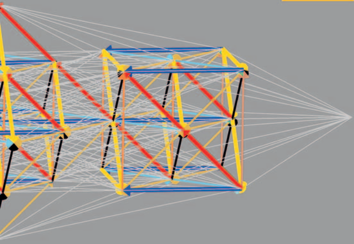
OXFORD



REALIZING REASON

A Narrative of Truth & Knowing

DANIELLE MACBETH

Contents

2

Ancient Greek Diagrammatic Practice

Mathematics, the oldest and most venerable of all the sciences, begins first and foremost with the ancient Greeks. Although many mathematical discoveries were made also in ancient India and ancient China, it was the Greeks who realized mathematics as a systematic science within which to demonstrate an extraordinary range of properties and relations of mathematical entities from the very simple to the remarkably complex.[1] Our aim is to understand the practice of this science, in particular, the use of diagrams in ancient Greek mathematical practice. That is, we want not merely to understand the mathematical results but also, and primarily, the means by which those results are established in Greek diagrammatic practice, how the practice works as mathematics. We will be especially concerned, first, with the sort of generality that is involved in ancient Greek diagrammatic reasoning, and also with the very idea of diagrammatic reasoning, what it means to reason *in* a diagram as opposed to reasoning on the basis of a diagram.

It is often remarked that mathematical knowledge is cumulative in a way that empirical knowledge is not, that there are no revolutions in mathematics comparable to those that have occurred in the empirical sciences. But although revolutions perhaps do not occur in mathematics, at least as they do in the empirical sciences, nevertheless, as Crowe (1975, 19) remarks, "revolutions may occur in mathematical nomenclature, symbolism, metamathematics (e.g. the metaphysics of mathematics), methodology (e.g. standards of rigor), and perhaps even in the historiography of mathematics."[2] Although the mathematics itself may not change, nevertheless, our understanding of the mathematics can and does change, sometimes very radically.[3] This idea that the mathematics itself does not change is furthermore often formulated

[1] See Bashmakova and Smirnova (2000).

[2] Crowe (1975) argues that there are no revolutions in mathematics comparable to those in the natural sciences. Later, in Crowe (1988), he reverses his view. Crowe (1975) has generated a great deal of further work. For a recent overview see François and Van Bendegem (2010).

[3] Putnam (1972, 5) suggests something similar for the case of logic. As he writes, for instance, of the law of identity or of inference in *barbara*, "even where a principle may seem to have undergone no change in the course of the centuries . . . the *interpretation* of the 'unchanging' truth has, in fact, changed considerably."

in terms of the idea that the earlier conceptual framework is always translatable into the later one.[4] And at least in some cases, the translations seem not merely to preserve the mathematics but to *reveal* the mathematics in a way that the original formulations could not.[5] It was, for example, widely argued in the nineteenth century that ancient Greek mathematics is algebra in geometric dress, that is, that it is really algebra though the method is geometrical.[6] And there is a sense in which this is true: certain parts of Greek mathematics achieve a kind of closure when translated into the language of algebra. But although it may be correct to describe the relevant *mathematics* as algebra, to think of Greek mathematical *practice* as "algebra in geometrical dress" neglects the fact that, although its fruits can be translated into the language of algebra, Greek mathematical practice is not algebraic. Although the *mathematics* (in relevant instances) may be algebra, the *language* and *practice* of ancient Greek mathematics is geometry.[7]

If in the nineteenth century Euclid was read through the lens of the sort of algebra that was first made possible by Descartes, today it is more common to read Euclid through the lens of twentieth-century quantificational logic and current mathematical practice (or at least what many philosophers think of as current mathematical

[4] Friedman claims, for example, that "revolutionary transitions within pure mathematics . . . have the striking property of continuously (and, as it were, monotonically) preserving what I want to call *retrospective* communicative rationality: practitioners at a later stage are always in a position to understand and rationally to justify—at least in their own terms—all the results of earlier stages" (Friedman 2001, 96); "in pure mathematics . . . there is a clear sense in which an earlier conceptual framework (such as classical Euclidean geometry) is always translatable into a later one (such as the Riemannian theory of manifolds)" (Friedman 2001, 99). Stein (1988, 238) similarly remarks, "a mathematician today, reading the works of Archimedes, or Eudoxus's theory of ratios in Book V of Euclid, will feel that he is reading a contemporary."

[5] This is because, as Wilson (1995, 111) notes, "the 'proper' definition of a mathematical term should not rest upon the brute fact that earlier mathematicians had decided that it should be explicated in such-and-such manner, but upon whether the definition suits the realm in which the relevant objects optimally 'grow and thrive'." And this is true even in the case of mathematical constants. For example, we take the number π, that is, the ratio of the circumference of a circle to its diameter, to be a basic and important mathematical constant, but Palais (2001) persuasively argues that it makes more mathematical sense to use the ratio of the circumference to the *radius* as the basic and important constant. It would have been natural for the ancients (whether Greek or Chinese or Indian) to fix on the diameter given that for them a circle is an object, a kind of two-dimensional geometrical figure. But once one has achieved the modern notion of a circle as given by the familiar equation $x^2 + y^2 = r^2$, it is clear that what matters is not the diameter, but r, that is, the radius. See also Hartl (2010), who suggests that the constant ought to be given the symbol *tau*, τ.

[6] We find such arguments in, for example, Georg Nesselman's *Die Algebra der Griechen* (1847), Paul Tannery's "*De la solution géométrique des problèmes du second degree avant Euclide*," *Mémoires scientifiques* (1882), and Hieronymous Zeuthen's *Die Lehre von den Kegelschnitten im Altertum* (1886). The interpretation was taken up by Heath in his English translation of the *Elements* first published in 1908, and was further applied also to Babylonian mathematics in the 1930s by Otto Neugebauer. Bartel L. van der Waerden, following Neugebauer, argues in his influential *Scientific Awakening* (first published in Dutch in 1950, then in English in 1954) that the Greeks clothed their algebra in geometrical dress under the duress of the discovery of irrational quantities.

[7] See Szabó (1978), especially the Postscript, and Unguru (1979a) and (1979b). See also Høyrup (2004) for a thoughtful discussion of the ways in which "mathematical concepts and conceptual structures are formed in interaction with tools within a practice" despite there being no "clear one-to-one correspondences between practices and mathematical conceptual structures" (Høyrup 2004, 134).

practice), to take the system of mathematics presented in the *Elements* to be an (imperfectly realized) axiomatic system in which theorems are proven and problems constructed though chains of diagram-based reasoning about instances of the relevant geometrical figures. But this reading, as a reading of ancient Greek mathematical practice, is as misguided as the algebraic reading. The *Elements* is not best thought of as an axiomatic system in our sense but is more like a system of natural deduction; its Common Notions, Postulates, and Definitions function not as premises from which to reason but instead as rules or principles according to which to reason. Furthermore, and especially important to the historical developments that are of concern to us here, demonstrations in Euclid do not involve reasoning about *instances* of geometrical figures, particular lines, triangles, and so on, but are instead general throughout. Equally importantly, the chain of reasoning is not merely diagram-based, its moves, at least some of them, licensed or justified by manifest features of the diagram. The reasoning is instead diagrammatic. One reasons *in* the diagram in Euclid, or so it will be argued.

2.1 Some Preliminary Distinctions

In the science of mathematics, whether ancient or modern, truths are established by way of some sort of demonstration. If one does not have a demonstration then, however obvious it is that one's claim is true, one has only a conjecture, not mathematical knowledge.[8] A (written) mathematical demonstration is, furthermore, self-standing: by contrast with (written reports of) empirical findings, a mathematical proof *itself* carries evidentiary force; the intentions, reliability, and trustworthiness of its author are irrelevant to its cogency.[9] In mathematics one must be able, at least in principle, to see for oneself that things are as the author argues they are. (This is also the case in philosophy.) Mathematical knowledge is, in this sense, a priori; it is a priori insofar as it does not rely on empirical evidence, whether the evidence of one's own senses or the testimony of another. Mathematical knowledge relies instead on self-standing proof.

In mathematics one can see for oneself how it goes. But this can mean different things in different cases. Suppose, for example, that you are given the task of dividing eight hundred and seventy-three by seventeen. You might solve the problem by engaging in a bit of mental arithmetic that you could report as follows.

[8] Of course, proof in general requires that something be given at the outset, whether in the form of axioms or something else. Not *everything* can be proved, even in mathematics. For the moment, we leave this obvious fact aside.

[9] This feature of mathematical proof is sometimes described as its "transferability"; and it has been used by Easwaran (2009) to argue that probabilistic (mathematical) proofs are essentially different from standard mathematical proofs precisely insofar as they are not transferable but depend, as do reports of experimental results generally, on one's recognition of the (Gricean) intention of the author to convey the information reported.

Figure 2.1 A paper-and-pencil calculation in Arabic numeration.

$$\begin{array}{r} \underline{\;51} \\ 17\sqrt{873} \\ \underline{85}\; \\ 23 \\ \underline{17} \\ 6. \end{array}$$

A hundred seventeens is seventeen hundred; so fifty seventeens is half that, eight hundred and fifty. Add one more seventeen—so that now we have fifty-one seventeens—to give eight hundred and sixty-seven. This is six less than eight hundred and seventy-three. So the answer is fifty-one with six remainder.

Alternatively, you might do, or imagine yourself doing, a standard paper-and-pencil calculation in Arabic numeration the result of which will (if you were schooled in North America) look something like the array displayed in Figure 2.1. Either way, one gets the answer that is wanted. But the two collections of marks serve very different purposes. The written English, it seems fair to say, provides a *description* of some reasoning, in particular a description of the steps a person might go through in order to find, by mental arithmetic, the answer to the problem that was posed. The description does not give those steps themselves. To say or write that a hundred seventeens is seventeen hundred is not to calculate a product. Even more obviously, to say or write the words 'add one more seventeen' is not to add one more seventeen. It is to instruct a hearer or reader: at this point in their reasoning, assuming they are following the instructions, they are to add seventeen.

The paper-and-pencil calculation is radically different. It is manifestly not a *description* of a chain of reasoning one might undertake; rather it *shows* a calculation—at least it does so to one familiar with this use of signs to divide one number by another. (Notice that the paper-and-pencil calculation does not show the steps that are described in the first passage; calculating in Arabic numeration does not merely reproduce mental arithmetic but involves its own style of computing.) Working through this collection of signs in Arabic numeration, or performing the calculation from scratch for oneself, just *is* to calculate. Here we have not a *description* of reasoning but instead a *display* of reasoning. One calculates *in* the system of Arabic numeration in a way that is simply impossible in natural language. To one who is literate in this use of the system of written signs, the calculation shows the reasoning; it embodies it.

This distinction between *describing* or *reporting* a chain of reasoning in some natural language and *displaying* or *embodying* a chain of reasoning, for instance, in the system of Arabic numeration, is at once intuitively obvious and elusive insofar as it seems natural in a way to say that one reasons in natural language more or less as one reasons in, say, Arabic numeration. Certainly it is true that when I do mental arithmetic I in some sense do it in English as contrasted with Chinese or Italian. But one does not reason *in* natural language in the same sense in which one reasons in (say) Arabic numeration. One cannot divide the word 'sixty-nine' by three any more than one can bisect the word 'line', but one *can* divide the numeral '69' by three (in a calculation) as one can bisect a drawn line (in geometry). Arabic numerals are designed to be operated on; written words are not. The function of written words is, first and foremost, to record speech. I can tell, by speaking or writing, what happens or what is the case, how I have reasoned or what is true, but the reasoning itself is something different, as a different example may help to make clear: the proof, known already to the ancient Greeks, that there is no largest prime. Here it is.

Suppose that there are only finitely many primes, and that we have an ordered list of all of them. Now consider the number that is the product of all these primes plus one. Either this new number is prime or it is not. If it is prime then we have a prime number that is larger than all those originally listed; and if it is not prime then, because none of the numbers on our list divide this new number without remainder (because it is the product of all those primes plus one), this new number must have a prime divisor larger than any of the primes on our list. Either way there is a prime number larger than any with which we began. Q.E.D.

This proof, like a bit of mental arithmetic, clearly does not rely on any system of written signs. It depends not on the capacity to write but on the capacity to reflect on ideas, and to think, that is, to reason or infer.[10]

A calculation in Arabic numeration is essentially written—though of course the writing can be merely imaginatively performed rather than actually performed. The proof that there is no largest prime is not. Although the words clearly do convey the line of reasoning, the proof is not *in* the words (whether spoken or written); it is not the words that one attends to in the proof that there is no largest prime, but instead the relevant ideas, central among them the idea of a number that is the product of a collection of primes plus one. The task of the proof is to think through what follows in the case of such a number. The words do not display the reasoning but only describe it.[11]

In the case of a calculation in Arabic numeration, a person proficient in the system literally sees how it goes; the reasoning is not merely reported but is itself embodied in the system of written signs. But not all systems of written signs support reasoning in the same way. In some cases, although one can reason *on* the signs (as we can put it),

[10] As Kant might think of it, whereas a calculation in Arabic numeration involves an intuitive use of reason, a construction (in pure intuition), the reasoning involved in the ancient proof that there is no largest prime instead makes a discursive use of reason directly from concepts. See Kant (1781), first section of the first chapter of the *Transcendental Doctrine of Method* (especially A712/B740–A723/B751). Kant's understanding of the practice of mathematics is taken up in Chapter 4.

[11] As we will see in section 6.4, the reasoning that is needed in proofs of theorems in current college-level textbooks is similarly described rather than displayed. This is furthermore characteristic of the reasoning that is involved in the mathematical practice that first emerged in the nineteenth century in Germany, and especially significant to our concerns in this work. We will need to understand why current mathematics has no (mathematical) language within which to work. We will also need to understand what such a language within which to display the sort of reasoning that is involved in contemporary mathematical practice might look like.

one cannot reason *in* the signs. This distinction is critical to our philosophical purposes, but also subtle. We will approach it by considering a range of cases.

Case One. I am a merchant selling eggs. I keep my eggs in boxes of a dozen each; that is, when I have twelve eggs I put them in a box. And when I have six boxes I put them in a crate for storage. Suppose now that I have three crates, five boxes, and eight loose eggs. I receive a delivery of two crates, two boxes, and seven eggs. I add the two crates to the three I already have and the two boxes to the five I already have. Because I now have seven boxes, I put six of them into a crate, which I then put with the other crates. Combining the seven newly delivered loose eggs with the loose eggs I already had, I see that I have enough to fill a box so I do that and put this new box with the one that was left over after I crated the six boxes, leaving me with three loose eggs. Looking around my shop I see that I now have six crates, two boxes, and three loose eggs.

Case Two. I am again an egg merchant with three crates, five boxes, and eight loose eggs, and again receive delivery of two crates, two boxes, and seven eggs. Because I have hired someone to do the menial labor around the shop, I do not now deal directly with the eggs. Nonetheless, I want to know how many I have and for this I have devised a system of written marks. Using strokes to stand for loose eggs, one for each, crosses to stand for boxes, one for each, and crosshatches to stand for crates, one for each, my record of what I already have in store looks like this:

♯♯♯ + + + + + ////////.

Now I add to my tally the appropriate signs standing for what has been delivered, namely, two crates, two boxes, and seven eggs:

♯♯♯♯♯ + + + + + + + ///////////////.

And much as before I did with the eggs and boxes, I now put twelve strokes "into" one cross and six crosses "into" one crosshatch. That is, I erase twelve strokes and add a cross, and similarly, replace six of the crosses with a crosshatch:

♯♯♯♯♯♯ + + ///.

Without going anywhere near my merchandize I see that I have six crates, two boxes, and three eggs.

Case Three. Having diversified my enterprise so that I now deal in eggs, various sorts of fruits, three kinds of grains, and a wide array of spices, I have given up trying to have different signs for all the different sorts of goods. I adopt instead a numeral notation that can be used for all, together with signs for each sort of foodstuff so as to be able to record, using separate sorts of signs, both how many and of what. I adopt, let us say, the system of Roman numeration: 'I' stands for one thing, 'V' for five, 'X' for ten, 'L' for fifty, 'C' for one hundred, 'D' for five hundred, and 'M' for one thousand. As it helps me to think of it, V is a bag of five things; two V-bags go in an X-box; five X-boxes make an L-crate; two L-crates go in a C-bin; five C-bins are a D-pallet; and two D-pallets form an M-store. As before, I have three crates, five boxes, and eight loose eggs. Each crate contains six times XII eggs: XXXXXXIIIIIIIIIIII, that is, LXXII eggs. So the number of eggs in three crates is that taken three times: LLLXXXXXXIIIIII or CCXVI eggs. Five boxes is XII taken five times: XXXXXIIIIIIIIII, or LX. I now can determine how many three crates, five boxes, and eight loose eggs is simply by putting all the signs together and replacing signs as necessary: CCLXXVVIIII, that is, CCLXXXIIII. The new delivery is, again, two crates, two boxes, and seven eggs. I know that one crate has LXXII eggs in it, so two crates is double that: LLXXXXIIII or CXXXXIIII. Two boxes is double one: XXIIII, and I get the total number of eggs delivered by putting together the numerals for crates, boxes, and eggs: CXXXXXXVIIIIIIIIII, that is, CLXXV. Adding that to my existing stock gives CCCLLXXXXXVIIII or CCCCLVIIII.[12]

Case Four. Having been taught the positional system of Arabic numeration together with the means of calculating in this system, and faced with the same problem of determining how many eggs I have in light of the new shipment, I now determine that my existing stock of three crates, five boxes, and eight eggs is 3 times 6 times 12, and 5 times 12, and 8. I know, because I have memorized it, that 3 times 6 is 18. To determine 18 times 12, I write '18' then just below it '12', and do what is to us a very familiar paper-and-pencil calculation that yields 216. A similar calculation shows me that 5 times 12 is 60. So I add, in the way we all learned as children, 216 and 60 and 8 to give 284. The new delivery is two crates, two boxes, and seven eggs: 2 times 6, that is, 12, times 12, plus 2 times 12, plus 7. A paper-and-pencil calculation tells me that 12 times 12 is 144 and that 2 times 12 is 24. The new delivery is then 144 plus 24 plus 7, which I add up the usual paper-and-pencil way to give 175. Now I need to add 284 and 175, which I do by more paper-and-pencil scribbling to give 459.

