rigor, and presentations much like Euclid's are still used today to introduce students to the notion of proof. In the 19th century, however, Euclidean reasoning fell from grace. Diagrammatic reasoning came to be viewed as imperfect, lacking sufficient mathematical rigor, and relying on a faculty of intuition that has no place in mathematics. The role of diagrams in licensing inference was gradually reduced, and then, finally, eliminated from most mathematical texts. Axiomatizations by Pasch (1882) and Hilbert (1899), for example, were viewed as 'corrections' or improvements to the flawed methods of Euclid, filling in gaps that Euclid overlooked.

As Manders points out in his contributions to this volume, this view of Euclidean geometry belies the fact that Euclidean reasoning was a remarkably stable and robust practice for more than two thousand years. Sustained reflection on the *Elements* shows that there are implicit rules in play, and norms governing the use of diagrams that are just as determinate as the norms governing modern proof. Given the importance of diagrammatic reasoning not just to the history of mathematics but to geometric thought today, it is important to understand how such reasoning works.

In order to clarify the types of inferences that are licensed by a diagram, Manders distinguishes between 'exact' conditions, such as the claim that two segments are congruent, and 'coexact' conditions, such as the claim that two lines intersect, or that a point lies in a given region. Coexact conditions depend only on general topological features of the diagram, and are stable under perturbations of a diagram, whereas exact conditions are not. Manders observes that only coexact claims are ever inferred from a diagram in a Euclidean proof; text assertions are required to support an inference

that results in an exact claim. Thus, diagrams serve as useful representations of certain types of coexact data, and these representations are used in regimented ways.

Observations like these have led Nathaniel Miller (2001) and John Mumma (2006) to develop formal systems for diagram-based reasoning that are more faithful reflections of Euclidean reasoning than Hilbert's and Pasch's axiomatizations. In both systems, diagrams and formulas together bear the burden of representing assertions, and precise rules govern the construction of diagrams and the inferences that can be drawn. In Miller's system, a diagram is a more abstract object, namely, a graph representing the topological data that is relevant to a Euclidean argument. In Mumma's system, a diagram consists of labeled geometric elements with coordinates on a grid of points.

Both systems face the problem of explaining how a diagrammatic proof can ensure generality. Suppose that to prove a general theorem about triangles, I begin by drawing a particular triangle, *ABC*. Aside from being imperfect, this triangle will have specific properties that play no role in the proof: it may, for example, be acute or obtuse, or (more or less) a right triangle. The problem to be addressed is how reasoning about this *particular* triangle can warrant conclusions about *all* triangles, even ones with different coexact properties.

Let us consider, for example, the second proposition in Book I of the *Elements*, which shows that if *BC* is any segment and *A* is any point, it is possible to construct a segment congruent to *BC* with an endpoint at *A*. Aside from the definitions, common notions, and postulates, the proof relies only

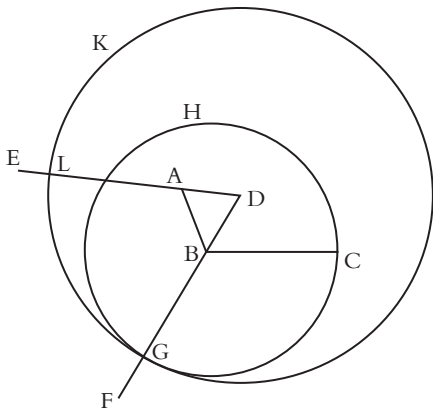

Fig. 12.1.

on the first proposition of Book I, which shows that it is possible to construct an equilateral triangle on a given side. The text below is taken from Heath's translation of Euclid. The example, and the gist of the subsequent discussion, are taken from Mumma (2006).

Proposition 2. *To place at a given point (as an extremity) a straight line equal to a given straight line.*

Proof. Let *A* be the given point, and *BC* the given straight line (Fig. 12.1). Thus it is required to place at the point *A* (as an extremity) a straight line equal to the given straight line *BC*.