In the first case, in which I deal directly with eggs, boxes, and crates, I am clearly not calculating in any arithmetically interesting sense. I am working with stock, not numbers. In the second case, I am working with signs rather than with stock. The signs picture the stock and I can manipulate the signs much as before I had manipulated the stock. Because in the third case I do not directly picture eggs, boxes, and crates but instead use a multipurpose system of numeration, it takes a little more work to record what I have in the notation, but here again the marks I use provide a kind of picture of the stock, of how many eggs I have, and I can manipulate the signs as before I had manipulated the stock. Although I keep my eggs in boxes of twelve eggs and crates of six boxes, in my reckoning I think of them instead as in bags of five eggs, boxes of two bags, crates of five boxes, and so on. In all three cases I literally add things together (and would similarly literally take things away if instead

12 As our example might suggest, it is very easy to make mistakes in this system, to miscount the various sorts of symbols or to add or remove too many or too few, which may help to account for the fact that numeration systems that picture collections of objects have not historically been used as we have used it here. Instead an abacus or counting board would have been used for the calculation.

of receiving a delivery I send off a shipment). To determine how many is three boxes I write the signs three times, and were I dividing, say, by three, I would look for three occurrences of a sign to separate into three. The notation thus has an immediacy and concreteness that makes it very intuitive and easy to learn.[13]

The same cannot be said of our last scenario involving the positional system of Arabic numeration. In this system, the signs do not simply stand in for objects and collections of objects that are put together to picture larger collections. The system is not additive: '475' does not mean four and seven and five. Not only the numerals but their positions in a given context of use together determine what number is meant. Of course this system is related to the earlier ones, and one might well try to explain how it works by talking of five units, seven tens, and four hundreds on the model of eggs, boxes, and crates. But it nonetheless works as a system of signs on very different principles from the straightforward picturing conventions of a numeration system such as the Roman system. One does not in the case of Arabic numeration operate on the system of signs in a way that mimics one's manipulation of the objects, and there is nothing particularly intuitive about the rules one does follow. One could know full well how to do a calculation in Arabic numeration without having any idea how or why it works, why doing things that way gives the right result. One could not, I think, in the same way know how to manipulate the signs of Roman numeration—how to put collections of signs together, or apart, how to pack and unpack signs by now putting a 'V' for five strokes and now five strokes for a 'V'—without at the same time seeing why it works. Knowing what the signs mean and how they function to stand for collections just is to be able to make sense of the manipulations with Roman numerals.

Over the past five and a half millennia more than a hundred different numeral notation systems, among them Roman numeration, have been developed for the purpose of recording how many. Indeed, "the primary function of numerical notation is always the simple visual representation of numbers. Most numeral notation systems were never used for arithmetic or mathematics, but only for representation" (Chrisomalis 2010, 30). Instead of using (say) Roman numerals to determine how many, as we did in the third case above, one would use, for instance, an abacus or a counting board; instead of manipulating the signs of Roman numeration, one would manipulate counters in a systematic and highly functional way to achieve the desired arithmetical results.[14] Here again, even more obviously than in the case of our manipulations of Roman numerals, one operates on the counters much as one would operate on the things themselves, and can do because the counters directly stand in for things and collections of things, that is, in effect, for eggs, boxes, and

[13] Lengnink and Schlimm (2010) provide some empirical evidence for this claim.

[14] In such a system, it is impossible to make some of the mistakes that are all too easy to make in manipulating the signs of Roman numeration because one is not rewriting but actually physically moving counters.

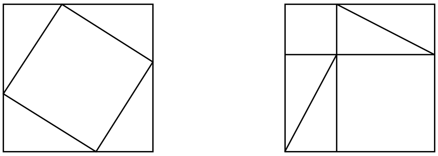


Figure 2.2 A picture proof of the Pythagorean Theorem.

crates. In all such cases, one operates *on* the system of signs rather than *in* it as one does in Arabic numeration.

Although it is possible, as Schlimm and Neth (2008) argue, to determine the answer to an arithmetical problem using Roman numeration, as it is of course possible to do this using Arabic numeration, it does not follow that one is in both cases *calculating* the answer, or at least calculating in just the same sense in the two cases. The fact that the Roman system directly pictures collections and so can be manipulated much as one manipulates objects such as eggs, boxes, and crates, whereas the Arabic does not in the same way directly picture collections and cannot be manipulated as one manipulates objects, matters crucially, at least for our philosophical purposes here. It is just this distinction that I aim to mark by saying that although one can reason *on* collections of Roman numerals, much as one reasons on an abacus or counting board, one cannot reason *in* that system as one does in Arabic numeration. And the same is true, for the same reason, of, for instance, representations of knots in knot theory that can be manipulated using Reidemeister moves. Because the representations directly picture knots, they can be manipulated much as the knots themselves can be. In our terminology, one operates on the system of signs but not in it. And the same would seem to be true of some picture proofs, for instance, the picture proof of the Pythagorean theorem shown in Figure 2.2. Here one directly pictures areas and imaginatively moves them around in a way that reveals that the square on the hypotenuse is equal in area to the sum of the squares on the other two sides. One reasons *on* the display. One of my principal aims will be to show that one does not in the same way reason *on* a diagram in Euclid. Instead one reasons *in* the diagram as one reasons in the written system of Arabic numeration.

Other picture proofs help to bring out another contrast that is relevant here, namely that between reasoning (generally one-step) that makes explicit something that is otherwise only implicit, and the sort of (multi-step) reasoning that one finds in Euclid and, it will be argued, does not merely make something explicit but instead actualizes a potential. Consider, for example, the familiar picture proof to show that $¼ + 1/16 + 1/32 + \ldots + 1/4^n + \ldots = 1/3$ displayed in Figure 2.3. Here the display formulates the information in the problem—that one is to add one quarter and one sixteenth and . . . —in a way that enables one to see that in the limit one will have taken one third of the unit square. Looking at the display one way enables one to see it as picturing the infinite sum; looking at it another way one sees it as marking off

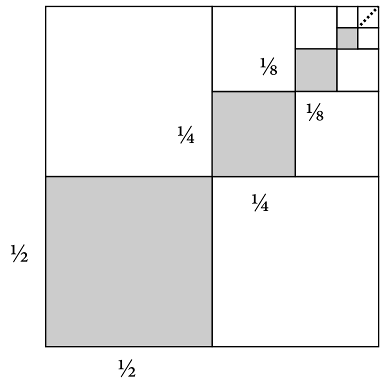


Figure 2.3 A picture proof showing that the series ¼ + 1/16 + . . . + $1/4^n$ + . . . sums to 1/3.

one third of the unit square. One can see the display either way and because one can, one comes to see that the equality given above is true. The signs flanking the equal sign express different senses but as the display shows, they designate one and the same thing. Here, it is the shift in one's perceptual focus that effects what we would otherwise think of as a step in reasoning, and what that shift serves to do is to make explicit information that is already implicit in the display.

Similarly in a Venn diagram, or an Euler diagram, one pictures the contents of the premises of a syllogism in a way that (again, by shifting one's perceptual focus) enables one to read off the valid conclusion if there is one.[15] In all three cases—the picture proof in Figure 2.3, a Venn diagram, and an Euler diagram—one pictures the given information in a way that serves implicitly to picture also the desired result. Shifting one's visual attention appropriately one makes that implicit content explicit and thereby sees that the result holds. Diagrammatic reasoning in Euclid, we will see, is not like this; a Euclidean diagram is not a picture proof.[16] Instead, much as a calculation in Arabic numeration does, a Euclidean diagram formulates content in a mathematically tractable way, in a way enabling one to reason in the system of signs in a step-wise fashion from the given starting point to the desired endpoint.

We have drawn three distinctions. The first was between describing or reporting some reasoning in (written or spoken) natural language as contrasted with displaying or embodying the reasoning in a written system of signs. The second distinction was that between reasoning *on* some collection of marks (or collection of counters) and

[15] For further discussion of these two systems, of their similarities and differences see Gurr, Lee, and Stenning (1998). For extensions of Venn-style diagrams to cases involving more than three, even indefinitely many, terms see Moktefi (2008) and Moktefi and Edwards (2011).

[16] We need, then, to distinguish between a broad sense of diagrammatic reasoning according to which it is any reasoning or problem solving involving pictorial representations—as in Chandrasekaran, Glasgow, and Narayanan (1995) and Larkin and Simon (1987)—and the narrower sense of diagrammatic reasoning that is characteristic of reasoning in Euclid and is our particular concern here.

reasoning *in* a system of signs. We saw that when a display directly pictures something, say, a collection of objects or a knot, it is possible to reason on the display as one might manipulate the relevant object or objects. One does not, in that case, reason *in* the system of signs. What is required in the case of a system of signs within which to reason remains to be clarified. Our last distinction was that between picture proofs, which for these purposes include Euler and Venn diagrams, on the one hand, and Euclid's diagrams, on the other. What is characteristic of picture proofs, at least those of concern here, is that they formulate information in a way that allows one to see in the display both the information with which one began and that one aims to discover. Looked at one way, one sees the starting point, and looked at another, one sees the desired result. The display encodes the starting point in a way that makes the result implicit in that encoding; to make it explicit one needs only to shift one's visual focus a bit. As we will soon see, Euclidean diagrams are much more complex, and fruitful, than any such display.

2.2 Euclid's Constructions

Euclid's geometrical practice involves the use of diagrams in proofs. Indeed Euclid's demonstrations involve diagrams even in cases, such as those in Books VII to IX regarding numbers, in which they play no argumentative role. And not only are diagrams used in Euclid's demonstrations, very often what is to be demonstrated just is that some figure can be drawn, some construction effected. Propositions in Euclid, that is to say, are of two sorts: problems in which it is shown how to construct something—say, an equilateral triangle on a given line (proposition I.1), or a point of bisection on a given line (proposition I.10)—and theorems in which it is shown that some claim, usually in the form of a generalized conditional, is true, for instance, that if two straight lines cut one another, they make the vertical angles equal to one another (proposition I.15). Furthermore, according to Proclus, head of the Academy in fifth-century Athens, such constructions are "rightly preliminary" to theorems. Euclid begins the *Elements*, for example, with three construction problems, among them that of constructing an equilateral triangle on a given straight line, and then turns, in proposition I.4, to a proof of a general theorem about triangles. Proclus explains why.

> Suppose someone, before these have been constructed, should say: "If two triangles have this attribute, they will necessarily also have that." Would it not be easy for anyone to meet this assertion with, "Do we know whether a triangle can be constructed at all?" . . . It is to forestall such objections that the author of the *Elements* has given us the construction of triangles . . . . these propositions are rightly preliminary [to the theorems about the congruence of triangles].[17]

Constructions must precede the demonstration of theorems about the entities constructed because otherwise it could be objected that we do not know whether the

[17] Quoted in Knorr (1983, 126).

objects involved can be constructed. But why is this an objection? What exactly is it that we do not know when we do not know how to construct a triangle? One very common answer is that we do not know, in that case, whether triangles exist.

It can seem obvious, at least to us, from the perspective of our current conceptions, that construction problems in Euclid serve as existence proofs, that they are included in order to show that the constructed object exists. Nevertheless, as Knorr and others have argued, the reading is anachronistic.[18] Although questions about existence did occasionally arise in Greek mathematics, "the thesis that the ancient constructions were intended as a form of existence proof accounts neither for the geometers' manner of treating problems of construction nor for their ways of handling issues of existence" (Knorr 1983, 135). For example, it is well known that many construction problems cannot be solved using only the resources provided by Euclid, among them, the problems of trisecting an angle and squaring a circle. The issue for the Greeks was not, however, whether trisecting an angle or squaring a circle is possible. It seemed to them obvious that the trisection or square existed; the problem was to find a method or procedure for constructing such a thing, "not to establish the *existence* of the solution (for that is simply assumed), but rather to discover the *manner* of its construction" (Knorr 1983, 140).[19] An illustrative analogy might be this. Suppose that you know, because you have counted them, that you have fifty-seven boxes, each containing twenty-three jars into each of which you have put exactly thirty-five olives. One would, in such a case, have no doubt that there is some number that is the total number of olives in all the boxes. Nevertheless, one would, in the absence of some means such as those provided in our examples above, be unable to *produce* the total number from the given information. (Of course one could just count the olives, but that is quite different from producing the number that is wanted from the given information.) Similarly, Greek mathematicians were in no doubt that there is, say, a square equal in area to a given circle despite having no way of producing, or constructing, that square from the circle given the resources Euclid provides.

As Knorr argues, construction problems do not function as existence proofs in Euclid. Might it nonetheless be the case that constructions are needed in the *course* of Euclidean demonstrations, whether of problems or of theorems, in order to show that the various entities required by the process of reasoning exist? Demonstrations in Euclid involve, in every case, both a *kataskeue* (construction of the diagram) and, following it, an *apodeixis* (proof).[20] Is the construction stage perhaps needed because,

[18] See Knorr (1983), also Lachterman (1989), and Harari (2003). Saito (2008, 808–9) also briefly discusses this issue.

[19] Knorr (1983, 132) further quotes Philoponus, who writes in his commentary on Aristotle's *Posterior Analytics* I 9 that "those who square the circle did not inquire whether it is possible that a square be equal to the circle, but by supposing that it can exist they thus tried to produce (γενναν) a square equal to the circle." As Knorr also notes in this context, Eutocius took it to be obvious that there is a straight line that is equal in length to the circumference of a circle, despite the fact that no one knew how to construct such a line.

[20] The translations are standard (see Heath, 1956, vol. I, 129); nevertheless, they may be misleading. On the question of the translation of '*kataskeue*' in particular, see Harari (2003, 2), also Lachterman (1989, 57).

as Mueller (1981, 15) suggests, in a system such as Euclid's "the existence of one object is always inferred from the existence of another by means of a construction"? Friedman makes a very similar claim.

> In Euclid existence assumptions are represented not via propositions formulated in modern quantificational logic but rather by constructive operations—generating lines, circles, and so on in geometry—which can then be iterated indefinitely. Such indefinite iteration of constructive operations takes the place, as it were, of our use of quantificational logic, and it is essentially different, moreover, from the inferential procedures of traditional syllogistic logic. (Friedman 1992, xiii–xiv)[21]

On this reading, Euclid's demonstrations, whether of problems or of theorems, are not strictly deductive, logically valid because they require the use of postulates, rules of construction, that cannot be formulated as axioms without the resources of the full polyadic predicate calculus. Lacking those resources, Euclid could not include among his axioms that, say, between any two points there is a third and therefore could not deduce the existence of the points and lines that are needed in the course of various proofs. The existence of such points and lines is instead established by constructions according to the rules laid out in the postulates. The constructions are needed because the logic is monadic; given a more powerful logic, one could reason deductively throughout.

On Friedman's account, Euclid's constructions serve to insure the existence of the points and lines that are needed in Euclid's demonstrations. Friedman explicitly connects the idea to the familiar objection that Euclid merely assumes, for example, in proposition I.1, that there exists a point of intersection where two lines cross. The complaint, Friedman suggests, is misguided insofar as Euclid does not *assume* that the relevant point exists but instead *constructs* it—or, as Friedman (1992, 61) says, "generates" it. Clearly, a point that is given at the crossing of two lines is not constructed in the sense in which, say, a triangle or even a point of bisection is constructed using Euclid's procedures; it could not be a *problem* of Euclidean geometry to construct a point at the intersection of two lines. That there is a point at the intersection of two lines is not something that could be demonstrated in a Euclidean construction because only a whole of parts in relation can be constructed and the point at the intersection of two lines is not a whole of parts in relation. Nor even is such a point *generated*, that is, produced, brought into existence, shown to exist, by the intersection of lines, as Friedman suggests. That two crossed lines have a point of intersection is not, and could not be, *justified* by the construction but must instead be presupposed by it. Euclid simply assumes without comment, perhaps in light of continuity considerations, that there is a point at the crossing of two lines (Knorr 1983, 133). Even within a construction, then, the aim cannot be to establish the existence of the points and lines that are needed.