From the point *A* to the point *B*, let the straight line *AB* be joined; and on it let the equilateral triangle *DAB* be constructed. Let the straight lines *AE*, *BF* be produced in a straight line with *DA*, *DB*; with centre *B* and distance *BC* let the circle *CGH* be described; and again, with centre *D* and distance *DG* let the circle *GKL* be described.

Then, since the point *B* is the centre of the circle *CGH*, *BC* is equal to *BG*. Again, since the point *D* is the centre of the circle *GKL*, *DL* is equal to *DG*. And in these *DA* is equal to *DB*; therefore the remainder *AL* is equal to the remainder *BG*. But *BC* was also proved equal to *BG*; therefore each of the straight lines *AL*, *BC* is equal to *BG*. And things that are equal to the same thing are also equal to one another; therefore *AL* is also equal to *BC*.

Therefore at the given point *A* the straight line *AL* is placed equal to the given straight line *BC*. (Being) what it was required to do. □

The conclusion of the proof is valid. But how does it work? As Mumma points out, the position of the point *A* with respect to *BC* is indeterminate, and different ways of placing that point would result in diagrams with different coexact properties. For example, if *A* were chosen so that *AB* is longer than *BC*, then the point *A* would lie *outside* the circle *CGH*. The diagram is used to license the conclusion that circle *CGH* intersects the line *DF*, and, in fact, this does hold in general. It is similarly used to license the conclusion that *GKL* will intersect the line *GE*. Moreover, to warrant the conclusion that *AL* is equal to *BG*, one needs to know that the point *A* lies between *D* and *L*, and that the point *B* lies between *D* and *G*. The question is, how can a particular diagram warrant these general conclusions?

In Miller's system, whenever a representational choice can affect the diagram in a way that may influence the inferences that can be drawn, one simply carries out a case split. Thus, in Miller's system, one needs to check the result against each topological configuration that can arise. But, as Mumma points out, this can yield a combinatorial explosion of possibilities, and such case splits are notably absent from the *Elements*. Euclid is, somehow, able to draw the

right general consequences from a specific representation, without having to distinguish irrelevant cases.

Mumma (2006) provides a more subtle explanation of how this works, by providing rules by which certain features of a diagram can be justified as general consequences of the construction, irrespective of the choices made. Note that Mumma is not claiming to describe the procedure by which Euclid and his peers justified their inferences. All we have, in the *Elements*, is a body of text, with no record of the cognitive mechanisms that produced them. But these texts provide us with a collection of inferences (which, today, we recognize as being sound for a certain semantics). Understanding Euclidean proof means, in part, being able to determine which inferences are allowed. Mumma has explained what this understanding amounts to, both by characterizing the allowable inferences, and providing a mechanism by which they can be validated.

What does all this have to do with automated reasoning and formal verification? Geometric theorem proving has, for the most part, followed modern mathematics in utilizing algebraic treatments of the subject. In other words, most modern theorem provers express geometric assertions in algebraic terms and use algebraic methods to justify them. In fact, via the usual embedding of Euclidean geometry in the Cartesian plane, a decision procedures for real closed fields is a decision procedure for geometry as well. Moreover, specialized algebraic procedures (Chou *et al.*, 1994; Wu, 1994) have proved to be remarkably effective in verifying ordinary geometric theorems.

But analyses like Miller's and Mumma's may point the way to developing computational methods for verifying Euclidean inferences on their own terms. Not only would this be interesting in its own right, but it would provide important insights into the inner workings of modern methods as well. For example, in the 16th century, the geometric method of 'analysis and synthesis' foreshadowed Descartes' algebraic treatment of geometry (Bos, 2000). In fact, this method of naming an object sought and then analyzing its properties can be seen as a geometric version of the method of Skolemization and unification described in Section 12.5. For another example, note that we often prove theorems in analysis—say, when dealing with Hilbert spaces, which are infinite-dimensional generalizations of Euclidean geometry—by drawing a diagram. So, even in light of the attenuated role of diagrams in modern mathematical proofs, understanding the nature of geometric inference is still an important part of understanding how such modern proofs work.