[21] See also Friedman (1992, ch. 1).

Ancient Greek mathematical practice suggests that neither construction problems nor the constructions that are appealed to in the course of demonstrations in Euclid's *Elements* serve to establish the existence of something. Indeed, problems regarding existence do not in general have the significance in ancient thought that they would come to have in the modern period. As Burnyeat (1982, 32) argues,

> Greek philosophy is perfectly prepared to think that reality may be entirely different from what we ordinarily take it to be .... But all these philosophers, however radical their scrutiny of ordinary belief, leave untouched—indeed they rely upon—the notion that we are deceived or ignorant about *something*. There is a reality of some sort confronting us; we are in touch with something, even if this something, reality, is not at all what we think it to be.

Roughly speaking, whereas modern philosophy (beginning with Descartes) takes essence (what a thing is) to precede existence (that it is), for the ancients, it does not. One cannot, on the ancient view, ask what something is, for example, what a triangle is—what it is to be a triangle (its essence)—unless one knows that triangles exist. And this is because, for the ancients, existence and essence are inextricably combined: to be is to be something in particular, that is, some sort of thing, something with a nature, form, or essence; and conversely, only what is has an essence, a nature or form. Whereas we find it natural to distinguish logically between the question what it is to be, say, a triangle or a circle, and the question whether there exist in reality any triangles or circles, no such distinction is made in ancient Greek thought: "the notions of existence and predication, which we distinguish as two separate logical or linguistic functions, are conceived in Greek as two sides of a single coin" (Kahn 1981, 123).[22] Thought, on such a view, is inextricably related to something that exists independent of it; it is inevitably directed on things without the mind. Indeed, this is just what we should expect given the intentional directedness that is enabled by natural language. As we have already seen, natural language is by its nature object oriented, and because it is, it will take a radical transformation, a *metamorphosis* in our mode of intentional directedness, for it to come to be so much as *intelligible* that, as Descartes suggests, the world without the mind might not exist.[23]

Much as mathematicians, both ancient and modern, distinguish between knowing that something is true and being able to prove it, so ancient geometers distinguish between knowing that something exists and being able to construct it. Indeed, they seem to have been primarily interested in constructions—that is, in finding ways of producing geometrical figures of mathematical interest beginning only with circles, points, and lines—insofar as, Knorr (1983, 140) suggests, in some cases theorems were proven only because they were needed to enable constructions. According to

[22] See also Owen (1965).

[23] Although skepticism about the world without the mind might have been conceivable to the ancient Greeks as an abstract or academic possibility, as a kind of limit case of sensory illusion, it could not, for reasons that will become clear in Chapter 3, actually be made intelligible before Descartes' revolutionary discoveries in mathematics.

Knorr (1983, 139), the solution of construction problems "constitutes in effect what the ancients *mean* by mathematical knowledge." Any adequate account of ancient Greek diagrammatic practice must be able to make sense of this.[24]

2.3 Propaedeutic to the Practice

The *Elements* opens with a series of definitions, postulates, and common notions. These are not, we will see, the starting points for the mathematics to follow, not if by that we mean the first, foundational truths from which other truths will be derived. Instead, they function as a propaedeutic to the practice of geometry. They belong not to geometry itself but only to the antechamber of geometry. They provide a "preamble" to the actual work of mathematics (Burnyeat 2000, 23),[25] one that amounts to "a constitution for Euclid's subject matter" (Reed 1995, 21) and enables one thereby "to 'read' or interpret diagrams which relate the parts of figures to one another" (Reed 1995, 52).[26]

Consider, first, the definitions. In current mathematical practice, definitions are understood as stipulations that fix the meaning of some newly introduced sign by setting out in the primitive (and perhaps previously defined) signs of the language what the newly introduced sign is to mean. Such definitions function, as the axioms of the system do, to provide premises for inferences.[27] One defines, for instance, the concept of a group using various primitive notions, and then one proves theorems about groups on the basis of that definition. The primitive notions cannot, of course, be similarly defined; and yet they must be understood. Instead they are elucidated in prefatory remarks, that is, in remarks that belong not to the system itself, in which starting points are set out and inferences drawn from them, but in the preparation for setting out the starting points. Euclid's definitions, we will see, function as elucidations in this sense. They belong to the preparation rather than to the system itself.

[24] Grosholz (2007), section 2.1.1, also emphasizes the importance of, as she puts it, the analysis of intelligible things in Euclid, "what makes a shape the shape it is" (Grosholz 2007, 36), and highlights as well the fact that in such an analysis the whole is not reduced to its parts: "no part has a relation to another part that is not mediated by the whole to which they belong." Nevertheless, it is important to recognize that although regarded one way the whole is prior to the part, regarded another the part is intelligible independent of the whole. Because the side of a triangle is a line, it is intelligible independent of the triangle, despite the fact that regarded as a side it is intelligible only in relation to the triangle as a whole. Such a whole is, in other words, neither an essential unity, as a living body is, the parts intelligible *only* in relation to the whole, nor an accidental unity, the whole *reducible* to the parts in relation. It is an *intelligible unity* of parts within the whole, neither of which is prior to the other.

[25] As Netz (1999, 95) puts it for the case of Euclid's definitions in particular, they are "simply part of the introductory prose . . . Before getting down to work, the mathematician describes what he is doing—that's all."

[26] Reed (1995, ch. 1) explains in detail what this might mean.

[27] Perhaps it will be objected that definitions are merely stipulations about abbreviations, that they have no essential role to play in the reasoning analogous to that played by the axioms. This is the standard view of philosophers. It is also, we will eventually see, mistaken.

As we now understand them, definitions concern words, or other signs; a definition sets out the meaning of a word or newly introduced sign. According to the ancient Greeks, definitions are not of words (or signs) but of what words name. That is, as Aristotle explains in the *Metaphysics* VII.4, "we have a definition not where we have a word and a formula identical in meaning ... but where there is a formula of something primary [i.e., substance]" (1030a7–10).[28] "Only substance is definable" (1031a1). And only substance is definable because, as Aristotle explains in *Posterior Analytics* II.3, "definition is of the essential nature or being of something" (90b30); "definition reveals essential nature" (91a1). On Aristotle's view, which is that afforded by natural language, what most fundamentally exist are substances, that is, beings with natures, paradigmatically, living things. Again, to be is to be some particular sort of thing, something with a nature. Definitions as understood by the ancient Greeks set out what it is to be for such beings.

It is furthermore clear, as already indicated, that on Aristotle's view only what exists has such a nature: "he who knows what human—or any other—nature is, must know also that men exist; for no one knows the nature of what does not exist—one can know the meaning of the phrase or name 'goat-stag' but not what the essential nature of a goat-stag is" (*Posterior Analytics* 92b4–7). We know that nothing is a goat-stag because such a thing would have to be at once a goat and not a goat (because a stag) and nothing could be that. Because there is (demonstrably) nothing that can correctly be called a goat-stag, the term 'goat-stag' is not really a *name*. And hence, although the word has a perfectly clear meaning, no definition can be given of a goat-stag. Existence is in this way a precondition of essence, and hence of definition. But if so, then the definitions with which Euclid begins are not of words and do not leave it open whether the defined entities exist. Their existence is presupposed. What the definitions provide is only an account of what it is to be this or that geometrical entity, that is, their essential natures.[29]

According to Aristotle, definition is of the essential nature of (existing) things. It sets out what a thing most fundamentally is, its nature. In the case of a living being nature is given by reference to a form of life, the form of life of a rational animal, say. In the case of the elements, that is, earth, air, fire, and water, nature is given by certain sensory properties, the hot and the cold, and the wet and the dry: earth is cold and dry, fire hot and dry, and so on. In the cases of concern to Euclid, nature is given by parts in relation. In these cases, then, the definitions set out what parts in relation are required if something is to be a particular geometrical object, or a part thereof, what parts in relation are required for a particular property or relation of geometrical

[28] See also Le Blond (1979).

[29] This is not to say that there were not dialectical challenges to the idea that the objects of mathematics exist. There were. (See, for instance, Mueller 1982.) Such challenges were, however, of no concern to the practicing mathematician. And of course there were known cases in which alleged mathematical entities, such as the numerical ratio of the diagonal to the side of a square, demonstrably do not exist. These were, however, exceptions to the rule in Greek mathematics.

entities to obtain. We are told, for example, that a triangle is a rectilineal figure contained by three lines, rectilineal, figure, and line all having previously been defined. Similarly we are told that a number is a multiple composed of units, a unit having been previously defined as that in accordance with which each of the things that exist is called one. Definitions in terms of relations of parts are given of properties: we are told, for example, what it is for a line to be straight (it is a line that lies evenly with the points on itself), and for a number to be prime (it is a number that is measured only by a unit). And definitions in terms of relations of parts are provided, finally, of what it is to bear some particular relation, for instance, of magnitudes what it is to have a ratio one to another (just in case they are capable, when multiplied, of exceeding one another), or of circles what it is to touch one another (just in case they meet but do not cut one another). Because it is in the nature of such entities to be wholes of parts in this way, these sorts of things are, as Kant was perhaps the first to note explicitly, constitutively such as to be iconically representable. Drawings of geometrical entities "show in their composition the constituent concepts of which the whole idea . . . consists" (Kant 1764, 251; AK 2:278). The drawings like the things drawn are wholes that are made up of parts in relation. As Kant also sees, such wholes can then further be combined to show "in their combinations the relations of the . . . thoughts to each other" (Kant 1764, 251; AK 2:278).[30] In geometry—indeed, Kant thinks, in mathematics generally—one combines in a diagram the wholes that are created out of simples into larger wholes that exhibit relations among them. We will return to this.

In Book I of the *Elements*, after the definitions have been set out, Euclid lists some postulates and common notions. We know already from the definitions that, for example, the extremities of a line are points; now it is postulated "to draw a straight line from any point to any point." We know already that when a straight line is set up on a straight line so as to make the adjacent angles equal then those equal angles are right; now it is postulated that all right angles are equal to one another. And we know already that parallel lines are straight lines in a plane that never meet no matter how much they are extended; now it is postulated that if the interior angles on the same side of a transverse line are both acute, that is, less than a right angle, then the two lines crossed by the transverse line will meet if extended far enough. It is also postulated that a straight line can be extended and a circle described given a center and distance. The common notions then set out fundamental features of equality, for instance, that things equal to the same are equal, that things coinciding with one another are equal, and that the whole is greater than the part. What is at issue is whether these postulates and common notions function as premises *from which* to reason or instead as rules or principles *according to which* to reason.

[30] Kant is in fact describing what the words of natural language that are used in philosophy cannot do. It is clear that he means indirectly to say what the marks used in mathematics can do.

In an axiomatic system, a list of axioms is provided (perhaps along with an explicitly stated rule or rules of inference) on the basis of which to deduce theorems. Axioms are judgments furnishing premises for inferences. In a natural deduction system one is provided not with axioms but instead with a variety of rules of inference governing the sorts of inferential moves from premises to conclusions that are legitimate in the system. In natural deduction, one must furnish the premises oneself; the rules only tell you how to go on. The question whether Euclid's system is an axiomatic system or not is, then, a question about how the postulates, and common notions that are laid out in advance of Euclid's demonstrations actually function, whether as *premises* or as *rules* of construction and inference. Do they function to provide starting points for reasoning or do they instead govern one's passage, in the *kataskeue*, from one construction to another, and in the *apodeixis*, from one judgment to another? Inspection of the *Elements* strongly suggests the latter. In Euclid's demonstrations, the common notions and postulates are not treated as premises; instead they function, albeit only implicitly, as rules constraining what may be drawn in a diagram and what may be inferred given that something is true.[31] They provide the rules of the game, not its opening positions.

Consider, for example, the first three postulates. They govern what can be drawn in the course of constructing a diagram: (1) if you have two points then a line (and only one) may be produced with the two points as endpoints; (2) a finite line may be continued; and (3) if you have a point and a line segment or distance then a circle may be produced with that point as center and that distance as radius. In each case, one's starting point—points and lines—must be supplied from elsewhere in order for the postulate to be applied. And nothing can be done in constructing a diagram, at least at first, that is not allowed by one of these postulates. But once they have been demonstrated in problems, various other rules of construction can be used as well. For instance, once it has been shown, using circles, lines, and points, that an equilateral triangle can be constructed on a given finite straight line (proposition I.1), one may in subsequent constructions immediately draw an equilateral triangle, without any intermediate steps or constructions, provided that one has the appropriate line segment. Propositions such as I.1 that solve construction problems function in Euclid's practice as derived rules of construction. Once they have been demonstrated, they can be used in the construction of diagrams just like the postulates themselves, as rules governing those constructions.

Euclid's common notions, and again most obviously the first three, also govern moves one can make in the course of a demonstration, in this case in the course of the *apodeixis*. They govern what may be inferred: (1) if two things are both equal to a third then it can be inferred that they are equal to one another; (2) if equals are added to equals then it follows that the wholes are equal; (3) if equals be subtracted from equals,

[31] As we will see, Euclid in fact almost never invokes his definitions, postulates, and common notions in the course of a demonstration. They are nevertheless readily identifiable as warranting the moves that are made.

then the remainders are equal.[32] These common notions manifestly have the form of generalized conditionals; that is, they have the form of rules of inference.[33] Furthermore, in this case as well, theorems, once demonstrated, can function in subsequent demonstrations as derived rules of inference. Once it has been established that, say, the Pythagorean theorem is true (I.47), one may henceforth infer directly from something's being a right triangle that the square on the hypotenuse is equal to the sum of the squares on the sides containing the right angle. Euclid's *Elements* provides in this way the elements, the rules of construction (in its demonstrations of problems) and inference (in its demonstrations of theorems), for more advanced mathematical work.

But not everything that happens in the course of a Euclidean demonstration is governed by a stated rule, whether primitive or derived. There are two sorts of cases. First, Euclid draws (explicitly or implicitly) various obviously valid inferences, such as that two things are not equal given that one is larger than the other, despite the fact that the rule governing the passage is nowhere explicitly stated. Because rules such as this do not belong to mathematics in particular, but are simply a part of our understanding of natural language, no special mention is made of them.[34] The second sort of case is more interesting. It is well known that in order to follow a demonstration in Euclid, one must read various things off the relevant diagrams. For example, as we have already seen, given two lines that intersect in a diagram, Euclid assumes that there is a point at their intersection. The point of intersection seems simply to "pop up" in the diagram as drawn, and is henceforth available to one in the course of one's reasoning.[35] This happens not once but twice in the very first proposition of Book I of the *Elements*, to construct on a given finite straight line an equilateral triangle.

The demonstration begins with the *ekthesis* (setting out): let there be a straight line, AB.[36] A statement of what is to be done follows: to construct an equilateral triangle on AB. Then the *kataskeue* is given:

[C1.] With center A and distance AB let the circle BCD be described. [This is licensed by the third postulate, though Euclid does not mention this.]
[C2.] With center B and distance BA let the circle ACE be described. [Again, the warrant for this, the third postulate, is not mentioned.]
[C3.] From point C, in which the circles cut one another, to the points A, B let the straight lines CA, CB be joined. [This is implicitly licensed by the first postulate on the assumption that there is such a point C.]

32 These are, of course, not formally valid rules of inference; they are instead what Sellars has taught us to call materially valid rules.

33 As Ryle (1950) argues, rules of inference are inherently conditional in form and essentially general.

34 Why, then, does Euclid include among his common notions that the whole is greater than the part? Is this not also something one learns as one learns the language one speaks? The answer may be that Zeno had by then rendered the relation of whole and part more problematic in mathematics than natural language usage might suggest.

35 I borrow this use of the expression "pop up" from Kenneth Manders, to whom I am indebted for helping me to appreciate just how important is this feature of Euclidean diagrams. See Manders (1996) and (2008).

36 I here follow Netz (1999, 43–4).

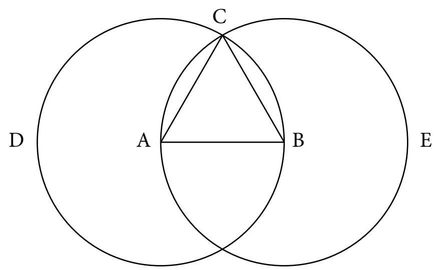


Figure 2.4 The diagram of proposition I.1 of Euclid's *Elements*.

The diagram that results is shown in Figure 2.4. And the *apodeixis* then follows:

[A1.] Given that A is the center of circle CDB, AC is equal to AB. [This is licensed, without mention, by the definition of a circle.]
[A2.] Given that B is the center of circle CAE, BC is equal to BA. [This again is implicitly licensed by the definition of a circle.]
[A3.] Given that AC equals AB and BC equals BA, we can infer that AC equals BC because what are equal to the same are equal to each other [that is, Common Notion 1].
[A4.] Given that AB, BC, and AC are equal to one another, the triangle ABC is equilateral. [This is warranted by the definition of equilateral triangle, on the assumption that there is such a triangle ABC.]
This triangle was constructed on the given finite straight line AB as required, and so we are done.

In the course of this demonstration, first a point pops up at the intersection of the two drawn circles, and then later a triangle pops up, formed from the radii of the two circles. This sort of thing is, furthermore, ubiquitous in ancient Greek geometrical practice. One simply reads the relevant geometrical objects off the diagrams, apparently without any explicitly stated warrant for doing so. Although nothing can be put into a diagram that is not licensed by one of the given postulates or by a previous construction, there seem to be no stated rules governing what, in the way of pop-up objects, can be taken out of it.[37]

We have seen that Euclid's definitions are not of words but of what words name, and that they function more like our elucidations of primitive expressions than like our definitions. And Euclid's common notions and postulates function, we have seen, essentially like rules of a natural deduction system. They are not axiomatic starting

[37] Could Euclid's definitions serve this purpose? Can it be inferred, for example, from the fact that a line is a breadthless length that at the cut of two lines there is a point (i.e., something that has no parts)? Perhaps, but more would need to be said.

points for inference but instead rules governing one's passage from the starting points given by the proposition to be demonstrated to the desired conclusion. But although both common notions and postulates are rules on this view, they are rules governing two very different sorts of activities. Postulates, at least the first three, govern constructions, what can be put into a diagram given the starting points provided by the proposition in question. Common notions instead govern one's reasoning in the course of the demonstration, what may be inferred given something one has already established. Problems and theorems function, respectively, as derived postulates and common notions so conceived. Once a problem has been demonstrated, the relevant construction can immediately be effected if one has the appropriate starting point; once a theorem has been demonstrated, the relevant inference can immediately be drawn given the relevant premise. But, as we also have seen, not everything that is inferred in the course of a Euclidean demonstration is governed by a rule. We have yet to understand the nature and legitimacy of pop-up objects in Euclid. We will return to them when we take up again the Kantian idea that in a diagram parts are related in wholes that are in turn combined in larger wholes.

2.4 Generality in Euclid's Demonstrations

Euclid's demonstration that an equilateral triangle can be drawn on a given finite straight line begins with a "setting out" (*ekthesis*): let there be a straight line, AB. And this is standard in Euclid's demonstrations. Although what is to be demonstrated is something wholly general, the demonstration proceeds, at least in many cases, by way of such a setting out. To anyone familiar with proofs in standard quantificational logic, it is very easy to take this setting out as an analogue of the rule of Universal Instantiation. In quantificational logic, in order to prove that (say) all A is C given that all A is B and that all B is C, one first turns to an instantiation of the premises in order that the rules of the propositional calculus may be applied. One reasons, in effect, about a particular case, and then at the end of the proof one takes what has been shown to apply generally on the grounds that no inference was drawn in the course of the proof that could not have been drawn were any other instance to have been considered instead. We tend to assume that demonstrations in Euclid work the same way.[38]

There is, however, little reason to think that the ancient Greeks understood generality as it is understood in quantificational logic.[39] Although they did

[38] We read, for instance in Russell (1908, 64) that "the general enunciation tells us something about (say) all triangles, while the particular enunciation takes one triangle and asserts the same thing of this one triangle. But the triangle taken is any triangle, not some one special triangle; and thus although, throughout the proof, only one triangle is dealt with the proof retains its generality." Netz (1999, 246) also takes this view: what is to be demonstrated is general but the demonstration itself (that is, the setting out, construction, and *apodeixis*) is particular; its generalizability is "a derivative of [its] repeatability." See also Netz (1999, 262).

[39] The notion of a quantifier, and more generally the quantificational conception of generality, are explored in some detail in section 6.1.

distinguish between general sentences about all or some objects of some sort, and sentences about particular objects, this distinction was not for them, as it is for us, a *logical* distinction. Aristotle's logic reflects this. It is a term logic in which no logical distinction is drawn between referring and predicative expressions (required for quantificational logic's distinction between singular and general sentences). Terms as Aristotle understands them have both referential and predicative aspects; they are essentially object involving (unlike the predicative expressions of standard logic), and also predicative (as referring expressions, as conceived in standard logic, are not). Terms in Aristotle's logic are, as we might think of it, what things are *called*, for instance, 'Socrates', 'man', 'snub-nosed', and so on. It is for precisely this reason that it is legitimate in traditional logic, though not in standard quantificational logic, to infer, on the basis of the fact that all (no) S is P, that some S is (not) P. Again, the ancients did recognize terms, such as 'goat-stag', that nothing is called, but such terms are, for just that reason, not really names. (There are no judgments that can be made about goat-stags.) Ancient Greek thought is inherently world directed; lacking the essentially modern notion of a concept as predicative rather than referential, the Greeks had no way of asking, in general, whether anything answering to one's concepts exists, no way of calling into question the existence of the "external" world.

According to the quantificational reading, a Euclidean demonstration proves something general by proving it for the particular instance that is introduced in the setting out portion of the text with which the demonstration begins. The setting out is essential on this reading insofar as the only way to prove something general in quantificational logic is by way of an instantiation. The diagram, as contrasted with the (textual) setting out, is not required in the same way. The diagram is needed, when it is, only to ensure the existence of the various points, lines, and so on, that are required for the cogency of the course of reasoning. And yet, in the *Elements*, diagrams are invariably included even in cases in which no construction is carried out on them, as in the demonstrations regarding numbers. Proposition VII.4, that "any number is either a part or parts of any number, the less of the greater," is a case in point. The text of the demonstration is this (Euclid 1956, I, 303):

> Let A, BC be two numbers, and let BC be the less; I say that BC is either a part, or parts, [i.e., a submultiple, or proper fraction] of A. For A, BC are either prime to one another or not. First, let A, BC be prime to one another. Then, if BC be divided into the units in it, each unit of those in BC will be in some part of A; so that BC is parts of A. Next, let A, BC not be prime to one another; then BC either measures, or does not measure, A. If now BC measures A, BC is a part of A. But, if not, let the greatest common measure D of A, BC be taken; and let BC be divided into the numbers equal to D, namely BE, EF, FC. Now, since D measures A, D is a part of A. But D is equal to each of the numbers BE, EF, FC; therefore each of the numbers BE, EF, FC is also a part of A; so that BC is parts of A. Therefore, etc. Q.E.D.

The "diagram" that is provided is shown in Figure 2.5. There is in this case no construction beyond the diagrammatic "setting out" of the numbers A, BC, and D;

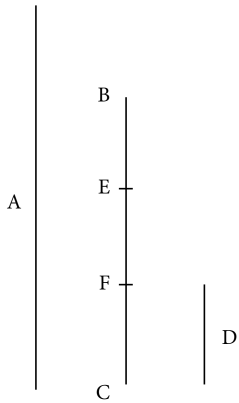


Figure 2.5 The diagram of proposition VII.4 of Euclid's *Elements*.

nothing further is added to the diagram, and nothing is inferred on the basis of it. And yet the diagram is included in this, and in every other, case.

Diagrams are inevitably included in Euclid's demonstrations, even in cases in which they seem to play no role in the demonstration. Surprisingly enough, at least from the quantificational perspective, the setting out is *not* invariably included. In some demonstrations, there is simply no moment of "universal instantiation." In proposition IV.10, for example, there cannot be any such moment because nothing is given as the basis on which to carry out the needed construction. One is merely to construct, out of thin air as it were, an isosceles triangle of a certain sort. The construction is given, and the fact that the triangle so constructed meets the requirement set is shown. But there is no logical moment of setting out; one simply starts with a drawn line and goes from there.[40]

A Euclidean demonstration apparently does not need to have a setting out; but it must have a diagram, and it must have one even in cases in which the diagram seems to serve no purpose in the demonstration. Both features of such demonstrations are anomalous on a quantificational interpretation of them; on the quantificational interpretation it is the setting out, universal instantiation, that is essential and the diagram that is optional, needed only in certain cases. It is furthermore worth noting that in Aristotle's term logic one infers *directly* from the fact that all A is B together with the fact that all B is C to the conclusion that all A is C. The reasoning, that is to say, is general throughout, though not quantificationally general. Perhaps, then, the same is true in the case of reasoning in Euclid. Grice's analysis, in "Meaning" (1957),

[40] Mendell (1998, 179) notes this case and others like it. A different sort of case in which the setting out is absent, noted also by Mendell, is that in which it is to be shown that an object lacks a certain property, for instance, that a circle does not cut a circle at more than two points (III.10). Because the reasoning in this case is by reductio, there is no setting out; one just starts with the needed construction, and thereby with the supposition that a circle can cut a circle at more than two points. This example will be discussed in greater detail in section 2.5.

of the distinction between (as Grice puts it) natural and non-natural meaning can help us to understand how this might work.[41]

Grice points out that there is an intuitively clear distinction between, for example, the sense in which a certain sort of spot on one's skin can mean measles and the sense in which three rings on the bell of a bus can mean that the bus is full.[42] The task is to provide an analysis of the difference between the two senses of 'means', natural and non-natural, respectively; and one of Grice's examples to that end is particularly revealing for our purposes. Grice (1957, 282–3) asks us to compare the following two cases:

(1) I show Mr. X a photograph of Mr. Y displaying undue familiarity with Mrs. X.
(2) I draw a picture of Mr. Y behaving in this manner and show it to Mr. X.

He immediately remarks:

> I find that I want to deny that in (1) the photograph (or my showing it to Mr. X) meant_{NN} [that is, non-naturally] anything at all; while I want to assert that in (2) the picture (or my drawing and showing it) meant_{NN} something (that Mr. Y had been unduly familiar), or at least that I had meant_{NN} by it that Mr. Y had been unduly familiar. What is the difference between the two cases? Surely that in case (1) Mr. X's recognition of my intention to make him believe that there is something between Mr. Y and Mrs. X is (more or less) irrelevant to the production of this effect by the photograph .... But it will make a difference to the effect of my picture on Mr. X whether or not he takes me to be intending to inform him (make him believe something) about Mrs. X, and not to be just doodling or trying to produce a work of art.

Although the photograph can serve to convey to someone the fact that Mr. Y is unduly familiar with Mrs. X independent of anyone's intending that it so serve, the drawing cannot. To take the drawing as conveying a message about Mr. Y's behavior, rather than as a mere doodle or as a work of art, essentially involves taking it that someone produced it with the intention of conveying such a message, and that that person did so with the intention that that intention be recognized (and play a role in the communicative act). Only in that case does the drawing mean (non-naturally) that Mr. Y is unduly familiar with Mrs. X.

A photograph has natural meaning in virtue of a (causally induced) resemblance between the image in the photograph and that of which it is a photograph. A drawing, Grice argues, can in certain circumstances have instead non-natural meaning. That is, it can have the meaning or content that it does in virtue of one's intending that it have that meaning or content (and intending that that intention be recognized and play a certain role in the communicative act). So, we can ask, does a drawn figure in Euclid, a triangle, say, have Gricean natural meaning or instead

[41] See also Dipert (1996).

[42] In the opening paragraphs of his essay, Grice lists five different ways that the use of 'means' differs in the two cases.

Gricean non-natural meaning?[43] If it is a drawing of an instance, a particular geometrical figure, then it has natural meaning. It is in that case a semantic counterpart.[44] It is the thing, say, the particular triangle ABC, that is referred to in the text of a demonstration when one judges of triangle ABC that it is thus and so. But perhaps the drawing, like Grice's drawing, instead has non-natural meaning. Perhaps it is not (say) a circle but instead, as Leibniz (1969, 84) says, is taken for one.[45] If so, then the Euclidean diagram can mean or signify some particular sort of geometrical entity only in virtue of someone's intending that it do so and intending that that intention be recognized. One's intention in making the drawing—an intention that can be seen to be expressed in the setting out (in those cases in which there is one) and throughout the course of the *kataskeue*—is, in that case, indispensable to the diagram's playing the role it is to play in a Euclidean demonstration.

A Euclidean diagram, more exactly, a figure that is discerned within such a diagram, can be interpreted either as having natural or as having non-natural meaning in Grice's sense. If it has natural meaning then it does so in virtue of being an instance of a geometrical figure. But if it has non-natural meaning then it can be inherently general, a drawing of, say, an angle without being a drawing of any angle in particular. The geometer draws some lines to form a rectilineal angle, in order, say, to show that an angle can be bisected. The geometer does not mean or intend to draw an angle that is right, or obtuse, or acute. He means or intends merely to draw *an angle*. Hence he says, for example, in proposition I.9, that the angle he draws is "the given rectilineal angle," not also that it is right, or obtuse, or acute. That which he draws, regarded as something having natural meaning, will necessarily be right, or acute, or obtuse; but regarded as having non-natural meaning, it will be neither right nor acute nor obtuse. It will simply be *an angle*.[46] One can no more infer on the basis of the drawing so conceived that the angle in question is (say) obtuse, based on how it looks, than in Grice's case one could infer (say) that Mr. Y has put on a little weight recently, based on how the drawing of him looks. A drawing of Mr. Y could have as its non-natural meaning that he has put on weight recently, just as in the context of a Euclidean demonstration a drawing of an angle can have as its non-natural meaning that it is (say) obtuse, but that would require, in both cases, that quite different intentions be in play.

[43] We need to distinguish here between the role of Gricean intentions in someone's coming to believe something and the role of such intentions in what is meant, whether or not it is believed. Because proofs in mathematics are self-standing, in the sense explicated in section 2.1, Gricean intentions play no role in one's coming to accept a mathematical proof. They can nonetheless play a role in what is meant. It is only the latter idea that is invoked here.

[44] The terminology is Manders' (1996, 391).

[45] Quoted in Manders (1996, 391).

[46] The use of lines in Books VII–IX for numbers reinforces the point. Had, say, collections of points or strokes been used instead, the generality of the demonstration would have been lost insofar as any collection of points or strokes must involve some number of points or strokes in particular. By using lines for numbers, it can be left unspecified what the unit is that measures the line. The line so understood gives a number but no number in particular. The demonstration is general throughout.

If the figures drawn in a Euclidean diagram have non-natural rather than natural meaning, then they can, by intention, be essentially general.[47] It seems furthermore clear that they would in that case function as icons in Peirce's sense, rather than as symbols or indices, because they would in some way *resemble* that which they signify.[48] But the resemblance is *not*, at least not merely and not directly, a resemblance in appearance. As Peirce (1932, 159) notes, "many diagrams resemble their objects not at all in looks; it is only in respect to the relations of their parts that the likeness consists." We can, then, think of a drawn figure in Euclid as an icon that (though it may also resemble its object in appearance) signifies by way of a resemblance, or similarity, in the relations of parts, that is, in virtue of a homomorphism. On such an account a drawn circle serves as an icon of a geometrical circle not in virtue of any similarity in appearance between the two but because there is a likeness in the relationship of the parts of the drawing, specifically in the relation of the points on the drawn circumference to the drawn center, on the one hand, and the relation of the corresponding parts of the geometrical figure, a circle, on the other.[49]

A drawn circle is roughly circular; it looks like a circle just as a dog looks like a dog. But a dog looks like a dog because it *is* a dog, that is, a particular instance of dog nature. A drawn circle, I have suggested, can look like a circle for *either* of *two* reasons. It can look like a circle for the same reason that a dog looks like a dog, namely, because it is a circle, a particular instance of circle nature. Or it can look like a circle because it is an icon with non-natural meaning that is intended to resemble a circle first and foremost in the relation of its parts. Because what it is an icon of is circle nature, and because what is essential to a circle's being a circle is (as Euclid's definition reveals) that all points on the circumference are equidistant from the center, and it is this relationship of parts that is to be iconically represented, the icon itself comes to look roughly circular. The appearance of circularity is induced in this case by the intended higher order resemblance rather than being something that is there in any case (as circularity is there in any case in a drawing of a particular instance of a circle).

It is a familiar fact that Euclid never mentions any tools to be used in the constructions that are needed for his demonstrations. If drawings in Euclid provided

[47] Saito (2009, 821) makes a similar point using a numerical example, proposition IX.36.

[48] Peircean icons can have either natural or non-natural meaning. In particular, individual instances of geometrical figures can be icons of the relevant sorts of things; they have in that case natural meaning. A drawn circle regarded as an instance of a circle is an icon of a circle that has natural meaning because it so functions independent of anyone's intention that it do so. But a drawn circle can also function as an icon with non-natural meaning. In that case it can be essentially general, an icon of a circle not further specified.