12.9 Understanding numeric substructures

As a fourth and final example, I would like to consider a class of inferences that involve reasoning about the domains of natural numbers, integers, rational numbers, real numbers, and complex numbers, with respect to one another. From a foundational perspective, it is common to take the natural numbers as basic, or to construct them from even more basic objects, like sets. Integers can then be viewed as equivalence classes of pairs of natural numbers; rational numbers can be viewed as equivalence classes of pairs of integers; real numbers can be viewed, say, as sequences of rational numbers; and complex numbers can be viewed as pairs of reals. As McLarty points out in the first of his essays in this collection, however, this puts us in a funny situation. After all, we think of the natural numbers as a *subset* of the integers, the integers as a subset of the rationals, and so on. On the foundational story, this is not quite true; really, each domain is *embedded* in the larger ones. Of course, once we have the complex numbers, we may choose to baptize the complex image of our original copy of the natural numbers as our new working version. Even if we do this, however, we still have to keep track of where individual elements 'live'. For example, we can apply the principle of induction to the complex copy of the natural numbers, but not to the complex numbers as a whole; and we can divide one natural number by another, but the result may not be a natural number. If we think of the natural numbers as embedded in the complex numbers, we have to use an embedding function to make our statements literally true; if we think of them as being a subset of the complex numbers, all of our statements have to be carefully qualified so that we have specified what types of objects we are dealing with.

The funny thing is that in ordinary mathematical texts, all this happens under the surface. Textbooks almost never tell us which of these two foundational options are being followed, because the choice has no effect on the subsequent arguments. And when we read a theorem that combines the different domains, we somehow manage to interpret the statements in such a way that everything makes sense. For example, you probably did not think twice about the fact that Lemma 1 above involved three sorts of number domain. The fact that the series is indexed by n means that we have to think of n as a natural number (or a positive integer). Similarly, there is an implicit indexing of terms by natural numbers in the use of '…' in the expression for the logarithm. The proof, in fact, made use of properties of such sums that are typically proved by induction. In the statement of the theorem, the variable z is explicitly designated to denote a complex number, so when we divide z by n in the

expression z/n, we are thinking of n as a complex number as well. But the absolute value function converts a complex value to a real value, so expressions like $|z/n|$ denote real numbers, and many of the products and sums in that proof denote real multiplication and addition. In the proof of the lemma, it is crucial that we keep track of this last fact: the $\leq$ ordering only makes sense on the real numbers, so to invoke properties of the ordering we need to know that the relevant expressions are real-valued.

The elaborate structure of implicit inferences comes to the fore when we try to formalize such reasoning. With a formal verification system, these inferences need to be spelled out in detail, and the result can be painstaking and tedious (see Avigad *et al.*, 2007). In Isabelle, for example, one has to use a function *real*(n) to cast a natural number, n, as a real. Coq has mechanisms that apply such 'coercions' automatically, but, in both cases, the appropriate simplifications and verifications are rarely carried out as automatically as one would like.

The fact that methods of reasoning that we are barely conscious of when we read a mathematical proof requires so much effort to formalize is one of the scandals of formal verification, and a clear sign that more thought is needed as to how we understand such inferences. I suspect that the structural perspective described in McLarty's article, combined with the locale mechanisms described in Section 12.7, holds the germ of a solution. When we prove theorems about the natural numbers, that is the structure of interest; but when we identify the natural numbers as a substructure of the reals, we are working in an expanded locale, where both structures are present and interact. To start with, everything we know in the context of the natural numbers and the real numbers individually is imported to the combined locale, and is therefore available to us. But there are also new facts and procedures that govern the combination of the two domains. Figuring out how to model such an understanding so that proof assistants can verify the proof of Lemma 1, as it stands, will go a long way in explaining how we understand proofs that make use of mixed domains.

12.10 Conclusions

I have described four types of inference that are found in ordinary mathematical proofs, and considered some of the logical and computational methods that have been developed to verify them. I have argued that these efforts are not just pragmatic solutions to problems of engineering; rather, they address core issues in the epistemology of mathematics, and should be supported by broader philosophical reflection.