[49] Consider also the fact that in the case of the geometry of three-dimensional objects what is needed is not a perspectival drawing but something more schematic, something "suggesting objective geometrical relations rather than subjective optical impressions" as Netz (1999, 17, n. 24), following Burnyeat, puts it. As Netz also remarks (1999, 18, n. 28), "Greek diagrams are . . . 'graphs' in the mathematical sense [though not in the sense of graph theory, as he notes (1999, 34)]. They are not drawings." Nevertheless, as already indicated, Netz himself does not seem to consider the possibility that the demonstration is constitutively general.

instances of various figures, however, this is just what one would expect.[50] If one wishes to draw a particular straight line then the best way to do that is with a straight-edge, and similarly in the case of a circle: if one wants to draw a particular circle then one is best off using a compass. If, however, one's aim is to draw something with non-natural meaning, in particular, an icon of, say, a line or circle (that is, something essentially general, merely *a line* or *a circle*), then all that matters is that one's drawing is able to produce the desired effect, to convey one's intention. In this case there is no reason at all to mention some particular means of producing the drawing precisely because the drawing, to serve the purpose it is to serve, need not *look* very much like that for which it is an icon. So long as it serves the role it is intended to play then the resemblance (with respect to the relevant relations of parts) is good enough. This is not true of an instance: an instance ought as far as possible to look like what it is. It follows that instances are harder to draw than icons with non-natural meaning, and this is generally true. It is, for example, much easier to draw stick human figures and "smiley" faces, which are of course inherently general, than it is to draw someone in particular, the way some particular person actually looks. Even very small children can do the former; most of us even as adults cannot do the latter very well at all.[51] There is no need for mechanical aids in drawing Euclidean diagrams if those diagrams function as has been suggested here, as essentially general icons with non-natural meaning.

The idea that diagrams are iconic and general by intention can also explain why ancient Greek mathematicians almost never make the sorts of mistakes they would be expected to make if their reasoning were based instead on an instance. If one were reasoning about an instance, particularly about a drawn instance, it would be critical to distinguish carefully between what can and cannot be inferred on the basis of one's consideration of that instance. And this ought, in principle, to be quite difficult. It is, for instance, much harder to learn what a particular person looks like (so as to be able to re-identify that person) from a chance photograph than it is to find this out from, say, a caricature, because in the caricature the work of discovering what are the salient and characteristic features of the person's appearance has already been done. Similarly, if a Euclidean drawing were an instance (with natural meaning), it would be hard to distinguish between what Manders (1996, 392–3) has called co-exact features, ones on the basis of which inferences can be made, and exact features, which have no implications for one's course of reasoning.[52] But in fact, all the evidence suggests that this is not hard at all. As Mueller (1981, 5) remarks regarding the familiar diagram-based "proof" that all triangles are isosceles,

[50] Compare Lachterman (1989, 71).

[51] Vygotski (1978, 112) describes children's drawings as "graphic speech": "The schemes that distinguish children's first drawings are reminiscent in this sense of verbal concepts that communicate only the essential features of objects."

[52] See also the discussion in Greaves (2002, ch. 3).

perhaps a "pupil of Euclid" might stumble on such a proof; but probably he, and certainly an interested mathematician, would have no trouble in figuring out the fallacy on the basis of intuition and figures alone. And in the history of Euclidean geometry no such fallacious arguments are to be found. There are indeed many instances of tacit assumptions being made, but these assumptions were always true. In Euclidean geometry . . . precautions to avoid falsehood are really unnecessary.

This is unsurprising if the figures in the diagram function not as instances but as icons with non-natural meaning that is inherently general. The features that Manders identifies as exact are no more there in the diagram than information about the relative size of human limbs is there in a stick figure.[53]

We noted above that it is in the nature of the objects Euclid defines to have parts in relation and so to be iconically representable. Furthermore, we have seen, such icons can be essentially general, of the concepts themselves, what it is to be, say, a circle, as they contrast with instantiations of those concepts. Drawings of geometrical figures conceived as icons with non-natural meaning can resemble particular instances in appearance in virtue of the fact that there is a resemblance in the relation of the parts. A drawn circle, intended as an icon of a circle, can look very much like a particular instance of a circle. It is easy, then, based on such an appearance, to think that the drawing is such an instance, not such an icon at all. There is good reason nonetheless for thinking that in Euclidean demonstrations the diagram functions iconically (i.e., non-naturally) rather than to provide an instance about which to reason. First, as Manders (1996, 391) has noted, the latter idea "seems incompatible with the use of diagrams in proof by contradiction." To demonstrate, for instance, that a circle does not cut a circle at more than two points, one first sets out in a diagram that two circles do cut at more than two points, say, at four. This is not a situation that can obtain; there is no instance to be drawn. And yet the diagram is drawn—as we will see. If the diagram is read instead as a Peircean icon with non-natural meaning there is no difficulty. What the diagram means non-naturally, the content it exhibits, is exactly what one aims to show is impossible, that a circle cuts a circle at more than two points.

Nor is this the only sort of case in which it is impossible to draw an instance. We know, because Euclid tells us in the opening section of the *Elements*, that a point is that which has no parts, and that a line segment is a length that has no breadth (the extremities of which are points). Such entities are clearly not perceptible; there is nothing that a thing with no parts, or a length with no breadth, looks like. It follows directly that there is no way to draw an instance of either a point or a line. But the concepts of such things, and their relations one to another, can be iconically

[53] Kant offers an interesting variant on the idea in the first *Critique*. According to him, although the geometer constructs an instance (in pure intuition), the construction nonetheless enables a priori knowledge because nothing is ascribed to the figure "except what follows necessarily from what he [the geometer] himself put into it in accordance with its concept" (B xii). Although it is an instance that is drawn, according to Kant, it seems to be an arbitrary one insofar as it has only the properties that are common to all such entities. On the notion of an arbitrary object see Fine (1985).

represented. A drawn line length, for example, can formulate (be intended iconically to display) the content of the concept of a line with endpoints. A drawn dot can formulate (be intended iconically to display) the content of the concept *point*. A drawn circle is, again, a slightly different case because drawn circles do look roughly circular; that is, there is a look that geometrical circles can be said to have. But as in the case of Grice's drawing, the role of a drawn circle in the context of a Euclidean demonstration is not that of an instance but instead that of an icon with non-natural meaning, one that is intended to formulate the content of the concept *circle*, that is, the relation of the points on the circumference to the center, the fact that those points are equidistant from the center (whether or not in the figure as drawn the points on the circumference *look* equidistant from the center). All that is formulated in a Euclidean straight line (conceived as an icon) is a breadthless length lying evenly with the points on itself, that is, a certain relationship between the line and the points that may be found on it; and similarly, all that is formulated in a Euclidean circle (similarly conceived) is the relation of the points on its circumference to the center. In each case, what is important for the cogency of the demonstration is not what the figure looks like but instead what is intended by it whether as set out in Euclid's definitions and postulates, or as stipulated by the particular problem or theorem in question. The similarity in appearance (say, between the icon of a circle and an actual instance of something circular) does help to convey the intended meaning, just as it does in the example of Grice's drawing; nevertheless, what is meant is carried by the intention, not by the similarity in appearance.

A final indication that the object of investigation is not the individual drawn instance, which is a sensory object grasped in perception, is the fact that in the drawing of (say) the diagonal of a square, conceived as a drawing of an instance, the side and the diagonal of a square will not be incommensurable but instead commensurable. That is, as Mueller (1980, 115) has noted, that the diagonal of the square is incommensurable with its side is, in the case of any actually drawn instance, "always disconfirmed by careful measurement." Ancient Greek mathematicians do talk about their diagrams, but as Plato has Socrates remark in *Republic* (510d 7–8), "they do this *for the sake of* the square itself and the diagonal itself," that is, invariant being, however precisely the being of such being is to be understood.[54]

What one draws in Euclidean diagrams are not pictures or instances of geometrical objects but the relations that are constitutive of the various kinds of geometrical entities involved. A Euclidean diagram does not *instantiate* content but instead *formulates* it.[55] Obviously, then, the *ekthesis* does not provide an instance falling under a concept; the letters that it introduces are neither variables of quantification nor names of particular objects. They are nothing more than a means by which the various parts of the diagram can be referred to in the written text. Were the

[54] The translation is Burnyeat's in (1987, 219). [55] See Harari (2003).

demonstration live, that is, actually given by one geometer to another, no letters would be needed. One would in that case merely point as needed now to this bit of the diagram and now to that. The *sumperasma* in which the conclusion is stated is, correlatively, not inferred from a particular case. As Mueller (1974, 42) argues, although "in ancient logic the *sumperasma* is the conclusion of an argument," in Greek geometry, "the *sumperasma* is not so much a result of an inference as a summing up of what has been established." "The word *sumperasma* can … mean 'completion' or 'finish' … the *sumperasma* merely sums up what has taken place in the proof…. It merely completes the proof by summarizing what has been established" (Mueller 1981, 13). The move in a Euclidean demonstration from a claim involving letters to the explicitly general claim is not an application of the rule of universal generalization but instead a move from an implicit generality, such as that a horse is warm-blooded (inferred, say, from the claim that a horse is mammal together with the fact that mammals are warm-blooded), to an explicit generality, say, that all horses are warm-blooded. The demonstration is general throughout.

2.5 Diagrammatic Reasoning in the *Elements*

We have seen that Euclid's system is best thought of on analogy with a system of natural deduction as it contrasts with an axiomatization. And I have argued that figures in Euclidean diagrams do not provide instances but instead should be read as having (Gricean) non-natural meaning and to function, in particular, as icons formulating the contents of the concepts of concern to Euclid. Diagrams so conceived are constitutively general. What remains to be shown is that the reasoning involved in a Euclidean demonstration is not merely diagram-based but instead diagrammatic.

It is clearly true that in a Euclidean demonstration at least some steps in the chain of reasoning are in some way licensed by the diagram. Consider, for example, proposition I.5, that in isosceles triangles the angles at the base are equal to one another, and if the equal straight lines be produced further, the angles under the base will be equal to one another. The diagram is given in Figure 2.6.

It is then argued that because AF is equal to AG and AB to AC, the two sides FA, AC are equal to the two sides GA, AB, and they contain a common angle FAG. This

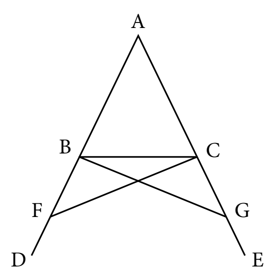


Figure 2.6 The diagram of proposition I.5 of Euclid's *Elements*.

inference, from the two equalities, AF to AG and AB to AC, to the double conclusion, first that the two sides FA, AC are equal to the two sides GA, AB, and also that both those same pairs of sides contain the angle FAG, might be thought to be diagram-based in the following sense. Having, first, drawn an isosceles triangle with sides AB and AC equal, and then having produced the straight lines BD and CE by extending (respectively) AB and AC, and having taken F on BD at random and cut AG off from AE to equal AF, one has thereby inevitably produced not only FA, AC equal to GA, AB, which is already more or less given, but also the angle FAG, which, as inspection of the diagram shows, is identical both to the angle FAC and to the angle GAB. It is no more possible to construct the required diagram without realizing this identity of angles than it is possible to embed one circle in a second circle and the second in a third without thereby embedding the first in the third—which is of course the principle behind Euler diagrams used to exhibit the validity of various valid syllogistic forms of reasoning, here *barbara*. Just as one can infer on the basis of an Euler diagram that some conclusion follows, so one can infer on the basis of the above diagram that the relevant claims are true. The diagram-based inference is, in both cases, valid, that is, truth-preserving, and it is truth-preserving precisely because there exists a homomorphism, a higher level formal identity, that relates relations among elements in the diagram with relations among the entities represented by those elements.

Reasoning using Euler diagrams provides a paradigm of what we might mean by diagram-based inference. In that case, one draws icons of things of a given sort, say, circles to represent collections of objects, and draws them in spatial relations that are homomorphic to relations of inclusion and exclusion over collections. If, for example, all dogs are mammals then one draws the circle representing the collection of dogs inside the circle representing the collection of mammals. If no fish are mammals then one draws the circle representing the collection of fish wholly separate from that representing the collection of mammals, without any overlap. Then one can read off the diagram what the relationship is between the collection of dogs and the collection of fish: the relevant icons are wholly disjoint so one infers that no dogs are fish. The inference, the passage from the premises about dogs, mammals, and fish, to the conclusion about dogs and fish is licensed by the diagram one draws. It is diagram-based. Do diagrams in Euclidean demonstrations function in the same way as the basis for an inference? We will see that although they can, in general they do not.

The first indication that diagrams in Euclid function differently from Euler diagrams is the fact that although Euler diagrams can make something that is implicit in one's premises explicit, they cannot in any way extend one's knowledge. Euclidean demonstrations, by contrast, do seem clearly to be fruitful, real extensions of our knowledge. And the reason they are is connected to the fact that objects can pop up in a Euclidean diagram. We find out that an equilateral triangle can be constructed on a straight line only because a triangle pops up in the course of the construction. Before that point there is nothing whatever about any triangles in anything we have to work with in the demonstration. Quite simply, there is no sense in which an equilateral triangle is *implicit* in a line, even given Euclid's axioms, postulates, and definitions. Nevertheless, as I.1 shows, such a triangle *is* there *potentially*: given what Euclid provides us, we can construct an equilateral triangle on a given straight line. The triangle just pops up when we draw certain lines. And this is true generally in ancient Greek mathematics: one achieves something new largely in virtue of this pop-up feature of its diagrams.[56] To understand the role of diagrams in Euclidean demonstrations, then, we need to understand these pop-up objects.

In a Euclidean demonstration, what is at first taken to be, say, a radius of a circle is later in the demonstration seen as a side of a triangle. But how *could* an icon of one thing become an icon of another? How could an icon of, say, a radius of a circle *turn into* an icon of a side of a triangle? Certainly nothing like this happens in the course of reasoning about an Euler diagram. The iconic significance of the drawn circles is fixed and unchanging throughout the course of reasoning. If the lines one draws in a Euclidean diagram had this same sort of iconic significance, the reasoning would clearly stall because in that case no new points or figures *could* pop up. It is also obvious that nothing could possibly be *both* an icon of (say) a radius of a circle and an icon of a side of a triangle at once, because these are incompatible. An icon of a radius essentially involves reference to a circle; radii are and must be radii of circles. An icon of a side of a triangle makes no reference to a circle. So nothing could at once be an icon of both. But one and the same thing *could serve* now (at time t) as an icon of a radius, and now (at time $t^* \neq t$) as an icon of a side. The familiar duck/rabbit drawing is just such a drawing; it is a drawing that is an icon of a duck (though of course no duck in particular) when viewed in one way and an icon of a rabbit (no one rabbit in particular) when viewed in another.

Some drawings, such as the duck/rabbit, can be seen in more than one way. It would seem, then, that we can say that in the diagram of proposition I.1 one sees certain lines now as icons of radii, as required to determine that they are equal in length, and now as icons of sides of a triangle, as required in order to draw the conclusion that one has drawn an equilateral triangle on a given straight line.[57] The cogency of the reasoning clearly requires both perspectives. But if it does then the Euclidean diagram is functioning in a way that is very different from the way an Euler diagram functions. Much as in the case of the duck/rabbit drawing, and by

[56] What then should we say about, for example, the ancient Greek demonstration that there is no largest prime? Here there is no diagram and there are no pop-up objects, and yet it seems right to say that the demonstration, like properly diagrammatic demonstrations in Euclid, extends our knowledge. As we will eventually see, understanding the ancient demonstration that there is no largest prime, which is a matter of deductive reasoning from concepts, can be understood only at the end of our story, in light of Frege's analysis of the nineteenth-century practice of reasoning from the contents of concepts. Only when mathematics came, in the nineteenth century, to be primarily a practice of deductive reasoning from concepts could the resources be developed that would enable us to understand such a practice.

[57] Euclid's definitions would seem to play a crucial role here: only if one can see in the diagram that the definition is satisfied can one find the relevant object in it.

contrast with an Euler diagram, various collections of lines and points in a Euclidean diagram are icons of, say, circles, or other particular sorts of geometrical figures, *only when viewed a certain way*, only when, as Kant would think of it, the manifold display (or a portion of it) is synthesized under some particular concept, say, that of a circle, or of a triangle. The point is not that the drawn lines underdetermine what is iconically represented, as if they had to be supplemented in some way. It is that the drawing as given, as certain marks on the page, has (intrinsically) the *potential* to be regarded in radically different ways (each of which is fully determinate albeit general, again, as in the duck/rabbit case). It is just this potential that is actualized in the course of reasoning, as one sees lines now as radii and now as sides of a triangle, and which begins to explain the enormous power of Euclidean diagrams as against the relative sterility of Euler diagrams.

In Euclidean diagrams, geometrical entities—points, lines, and figures—pop up as lines are added to the diagram. Cut a line AB with another line CD and up pops a point E as the point of intersection, as well as four new lines AE, BE, CE, and DE. Take the diagonal of a square and up pop two right triangles, and so on. We do not find this surprising in practice; indeed, one need have no self-conscious awareness that it is happening as one follows the course of a Euclidean demonstration. Nevertheless, as we have seen, it clearly *is* happening, and the cogency of the demonstration essentially depends on it. Such pop-up objects, I have suggested, depend in turn on our capacity perceptually to "re-gestalt" various collections of lines, to see them now one way and now another. And what this shows is that it is not the lines themselves that function as icons (even in light of one's intention that they be so regarded) but only the lines *when seen from a particular perspective*, when viewed one way rather than another equally possible way.

Consider again our two crossed lines AB and CD that cut at E, and suppose that they are functioning as icons independent of any perspective taken on them. One could then argue that the point E can belong to only one of the four segments AE, BE, CE, or DE, leaving the other three without an endpoint at one end. This would be a perfectly reasonable inference if we were dealing with a simple icon, one whose significance was fixed independent of any perspective taken on it, because it is true that if you divide a (dense) line by a point then that point can be the endpoint of only one of the two line segments, leaving the other to approach it indefinitely closely (because the line is dense), without having it as its endpoint. Suppose now that we offered this argument to an ancient Greek geometer. How would he respond? He would laugh us away, much as (according to Plato, *Republic* 525e) he laughs away those who would "[attempt] to cut up the 'one'" by pointing out that the line drawn to iconically represent the unit can surely be divided. The geometer laughs because one is in that case taking the diagram the wrong way, because one fails to understand how it works as a diagram in mathematical practice.

When two lines AB and CD cut at E, only one point pops up and it is the endpoint of all four lines AE, BE, CE, and DE, which also just pop up with the cut. And they can do because E is the endpoint of any one of these lines only relative to a way of regarding it, much as certain lines in the duck/rabbit drawing are ears, or a duckbill, only relative to a way of regarding those lines. Another example makes the same point. Suppose that I claimed on the basis of the drawing of a straight line segment ABCD that, contrary to what Euclid holds, two straight lines *can* have a common segment: AC and BD have BC in common. Again the geometer would laugh—and rightly so. Of course one can see the drawn line as iconically representing AC, or BD, or AD, or BC, and indeed one may need so to regard it in the course of a demonstration; but that does not show that the lines themselves have a common segment. This would be required only if the drawn line functioned iconically to represent these possibilities independent of any perspective that was taken on the drawing. Between two points there is only one straight line, not many; two lines can seem to have a common segment (as above) only if one, perversely, mistakes the way the diagram functions.

The lines in a Euclidean diagram, like the lines in a duck/rabbit drawing, function as icons of various sorts of geometrical objects only relative to a perspective that is taken on them. But Euclid's diagrams have a further feature as well, one that is not found either in the duck/rabbit drawing or in an Euler diagram: they have, as Kant already saw, three levels of articulation. Not only are icons of various geometrical figures constructed out of parts. Those icons are combined in turn in larger wholes. At the lowest level, then, are the primitive parts, namely, points, lines, angles, and areas, and their corresponding icons. At the second level are the (concepts of) geometrical objects we are interested in, those that form the subject matter of geometry, all of which are wholes of those primitive parts (and similarly for their icons). At this level we find points as endpoints of lines, as points of intersection of lines, and as centers of circles; we find angles of various sorts that are limited by lines that are also parts of those angles; and we find figures of various sorts. A drawn figure such as, say, a square has as parts: four straight line lengths, four points connecting them, four angles all of which are right, and the area that is bounded by those four lines. Of course, in the figure as actually drawn, the lines will not be truly straight or equal, and they will not meet at a point; the angles will not be right or all equal to one another. But this does not matter because the drawn square is not a picture or instance of a square but instead an icon of a square, one that formulates certain necessary properties of squares. At the third level, finally, is the whole diagram, which is not itself a geometrical figure but within which can be discerned various second-level objects depending on how one configures various collections of drawn lines within the diagram.

A Euclidean diagram is a whole of (intermediate) parts that are themselves wholes of (primitive) parts, and because it has this structure, one can reconceive parts of intermediate wholes in new wholes and discover thereby something new. It is furthermore this feature of the diagram that distinguishes it from simple picture proofs (such as that of the Pythagorean theorem given in section 2.1) and from representations of knots in knot theory and from Roman numerals by which to record how many. A Euclidean diagram is not merely a picture or record of something; instead it formulates the contents of concepts of various geometrical figures in a structured array that enables not merely reasoning *on* the diagram but instead reasoning *in* it. The three levels of articulation enable one to reconfigure parts of intermediate wholes into new (intermediate) wholes in the course of a Euclidean demonstration, and thereby demonstrate significant and often surprising geometrical truths.

We have seen that a Euclidean demonstration comprises a diagram and some text, and in particular, the *kataskeue* and the *apodeixis*. The *kataskeue* provides information about the construction of the diagram and is governed by what can be formulated in the diagram as legitimated by the postulates and any previously demonstrated problems. The *apodeixis*, which is governed by what can be read off the diagram as legitimated by the definitions, common notions, and previously demonstrated theorems, is generally taken to be the proof. This, as we will see in more detail below, is a mistake. The *apodeixis*, that is, that particular bit of the text in a (written) Euclidean demonstration, should be read, on our account, not as the proof but instead as a piece of text that provides instructions regarding how various portions of the constructed diagram are to be read, construed, or analyzed, that is, how they are to be carved up in the course of the demonstration. It is the *diagram* that is the site of reasoning, on our account, not the accompanying text.

Consider one last time the first demonstration in Euclid, that on a given straight line an equilateral triangle can be constructed. The diagram, once again, is shown in Figure 2.7. According to reasoning already rehearsed, we know that AC = AB because A is the center of circle CDB, and that BC = BA because B is the center of circle CAE, hence that AC = BC = AB, on the basis of which it is to be inferred that ABC is an equilateral triangle constructed on the given line AB, which was what we were required to show. Our task concerning this little chain of reasoning is to understand how, based on a claim about radii of circles, one might infer, even given the diagram, something about a triangle. Were the reasoning merely diagram-based, that is, were it reasoning *in* (natural) language but, at least in some cases, *justified by* what is depicted in the diagram then the problem seems intractable. No mere diagram of

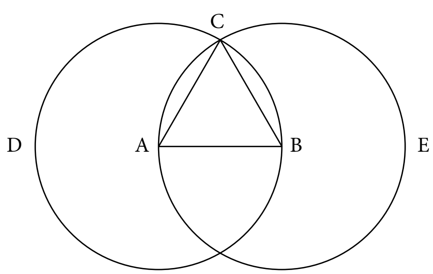


Figure 2.7 The diagram of proposition I.1 of Euclid's *Elements*.

some drawn circles, however iconic, can justify an inference from a claim about the radii of circles to a claim about a triangle. If, on the other hand, we take the diagram not merely as a collection of icons of various geometrical entities but instead as a display with the three levels of articulation outlined above, and so as a collection of lines various parts of which can be configured and reconfigured in a series of steps (as scripted by the *apodeixis*), the solution to our difficulty is obvious. One considers now one part of the diagram in some particular way and now another in some (perhaps incompatible, but perfectly legitimate) way as one makes one's passage to the conclusion.[58] In particular, a line that is at first taken iconically to represent a radius of a circle is later taken iconically to represent a side of a triangle. It is only because the drawing can be regarded in these different ways that one can determine that certain lines are equal in length, because they can be regarded as radii of a single circle, and then *conclude* that a certain triangle is equilateral, because its sides are equal in length. The demonstration is fruitful, a real extension of our knowledge, for just this reason: because we are able perceptually to take a part of one whole and combine it with a part of another whole to form an utterly new and hitherto unavailable whole, we are able to discover something that was simply not there, even implicitly, in the materials with which we began.

Consider now the Euclidean demonstration of proposition II.5: if a straight line be cut into equal and unequal segments, the rectangle contained by the unequal segments of the whole together with the square on the straight line between the points of section is equal to the square on the half.[59] The diagram, which we letter for ease of reference in our written discourse (were the presentation oral we could just point), is shown in Figure 2.8.

To understand this diagram as intended, that is, to know how to "read" it, how to configure its parts (at least to begin with), we need to know the provenance of its various parts, that is, the intention with which it is drawn. We are given that line AB is straight, and that it is cut into equal segments by C and into unequal segments by D. We know by the construction that CEFB is a square on CB, which we know from proposition I.46 is constructible, and that DG is a straight line parallel to CE and BF, KM parallel to AB and EF, and AK parallel to CL and BM, shown to be constructible in proposition I.31. That is, already we see here the various ways the drawn lines are taken to figure in various iconic representations of geometrical figures. CB, for example, is

[58] This idea, that a system of written marks might function to designate various entities only relative to a particular way of regarding it, will emerge again, most notably in our discussion of Frege's language *Begriffsschrift*. We can, then, put the point we have made here about how a Euclidean diagram functions in a demonstration in Fregean terms. Independent of a particular way of regarding collections of them, the drawn points lines, angles, and areas of a Euclidean diagram only express Fregean senses. It is only relative to a way of regarding a collection of them that the various signs designate anything.

[59] Szabó (1978, 334) persuasively argues that, though we might naturally interpret this theorem algebraically (see the second paragraph of Chapter 3), for the Greeks it "is a purely geometrical lemma needed for the solution of a purely geometrical problem . . . Proposition II.14 . . . to construct a square equal to a given rectilineal figure."

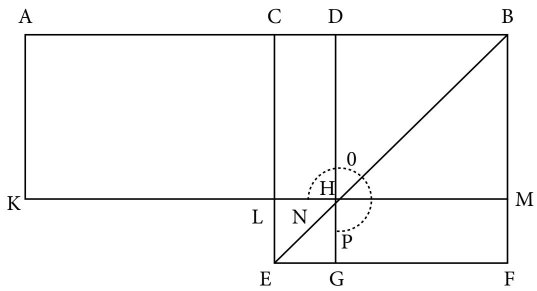
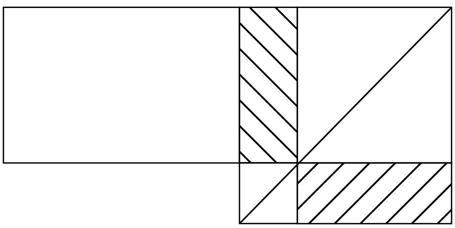


Figure 2.8 The diagram of proposition II.5 of Euclid's *Elements*.

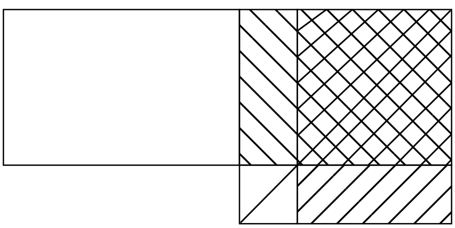
first taken to be the half of line AB, but then as a side of a square. BM is a line length equal in length to DB but also a part of the line BF, which is another side of that same square. It is in virtue of these various relations of parts iconically represented in the drawn diagram, and of these reconfigurings of parts, that the diagram can show that the theorem is true. It does so in a series of six steps scripted by the *apodeixis*.

1.



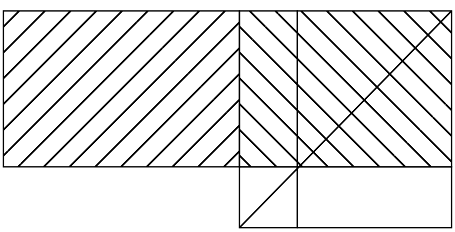
These differently shaded areas are equal, as is shown by prop. I.43: the complements of a parallelogram about the diameter are equal to one another.

2.



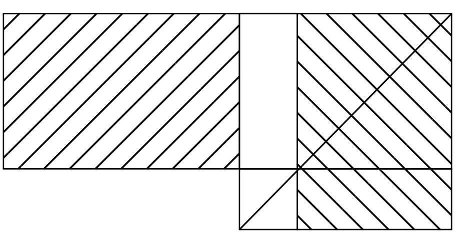
It follows from (1) that these areas are equal, because the same has been added to the same.

3.

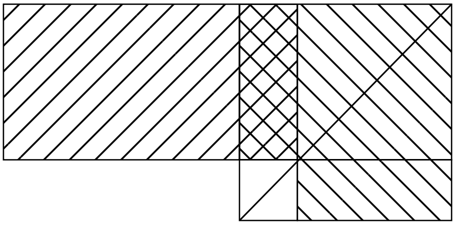


These areas are known to be equal on the basis of what we know about the relationships that obtain among the lines that form the boundaries of the two shaded areas.

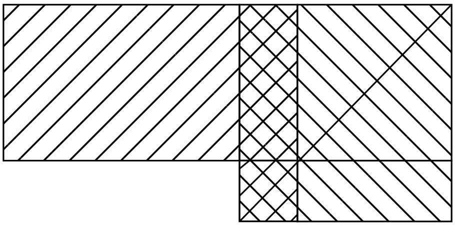
4.



It follows from (2) and (3) that these areas are equal because things equal to the same are equal to each other.

5. 

It follows from (4) that these areas are equal because the same has been added to equals.

6. 

It follows from (5) that these areas are equal because the same has been added to equals.

But that is just what we wanted to show: that the square on the half is equal to the rectangle contained by the unequal segments plus the square on the line between the points of section. By construing various aspects of the diagram in these various ways in the appropriate sequence one comes to see that the theorem is, indeed, must be, true. But if that is right, then (again) it is the diagram that is the site of reasoning in Euclid, not the text.[60] The text of the *apodeixis* is merely a script to guide one's words, and thereby one's thoughts, as one walks oneself through the demonstration in the diagram. Reasoning in Euclid is in this regard quite like doing a calculation in Arabic numeration, say, working out the product of thirty-seven and forty-two. In this latter case, one first writes the two Arabic numerals in a particular array, one directly above the other, then, in a series of familiar steps performs the usual calculation. As one writes down the appropriate numerals in the appropriate places, one may well talk to oneself ('let's see; two times seven is fourteen, so . . .'). Nevertheless, the calculation is not in the words one utters—if one does utter any words, even silently (and one need not)—but in the written notation itself. One calculates *in* Arabic numeration. In much the same sense, one reasons *in* a Euclidean diagram. The demonstration is not in the words of the *apodeixis*; it is in what those words help one to see in the diagram.[61] (This would perhaps be even more obvious if we considered someone

[60] Compare Peirce's remark (1976, 236) that "they [Greek writers] took it for granted that the reader would actively think; and the writer's sentences were to serve merely as so many blazes to enable him to follow the track of that writer's thought."

[61] That the words of the demonstration in Euclid are merely a record of the speech of someone presenting the demonstration to an audience is also indicated by the way the written text is presented in Euclid, and in ancient Greek generally: "unspaced, unpunctuated, unparagraphed, aided by no symbolism related to layout." "Script," in Euclid, "must be transformed into pre-written language" (Netz 1999, 163).

trying to discover a demonstration of some proposition. The inscribing would not be of words but of diagrams.)

As the demonstration of proposition II.5 shows, in order to follow (that is, to understand) a Euclidean demonstration, one must be able to see various drawn lines and points now as parts of one iconic figure and now as parts of another. The drawn lines are assigned very different significances at different stages in one's reasoning; they are *interpreted* differently depending on the context of lines they are taken, at a given stage in one's reasoning, to figure in. It is just this that explains the fruitfulness of a Euclidean demonstration, the fact that its conclusion is, as Kant would say, synthetic a priori. In Euclid, the desired conclusion is contained in the starting point, not merely implicitly, needing only to be made explicit (as in a deductive proof on the standard construal or in an Euler diagram), but instead only potentially. The potential of the starting point to yield the conclusion is made actual only through the construction of the diagram and the course of reasoning in it, that is, by a series of successive refigurings of what it is that is being iconically represented by various parts of the diagram. Parts of wholes must be perceptually taken apart and combined with parts of other wholes to make quite new wholes. And this is possible, first, in virtue of the three levels of articulation in the diagram, and also because the various parts of the diagram signify geometrical objects only relative to ways of regarding those parts. A given line must actually be construed now as an iconic representation of (say) a part of another line and now as an iconic representation of a side of a square, if the demonstration is to succeed. The diagram, more exactly, its proper parts, must be actualized, now as this iconic representation and now as that, through one's construal of them as such representations, if the result is to emerge from what is given. Only a course of constructing and thinking through the diagram can actualize the truth that it potentially contains.[62]

Another example, this time of the reductio proof of III.10, that a circle does not cut a circle at more points than two, reinforces the point. The diagram (again lettered to enable us to explain in the written text what it involves) is shown in Figure 2.15. The following is formulated iconically in the diagram. First, we have by hypothesis (for reductio) that the circle ABC cuts the circle DEF at the points B, G, F, and H; and here (again) it is especially obvious that we do not *picture* the hypothesized situation, which is of course impossible, but instead formulate in the diagram the (conceptual) content of that hypothesis. BH and BG are drawn (licensed by the first postulate) and are then bisected at K and L respectively. (We know that we can do this from I.10.) KC is then drawn at right angles to BH, and LM at right angles to BG. (We know from I.11 that we can do this.) Both KC and LM are extended, KC to A and LM to E, as permitted by the second postulate. What we have then are, first, two straight lines BH and BG, both of which are chords to both circles; that is, we can see BH as drawn

[62] We might, then, translate '*apodeixis*' not as 'proof ' but as 'reasoning'.