Mathematics guides our thought in deep and powerful ways, and deserves a philosophy that recognizes that fact. When we focus on particular features of mathematical practice, metaphysical concerns often seem petty and irrelevant, and we find, instead, a rich assortment of issues that have concrete bearing upon what we do and say about the subject. Our task is to develop a conceptual framework in which we can fruitfully begin to address these issues, and to narrow our focus to the point where discernible progress can be made. I hope the present chapter serves as encouragement.

Added in Proof. After the article had been sent to the publisher, I came across work by Jody Azzouni which bears directly on many of the issues raised here. Although, I cannot explore the points of agreement and disagreement now, the reader may wish to compare my views to those of Azzouni (2005) and Azzoumi (2006).

Bibliography

Aigner, Martin and Ziegler, Günter M. (2001), *Proofs from The Book*, 2nd edn (Berlin: Springer-Verlag).

Avigad, Jeremy (2006), 'Mathematical method and proof', *Synthese*, 153, 105–159.

Avigad, Jeremy, Donnelly, Kevin, Gray, David, and Raff, Paul (2007) 'A formally verified proof of the prime number theorem', *ACM Transactions on Computational Logic*, 9(1:2).

Avigad, Jeremy and Friedman, Harvey (2006), 'Combining decision procedures for the reals', *Logical Methods in Computer Science*, 2(4:4).

Azzoumi, Jody (2005), 'Is there a sense in which mathematics can have foundations?' In G. Sica (ed.), Essays in the Foundations of Mathematics and Logic, 9–47 (Monza: Polimetrica).

____(2006), Tracking Reason: Proof, Consequence, and Truth (Oxford: Oxford University Press).

Ballarin, Clemens (2006), 'Interpretation of locales in Isabelle: theories and proof contexts', in J. M. Borwein and W. M. Farmer (eds.), *Mathematical Knowledge Management: Proceedings of the Fifth International Conference, MKM 2006*, 31–43 (Berlin: Springer-Verlag).

Beeson, Michael (1998), 'Design principles of Mathpert: software to support education in algebra and calculus', in N. Kajler (ed.), *Computer—Human Interaction in Symbolic Computation*, 89–115 (Berlin: Springer-Verlag).

Bertot, Yves and Castéran, Pierre (2004), *Interactive Theorem Proving and Program Development: Coq'art: the Calculus of Inductive Constructions* (Berlin: Springer-Verlag).

Bos, Henk J. M. (2000), *Redefining Geometrical Exactness: Descartes' Transformation of the Early Modern Concept of Construction* (New York: Springer-Verlag).

CARR, David (1979), 'The logic of knowing how and ability', *Mind*, 88, 394–409.

CHOU, C. C., GAO, X. S., and ZHANG, J. Z. (1994), *Machine Proofs in Geometry* (Singapore: World Scientific).

DICKSON, Leonard Eugene (1966), *History of the Theory of Numbers*, vols. I–III (New York: Chelsea Publishing Co.).

EUCLID, *The Thirteen Books of Euclid's Elements Translated from the Text of Heiberg*, trans. with introduction and commentary by Thomas L. Heath, 2nd edn, vols. I–III (New York: Dover Publications).

GOLDFARB, Warren (1985), 'Kripke on Wittgenstein on rules', *Journal of Philosophy*, 82, 471–488.

GRAHAM, Ronald L., KNUTH, Donald E., and PATASHNIK, Oren (1994), *Concrete Mathematics: a Foundation for Computer Science*, 2nd edn (Reading, MA: Addison-Wesley).

HARDY, G. H., LITTLEWOOD, J. E., and PÓLYA, G. (1988), *Inequalities* (Cambridge: Cambridge University Press), repr. of 1952 edn.