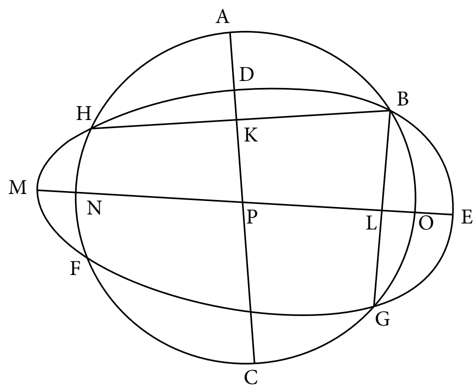
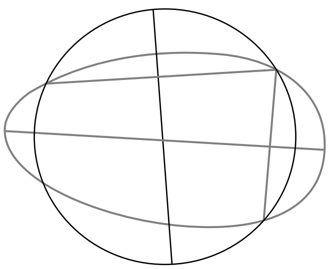


Figure 2.9 The diagram of proposition III.10 of Euclid's *Elements*.

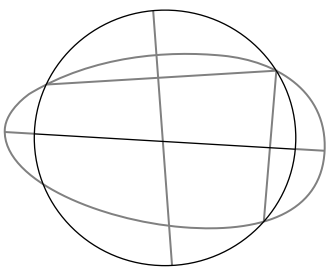
through ABC or as drawn through DEF, and similarly for BG. And we also know that AC bisects HB and is perpendicular to it, and similarly for ME and BG. Now the reasoning begins.

1.



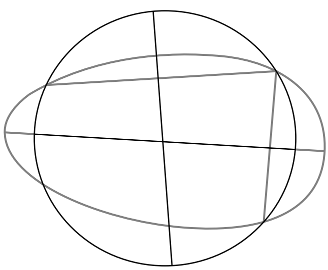
The center of the highlighted circle is on the highlighted bisecting line, by III.1 porism.

2.



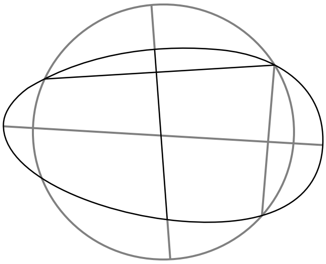
The center of the highlighted circle is on the highlighted bisecting line, again by III.1 porism.

3.



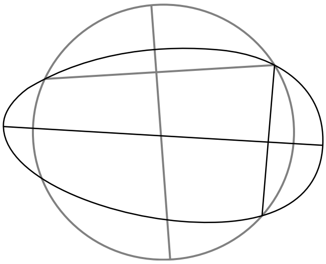
As the diagram shows, the two bisecting lines meet at only one point; so this point must be the center of the circle (because it is the only point that is on both lines).

4.



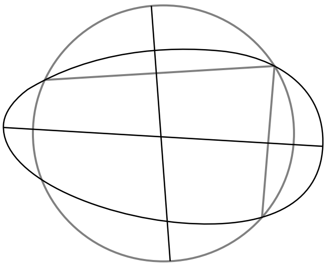
The center of the highlighted circle is on the highlighted bisecting line, by III.1 porism.

5.



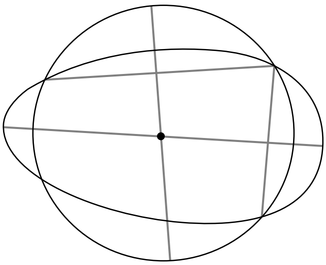
Again, the center of the highlighted circle is on the highlighted bisecting line, by III.1 porism.

6.



The two bisecting lines meet at only one point and so that point must be the center of the circle (because, again, that is the only point on both lines known to contain the center).

7.



The two circles have the same center because one and the same point that was shown first (3) to be the center of the circle ABC was shown also (6) to be the center of DEF.

But we know from III.5 that if two circles cut one another then they will not have the same center. Our hypothesis that a circle cuts a circle at four points is false. Furthermore, because the demonstration involves only three of the four points presumed to cut the two circles, it clearly is cogent no matter how many points one supposes two circles to cut at. Any number greater than two will lead to contradiction. It follows that a circle does not cut a circle at more points than two.

In this demonstration we are concerned with lines, circles, more exactly, the circumferences of circles, and their interrelationships. All that we assume about the various figures is what we are told about them in the construction, and all that we infer at each stage is what we know follows from what we know about the figures that are taken at that stage to be iconically represented in the relevant part of the diagram. By focusing on the various parts of the diagram, now this part and now that, and conceptualizing these parts appropriately, then applying what one knows to the relevant figure thereby iconically represented, one can easily come to realize that, given what is known and assumed, the two circles have one and the same center. But this is absurd given that they cut one another. The assumption that a circle can cut a circle at more than two points must be rejected. It is in just this way that one reasons by reductio in a Euclidean diagram, by formulating in a diagram what one aims to show is false and then showing by a chain of strict and rigorous reasoning in the diagram that that assumption leads to a contradiction.

We have seen that diagrams in Euclid are not merely images or instances of geometrical figures but are instead icons with Gricean non-natural meaning. As such, they are inherently general. But this alone is not sufficient to explain the workings of a Euclidean demonstration. In particular, Euclidean reasoning cannot be merely diagram-based, even where the diagram is conceived iconically, because if it were merely diagram-based, no account could be given of the pop-up objects that are essential to the cogency of Euclidean demonstrations. Euclidean reasoning is instead properly diagrammatic; one reasons *in* the diagram in Euclidean geometry, actualizing at each stage some potential of the diagram. This is furthermore made possible, we have seen, by the nature and structure of the signs that make up that system, and in particular by the fact that collections of signs can be seen now this way and now that within a diagram that has what we have described as three levels of articulation. Because a Euclidean diagram is a collection of primitive signs with three levels of articulation, the parts can be variously construed in systematic ways to mean, iconically represent, now this (sort of) geometrical figure and now that. The drawn diagram, then, is the site of reasoning but does not by itself give the desired result. The steps of the demonstration must actually be taken to realize that result. It is in just this sense that one reasons *in* the diagram in Euclid, that is, through lines, *dia grammon*, just as the ancient Greeks claimed. Not the words but "the construction . . . is the vehicle for the execution of the proof" (Knorr 1975, 74).[63]

2.6 Ancient Greek Philosophy of Mathematics

I have argued that diagrammatic demonstration in Euclid works in virtue of very distinctive features of the notation, in particular, the fact that there are three levels of

[63] See also Knorr's (1975, 72–4) discussion of various sources. As Netz (1999, 36, n. 61) has put this same point, "modern mathematicians prove with axioms; Greek mathematicians proved with lines."

articulation in the drawn diagram that enable various parts of the diagram to be regarded now as an icon of one geometrical figure and now as an icon of another. The diagram formulates content and does so in a way enabling reasoning in the system of signs. This reasoning is furthermore essentially general throughout. The diagram does not instantiate such content in an individual instance, but instead iconically exhibits *what it is to be*, say, a circle or triangle. The content that is formulated in the diagram is *conceptual* content. Nevertheless, because the intentionality of ancient thought is that enabled by natural language, and hence is constitutively object involving, ancient diagrammatic practice was understood by ancient Greek mathematicians and philosophers as similarly object involving. It was assumed that one was arguing about particulars of some sort.

This fact, that the intentionality of ancient Greek thought is constitutively object oriented, explains why we find diagrams even in those cases in which the diagram does no work, for instance, in the demonstration that there is no largest prime. Lines are drawn to be the numbers about which one reasons, and although the reasoning does not rely on the lines one draws, as it does in the cases we have discussed, still the lines are needed to give one objects about which to reason. These objects are not, in intention, particular numbers any more than the triangle one draws for the purposes of a demonstration is some particular triangle, but they are numbers and that about which one reasons. The lines give one a subject matter for one's demonstration. Again, one does not need a setting out in every case (though that is what one would expect if the generality involved in the demonstration were quantificational) but one does need a diagram in every case, even in those cases in which the reasoning is merely reported and is not diagrammatic (as in the demonstration that there is no largest prime), and this is explained by the fundamental object orientation of the intentionality of ancient Greek thought.

Mathematics is a science, and for the shape of spirit we find in the ancient Greeks this means that it is an inquiry into a domain of objects; insofar as it aims at the discovery of truths, mathematics aims at the discovery of truths about objects of various sorts. "All parties to the debate agree that mathematics is true. All parties are therefore committed to accepting that mathematicals exist. The dispute . . . is about their manner of existence" (Burnyeat 1987, 221). Furthermore, as Klein (1968) emphasizes, generality in the ancients' understanding of it is very different from the sort of generality that would come to characterize the essentially symbolic thinking of the early moderns. It is in terms of the *ancient* understanding of generality that we need to understand the philosophical problems they faced regarding the practice of mathematics as they knew it: "The problem of the 'general' applicability of method is therefore for the ancients the problem of the 'generality' (καθόλου) of the mathematical objects themselves, and this problem *they* can solve only on the basis of an ontology of mathematical objects" (Klein 1968, 122–3). The problem that divides Plato and Aristotle concerns the being of mathematical objects, the triangles, lines, and numbers that the mathematician studies. Mathematical truths clearly are not unqualifiedly true of everyday sensible objects. A drawn circle is not perfectly circular; the diagonal of a drawn unit square is not incommensurable with the side; and any physical object that is counted as one, a unit, can be divided. Physical objects do not have the mathematical properties and relations that are of concern to the mathematician, and do have properties, such as mutability and perishability, that proper objects of the understanding do not have. Furthermore, we know this; we know that the drawn circle is imperfectly circular. What is the nature of this knowledge? And what are the objects of mathematical science that will explain it? Are the objects of the science of mathematics ordinary sensory objects conceived in a particular way, as Aristotle thought, or are they peculiar, distinctively intelligible objects grasped with the eye of the mind, as Plato held?

According to Plato, mathematics, although needing to be completed by dialectic, is nonetheless a kind of paradigm of knowledge in large part because it is of what is, timeless and unchanging. It is through the study of mathematics that one's soul is first turned away from sensory things towards intelligible things, through the study of mathematics that one first comes to realize that the ordinary objects of everyday sensory experience and knowing cannot be the only or even the primary objects of knowledge there are. For, Plato argues, the being and intelligibility of the sensory realm of becoming is dependent on the being and intelligibility of the non-sensory realm of being. Our capacity to count sensory things, for instance, can be explained only by our prior and independent grasp of numbers as pure monads (Klein 1968, 71). Socrates in *Republic* tries to tell us a little about the nature of these other objects of knowledge—though of course he cannot say very much. What we learn is that they are intelligible rather than sensory, and that they depend in some way on the Forms. As to the Forms themselves we also know very little. The Form is that in virtue of which things appear similar, and it is by reflecting on the practice of mathematics that we come to expect that Forms are distinct from the sensible things of everyday experience. Drawn circles are roughly circular, none of them perfectly instantiating what it is to be a circle, namely, a plane figure all points on the circumference of which are equidistant from a center. Yet we do know what it is to be a circle, and what we grasp in such a case is, Plato thinks, the Form, Circle itself, something that is exactly what a circle is insofar as it is a circle, nothing more and nothing less.

In the imagery of the Line in *Republic* VI, Plato divides the intelligible realm into two sections. In the first, lower section one proceeds by means of images and assumptions; in the second, higher section one instead proceeds from assumptions to a "beginning or principle that transcends assumptions," making no use of images, "relying on ideas only" (510b). The first is the way of the mathematician, the second the way of the philosopher. That is, as Socrates explains,

> students of geometry and reckoning and such subjects first postulate the odd and the even and the various figures and three kinds of angles and other things akin to these in each branch of science, regard them as known, and, treating them as absolute assumptions, do not deign to

render any further account of them to themselves or others, taking it for granted that they are obvious to everybody. (510d)

And as Klein (1968, 73) points out, this "procedure by 'hypothesis' stressed by Plato is *not* a specifically 'scientific' method but the original attitude of human reflection prior to all science which is revealed directly in speech as it exhibits and judges things." The postulation that is involved in ancient mathematical practice is not the same as that we find in early modern scientific practice but is instead continuous with our everyday understanding, which always already knows a great deal about how things are and how they work.

Plato continues:

They take their start from these, and pursuing the inquiry from this point on consistently, conclude with that for the investigation of which they set out . . . . And . . . they further make use of the visible forms and talk about them, though they are not thinking of them but of those things of which they are a likeness, pursuing their inquiry for the sake of the square as such and the diagonal as such, and not for the sake of the image of it which they draw. And so in all cases. The very things which they mold and draw, which have shadows and images of themselves in water, these things they treat in their turn only as images, but what they really seek is to get sight of those realities which can be seen only by the mind. (510d)

Mathematicians use visible forms and talk about them; that is, they point to various parts of the drawn diagram, to lines, angles, areas, and figures, and say things about them, just as we have seen. But they are not, Plato holds, thinking about the drawn diagram "but of those things of which they are a likeness," of things that are more properly mathematical, things that can be grasped only by the mind's eye. And this notion of the mind's eye, of seeing with the mind, is no mere manner of speaking. As Frede (1987) explains, in Plato's later dialogues the verb '*aisthanesthai*' although usually translated as sense perception, often means more generally awareness of something, perhaps by way of sense perception but also perhaps by the mind, and the use of the verb in the latter case is not metaphorical:

It, rather, seems that all cases of becoming aware are understood and construed along the lines of the paradigm of seeing, exactly because one does not see a radical difference between the way the mind grasps something and the way the eyes see something. Both are supposed to involve some contact with the object by virtue of which, through a mechanism unknown to us, we become aware of it. (Frede 1987, 378)

Again, what we know first and foremost are not, for the ancient Greeks, facts (as would come to seem obvious with the advent of modernity) but objects. To know is to know things as they are, according to their natures, in some cases with the power of sight of the eyes and in others with the power of the sight of the mind.

These things that on Plato's account the mathematician thinks about, grasps with the eyes of the mind, are not themselves Forms. There is, for instance, only one Form Square, but in mathematical practice, we need to appeal to squares that are, however qualitatively similar, nonetheless numerically distinct. The point is perhaps easiest to appreciate in the case of numbers. According to the ancients numbers are collections, for instance, of eggs or boxes, or of pure units in the science of mathematics, collections that can be added to and subtracted from, taken many times, or divided. But for this one needs, for instance, many collections of five (given that five added to five is ten, that five times five is twenty-five, and so on). Each collection of five is five, but the different collections are numerically distinct. The Form is simply Five itself. The Five that is a Form is not a collection of units that can be combined with another number (collection of units) or divided into two and three. It is just Five. Exactly the same is true of triangles, circles, lines, and so on in geometry. To do geometry one needs to be able to call on many different squares, circles, and lines, all of which are ideal instances of Forms, but not themselves Forms. As Aristotle reports in *Metaphysics* I.6: "besides sensible things and Forms he [Plato] says there are the objects of mathematics, which occupy an intermediate position, differing from the sensible things in being eternal and unchangeable, and from Forms in that there are many alike, while the Form itself is in each case unique" (987b 14–18). One simply cannot do mathematics with Forms.[64] Nor, Plato seems to think, is the knowledge that results from mathematical inquiry knowledge about Forms, and for much the same reason. One cannot express mathematical knowledge by appeal only to Forms because, again, often one needs to invoke numerically distinct instances of one and the same Form. What one knows is, for instance, that a circle does not cut a circle at more than two points, and to know this, one might easily think, one needs to have in mind not one but two circles, however arbitrary, and not one but at least two (arbitrary) points.

Although Plato takes both the Forms and the mathematicals, that is, the ideal circles and points that are the objects of mathematical study, to exist separately from sensory objects, to be intelligible rather than sensory objects, Aristotle does not. According to Aristotle, mathematicians work with perceptible things but not as perceptible. Much as the perceptible property of a thing, say, the redness of the apple I see or the sound of the wind chime I hear, has no existence independent of the apple or sounding chime, so the triangularity of the triangle I draw or the fiveness of the fingers on my hand cannot exist independent of the drawn triangle or my hand (Klein 1968, 101). One studies, for instance, the drawn triangle but not *qua* drawn or perceptible; one studies it instead *qua* triangle. That is, as Aristotle explains in *de Anima* III.7:

> The so-called abstract objects the mind thinks just as, if one had thought of the snub-nosed not as snub-nosed but as hollow, one would have thought of an actuality without the flesh in which it is embodied: it is thus that the mind when it is thinking the objects of Mathematics thinks as

[64] See Burnyeat (1987, 229–36).

separate, elements which do not exist separate. In every case the mind which is actively thinking is the objects which it thinks. (431b 13–18)

One talks about (mathematical) triangles as if they were separable, indeed, as if they are separate and unchanging.[65] But in fact, according to Aristotle, they are not. As Burnyeat argues, this position that Aristotle defends is outlined, together with Plato's view, already in *Republic*.

What is at issue is, again, the ontological status of the objects of mathematical inquiry, and in the imagery of the cave, Burnyeat argues, we know both that the puppets on the wall, which are the second thing the prisoner sees (the first being shadows cast by those puppets), and the divine reflections, that is, the images that are the first to be seen outside the cave, are mathematicals.