HILBERT, David (1899), 'Grundlagen der Geometrie', in *Festschrift zur Feier der Enthüllung des Gauss-Weber Denkmals in Göttingen* (Leipzig: Teubner). Trans. by Leo Unger (1971) as *Foundations of geometry* (La Salle: Open Court).

HUNT, Warren A., KRUG, Robert Bellarmine, and MOORE, J. (2003), 'Linear and nonlinear arithmetic in ACL2', in Daniel Geist and Enrico Tronci (eds.), *Correct Hardware Design and Verification Methods, Proceedings of CHARME 2003*, 319–333 (Berlin: Springer-Verlag).

KRIPKE, Saul (1982), *Wittgenstein on Rules and Private Language* (Cambridge, Mass.: Harvard University Press).

MILLER, Nathaniel (2001), *A Diagrammatic Formal System for Euclidean Geometry*, Ph.D. thesis, Cornell University.

MUMMA, John (2006), *Intuition Formalized: Ancient and Modern Methods of Proof in Elementary Geometry*, Ph.D. thesis, Carnegie Mellon University.

NELSON, Greg and OPPEN, Derek C. (1979), 'Simplification by cooperating decision procedures', *ACM Transactions of Programming Languages and Systems*, 1, 245–257.

PASCH, Moritz (1882), *Vorlesungen über neueren Geometrie* (Leipzig: Teubner).

POINCARÉ, Henri (1908), *Science et Méthode* (Paris: Flammarion). Trans. by Francis Maitland in 1914 as *Science and Method* (London: T. Nelson and Sons, 1914), republished in 2003 (Mineola, NY: Dover Publications).

RYLE, Gilbert (1949), *The Concept of Mind* (Chicago, IL: University of Chicago Press).

STEELE, J. Michael (2004), *The Cauchy–Schwarz Master Class: an Introduction to the Art of Mathematical Inequalities* (Washington, DC: Mathematical Association of America).

TAIT, William W. (1986), 'Wittgenstein and the "skeptical paradoxes"', *Journal of Philosophy*, 83, 475–488.

TARSKI, Alfred (1951), *A Decision Procedure for Elementary Algebra and Geometry*, 2nd edn (University of California Press). Reprinted in B. F. Caviness and J. R. Johnson (eds.) *Quantifier Elimination and Cylindrical Algebraic Decomposition*, 24–84 (Vienna: Springer-Verlag, 1988).

TIWARI, A. (2003), 'Abstraction based theorem proving: an example from the theory of reals', in C. Tinelli and S. Ranise (eds.), *Proceedings of the CADE-19 Workshop on Pragmatics of Decision Procedures in Automated Deduction, PDPAR 2003*, 40–52 (Berlin: Springer-Verlag).

WHITTAKER, E. T. and WATSON, G. N. (1996), *A Course of Modern Analysis*, (Cambridge: Cambridge University Press), reprinting of the 4th edn, 1927.

WIEDIJK, Freek (2006), *The Seventeen Provers of the World* (Berlin: Springer-Verlag).

WITTGENSTEIN, Ludwig (1953), *Philosophical Investigations* (New York: Macmillan), ed. and trans. G. E. M. Anscombe, 3rd edn (2001) (Oxford: Blackwell Publishers).

——(1956), *Remarks on the Foundations of Mathematics*, ed. G. H. von Wright, R. Rhees and G. E. M. Anscombe, trans. G. E. M. Anscombe (Oxford: Blackwell). Revised edition (1978) (Cambridge, MA: MIT Press).

——(1969), *On Certainty*, ed. G. E. M. Anscombe and G. H. von Wright, trans. G. E. M. Anscombe and D. Paul (Oxford: Blackwell).

WU, Wen Tsün (1994), *Mechanical Theorem Proving in Geometries*, trans. from the 1984 Chinese original by Xiao Fan Jin and Dong Ming Wang (Vienna: Springer-Verlag).

WUSSING, Hans (1984), *The Genesis of the Abstract Group Concept: A Contribution to the History of the Origin of Abstract Group Theory*, trans. Abe Shenitzer and Hardy Grant (Cambridge, MA: MIT Press).