> But with the divine reflections, the escaped prisoner is apparently aware that he is looking at images of things that he cannot yet see directly (516ab); with the puppets, the opposite is the case (515d). Some awareness of a dependence of mathematicals on Forms is part of the decisive transition from inside to outside the cave. (Burnyeat 1987, 227)

> To rest content with C2 [the puppets] would be to accept, as Aristotle does, that a philosophical account of mathematics need probe no further than the diagram on the page and what the practicing mathematician does with it. To move out of the cave to C3 [the divine reflections] is to come to regard mathematicals no longer as abstractions from the sensible world but as things which exist independent of it. (Burnyeat 1987, 229)

And here we come to an aporia that will dog the philosophy of mathematics for millennia. Plato is right: no adequate account of mathematical *truth* can be achieved unless we recognize that mathematicians in their practice somehow transcend the sensory world of ever-changing things. But so is Aristotle: no adequate account of mathematical *knowledge* can be achieved by talk of mathematicians transcending, in their practice, the sensory world of becoming. The problem of truth and knowledge in mathematics, made familiar recently by Benacerraf (1973), is as old as the practice of mathematics itself. It will finally be resolved only when, through the course of our intellectual history, the power of reason is fully realized as a power of knowing.

2.7 Conclusion

A Euclidean demonstration, as we have come to understand it, is not a proof in the standard sense; it is not a sequence of sentences some of which are premises and the rest of which follow in a sequence of steps that are deductively valid, or diagram-based. Indeed, the demonstration does not lie in *sentences* at all. The demonstration consists in a certain activity, in a course of reasoning that at least in some cases, those

[65] See Mueller (1979), Lear (1982), and Mendell (1998).

we have been concerned with here, is focused directly on a diagram. One does not merely *report* the reasoning, as one does in the demonstration that there is no largest prime, but nor does one display the reasoning as one does in a calculation in Arabic numeration. What is displayed in ancient Greek geometry is a Euclidean diagram that must be reasoned through in a series of steps scripted by the *apodeixis*.

One reasons *in* the diagram rather than merely *on* it in Euclidean diagrammatic practice. And one *can* reason in the diagram because the diagram does not merely picture some objects or some state of affairs, as, we saw, a Roman numeral or picture proof does. Because the diagram formulates the *contents of concepts* in the way that it does, namely, by combining primitive parts into wholes that are themselves parts of the diagram as a whole, various parts of the diagram can be conceived now this way and now that in an ordered series of steps. One does not in this case merely shift one's gaze in order to see explicitly something that was already implicit in the diagram as drawn, as is the case in an Euler or Venn diagram and in some picture proofs. Because what is displayed are the contents of concepts the parts of which can be recombined with parts of other concepts, something new can emerge that was not there even implicitly in that with which one began—though, of course, it had the potential to be discovered. Because parts of different wholes can be perceptually combined in new ways, the demonstration can reveal hitherto unknown relations among the figures of interest to Greek geometry.

Diagrams serve in ancient Greek geometrical practice as a medium of reasoning by enabling the display of the contents of the concepts of various sorts of geometrical figures as they matter to diagrammatic reasoning. To demonstrate a truth, or a construction, in Greek mathematical practice just is to find a diagram, constructed according to the rules set out in the postulates and any previously demonstrated problems, that provides a path from one's starting point to the desired endpoint. To discover such a diagram is to reveal a connection between concepts that is made possible by the definitions, postulates, and common notions that Euclid sets out but is not already there, even if only implicitly, in those definitions, postulates, and common notions. It is only the diagram, itself fully actualized as the diagram it is as one reasons through it, regarding aspects of it now this way and now that as scripted by the *apodeixis*, that actualizes the potential of Euclid's starting points to yield something new.

Proclus claims that problems are preliminary to theorems in geometry, that one needs to know that, say, a triangle can be constructed before one is in a position to know theorems about attributes of triangles. What Proclus does not explain is why that is, what precise role such constructions play in Greek geometrical practice. We now have an answer. Constructions encode or formulate information about the essential natures of geometrical entities in a form that is usable in diagrammatic reasoning. To know how to construct a particular figure is to know how to formulate the content of the concept of that figure, as set out in a definition, in a diagram; to know how to construct figures in Euclidean geometry is to be able to *display* the contents of mathematical concepts in written marks in a way that allows one to reason rigorously in the system of signs. Knowing how to construct something given Euclid's primitive resources is, in other words, quite like knowing how to write a number in Arabic numerals. It is to be able to formulate in a visual display content as it matters to inference. Much as Arabic numeration is a language within which to calculate in arithmetic, so constructions in Euclid provide a language within which to reason in geometry. Only what can be constructed in a Euclidean diagram can figure in demonstrated theorems precisely because and insofar as the diagram is the site of reasoning.

Because a construction displays content in a mathematically tractable way, in a way enabling mathematical reasoning regarding that content, to know how to construct a given figure is to be able to reason about such figures in Euclidean geometry. But that alone does not explain the full significance of constructions in ancient Greek geometry. First, much as an axiomatization sets out a small collection of primitive truths on the basis of which to derive all the other truths in some domain of knowledge, and a small collection of primitive concepts can provide the basis on which to define more complex concepts, so constructions show one how to produce complex geometrical figures on the basis of a few primitive ones, more exactly, how to display the contents of complex concepts given that one knows how to display the contents of primitive ones. And much as in the case of an axiomatization or definition, to have a construction of some complex figure given only a few primitives is to achieve an important kind of systematic knowledge. Furthermore, because these constructions display *what it is to be* a Euclidean geometrical figure, by exhibiting its parts in relation, and knowledge just is, for the ancient Greeks, of things as what they most essentially are, in their natures, to know how to construct some geometrical figure *is* to know it in its nature. It is for precisely this reason that, as Knorr (1983, 139) points out, knowledge of constructions "constitutes in effect what the ancients *mean* by mathematical knowledge."

Constructions are the starting point for reasoning in ancient diagrammatic practice. It is the construction, not the definition, that formulates the contents of the concepts of geometry in a mathematically tractable way, in a way enabling one to reason rigorously in the system of signs to conclusions about those concepts. And the construction can do this because it exhibits in a way that words cannot the natures of the figures involved, what it is to be an equilateral triangle, say, or the bisection of an angle. In sum, the language of Greek mathematical practice is not natural language but constructions—drawn Euclidean diagrams. It is by means of constructions, diagrams, that, as ancient Greek mathematicians realized, it is possible to discover and explore the myriad necessary relationships that obtain among geometrical concepts, from the most obvious to the very subtle. It was an extraordinary achievement, one that would not be surpassed for nearly two millennia.

3

A New World Order

Although it demonstrates a priori, timeless truths winning thereby the title of *mathesis*, ancient mathematical practice operates within the horizon afforded by natural language. The concepts of ancient mathematics are concepts of objects with their characteristic natures in virtue of which they have properties and relate to one another in discoverable ways. But not all concepts of interest to mathematicians are concepts of objects together with their properties and relations; and not all geometrical problems that mathematicians can formulate can be solved diagrammatically. In Descartes' *Geometry* (1637) mathematical practice is transformed through the achievement of a distinctively symbolic language, and the achievement thereby of both a new mode of intentional directedness and a new understanding of the reality we seek to know.[1] Our task is to understand the emergence of this new mathematical practice, the conception of being it embodies, and its relationship to Descartes' radically new metaphysics.[2]

Where an ancient Greek geometer demonstrates by means of a diagram, Descartes computes in the symbolic language of elementary algebra.[3] For example, Euclid uses a diagram to show that if a line be cut into equal and unequal segments then the rectangle contained by the unequal segments of the whole together with the square on the line between the points of section is equal to the square on the half.[4] Descartes, and we following him, would instead approach the problem algebraically. We are given a line AB that is cut into equal segments at C and unequal segments at D as shown in Figure 3.1. We first assign names to the three lengths, say, a to AC, b to CD, and c to DB. We know, then, that $a = b + c$; that is, we interpret the claim that a line is

[1] Here I reverse the order suggested by Klein (1968) insofar as he holds that it was the new mode of intentionality that made possible the new form of mathematics. Where I can and do wholly concur with Klein is in his claim (Klein 1968, 121) that "the nature of the modification which the mathematical science of the sixteenth and seventeenth century brings about in the conceptions of ancient mathematics is *exemplary* for the total design of human knowledge in later times"—or at least, we will see, until the nineteenth century.

[2] The discussion of the significance of Descartes' mathematics in this chapter has been helped not only by Klein (1968) but also by Lachterman (1989). A good introduction to the mathematics of Descartes' *Geometry* can be found in Mancosu (1992).

[3] This is an oversimplification but for present purposes close enough. Later we will see in some detail the nature of Descartes' mathematical practice.

[4] This is proposition II.5 of Euclid's *Elements*. We considered its demonstration in section 2.5.

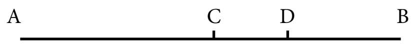


Figure 3.1 A line cut into equal and unequal segments.

cut into equal and unequal segments as a claim about an arithmetical relationship between the lengths of the three segments that are generated by the two cuts. What is to be shown is similarly interpreted. The idea of a rectangle contained by the unequal segments is interpreted as $(a + b)c$; the square on the line between the points of section becomes b^2, and the square on the half is a^2. What is to be shown, then, is that $(a + b)c + b^2 = a^2$, given that $a = b + c$. This is easily done: simply replace all occurrences of 'a' in what is to be shown by '$b + c$', and "do the math," that is, perform the appropriate symbol manipulations until the expressions on both sides of the equal sign are identical.

But how is it that we come to interpret an expression such as "the rectangle contained by the unequal segments" as '$(a + b)c$'? A rectangle is an *object*, a geometrical figure with a nature and a characteristic look. How does such an object come to be transformed into or conceived as something expressible using the language of elementary algebra? As obvious and natural as it may seem to us, this use of symbols was not at all easy to achieve.[5]

As we have noted already, the symbolic languages of mathematics are quite unlike natural languages. Neither narrative nor sensory (at least not in the way natural language is sensory), they are special purpose instruments designed for particular purposes and useless for others. They are not constitutively social and historical, and they have no inherent tendency to change with use. Unlike natural languages, they also can (at least in some cases) be used merely mechanically, without understanding, used, that is, not as languages at all but as useful calculating devices. The positional Arabic numeration system, for example, was used merely as a tool, as a device for solving arithmetical problems, for centuries after its introduction around the tenth century. Although used throughout the first half of the second millennium as an alternative technology to the counting board in the performance of calculations, Arabic numeration was not at first used to record the results of calculations. Records were instead kept in Roman numeration. Only well into the sixteenth century did shopkeepers and merchants begin to keep their records in Arabic numeration; and universities and monasteries continued to use Roman numerals for record keeping even after that (Ball 1889, 7).[6]

[5] Compare this remark of Thom (1971, 74): "geometry is a natural and possibly irreplaceable intermediary between ordinary language and mathematical formalism, where each object is reduced to a symbol and the group of equivalences is reduced to the identity of the written symbol with itself. From this point of view the stage of geometric thought may be a stage that it is impossible to omit in the normal development of man's rational activity."

[6] For further discussion of the use of Roman and Arabic numeration in Europe in the first half of the second millennium see also Swetz (1987) and Chrisomalis (2009, 504).

The significance of this fact, that for centuries both merchants and academics were using paper-and-pencil calculations in Arabic numeration to solve arithmetical problems but would nonetheless record their results in Roman numeration, should not be underestimated. In particular, the fact that even merchants, who surely have little concern for theoretical questions about what numbers (really) are, engaged in such a time-consuming and apparently unmotivated practice, and not merely for a few generations but for centuries, cannot be set aside as a mere historical curiosity. Rather it suggests that they simply could not see Arabic numerals as signs for numbers at all.[7] While it is surely correct to distinguish between everyday practice with numbers, a practice that does not distinguish between the unit, one, and the other numbers, and a philosopher's theoretical account of what numbers "must" be (namely, collections of units, in which case one is not a number),[8] here it is the practice that suggests something fundamental about the pre-modern conception of number. Here it really does seem that there was something they could not do that we can. Why after centuries of recording their results in Roman numeration did the fashion suddenly begin to change in the late sixteenth century? The answer, I think, can only be that the most basic understanding of what numbers are was changing towards the middle of the second millennium, first for practical purposes, for instance, in the introduction of negative numbers in the recording of debts, and finally, after Descartes, in theory as well.[9]

For centuries after its introduction, Arabic numeration was conceived *only* as an instrument, a tool that although useful for calculating results could not be employed in the statement of those results. The signs of Arabic numeration were not conceived as representatives of numbers, as a perspicuous notation of how many. And nor could they be so long as numbers were conceived in terms of the question 'how many?', that is, as collections of units. Similarly, we will see, although François Viète devised a notation suitable for algebraic manipulations, that notation had for him merely instrumental value. It was useful, but not itself a language within which to express mathematical ideas. Only with Descartes would we acquire the eyes to read the symbolism of arithmetic and algebra as itself a language, albeit one of a radically new sort.

Descartes' new mathematical practice and transformed vision both of ourselves as knowers and of the reality known, fundamentally changed who we are. But of course

[7] This may also help to explain the otherwise surprisingly heated debates between abacists and algorismists that persisted even into the seventeenth century (Chrisomalis 2009, 509–10).

[8] As Høyrup (2004, 143) remarks, "if it was necessary to explain so often that unity was no number, then the temptation must have been great to see it as one." "The conceptual otherness that is reflected in the sermons about the nature of number is not caused by any inability to think otherwise; the sermons censure an ever-recurrent tendency to neglect in mathematical practice taboos resulting from philosophical critique" (Høyrup 2004, 144).

[9] The invention of double-entry bookkeeping and the replacement of Roman numeration by Arabic numeration happened almost simultaneously during the fourteenth and fifteenth centuries (Urton 2009, 30–1).

Descartes did not appear out of nowhere. As just indicated, Viète had already developed a notation for basic algebra, and already European culture had become, at least in some respects, "modern." Already nature was coming to be conceived as a kind of clockwork, that is, a mechanical device rather than a living being.[10] These earlier developments, briefly considered in the next two sections, will serve to highlight both the respects in which Descartes continues the tradition he inherits and the extent to which he completely transforms our most fundamental understanding of the being of things in the world and of our cognitive relationship to them.

3.1 The Clockwork Universe

Due in part to the recovery of various classics of ancient Greek philosophy, mathematics, and science, the intellectual culture of Europe was fundamentally transformed over the four centuries preceding Descartes' birth in 1596. With the development of musical notation beginning in the thirteenth century, and the construction of the first clocks, time would, by the fifteenth century, come to be seen as something in its own right, independent of motion, and as itself measurable or quantifiable. The "quantification" of space is similarly manifested in the development of the renaissance conception of what constitutes a "faithful" depiction of things in space, both in the new perspective drawings and paintings, and in the new sorts of maps that began to appear in the fifteenth century. With this quantification of space and time came, finally, the new Galilean conception of motion as at once measurable and subject to laws.[11] What did not change amidst all these developments was our most fundamental conception of being. A grid was being laid over all reality; everything, it was coming to seem, can be measured and counted, or at least ordered, in tables, columns, and graphs. Not living things but clockworks, mechanical devices, were coming to be seen as the paradigm of being. The problem was that no one before Descartes really understood just what that might mean.

In our everyday lives we keep track of where we are both in space and in time. One is in the marketplace or at home, in town or at the seaside; it is morning or night, summer or winter. But although the further idea of measuring spaces, that is, lengths, surfaces, and volumes, is quite natural to us, the further idea of measuring times is not. Time is most naturally understood in terms of motion, for instance, the motion of the sun from its rising to its setting. Because motions can be faster or slower, so, it at first seemed, time can go more or less quickly, the hours themselves expanding and contracting as the seasons change. And the first mechanical clocks did nothing to change this conception. Early clocks were a means of tracking time, the changing

[10] Compare Lachterman's (1989, 125) suggestion that "the advent of radical modernity, at least in its Cartesian figure, might be characterized by this pairing of two tactics for outwitting 'Nature', mechanization and symbolization." See also Mahoney (1998) and Shapin (1996, ch. 1).

[11] And there were other important developments as well. See Crosby (1997), also Foucault (1970).

hours and days, seasons and years, but not yet a mechanism by which to *measure* time. They were nothing more than mechanical reproductions of the motions of the sun and passing of the seasons. The idea of measuring time, as opposed merely to keeping track of it, began to be developed only in the thirteenth century, and only in Europe. And it was made possible at all because musicians got the idea of writing music in addition to performing it.[12]

The earliest musical notation was devised for Gregorian chant, a form of singing that is monophonic and rhythmic without being metrical; in Gregorian chant, all singers sing the same notes at the same time, and although tones can be longer or shorter, they are not measured against any independently fixed temporal unit. There is no measure, no beat, only the longer and shorter tones themselves. As the number of these chants grew, and with them the time needed to learn them, the idea of trying to record the chants suggested itself. Much as written natural language records the sounds speakers make, so the sounds singers make could be recorded, not as in a written language such as English that uses arbitrary marks to stand for sounds, but by the relative placement of marks, higher or lower corresponding to the higher or lower pitch of the notes sung. The result was a two-dimensional display of the notes to be sung according to their pitch along the vertical and in their order along the horizontal. This notation hastened in turn the development of polyphonic music in which different singers sing different notes with different durations, and the development of notations adequate to record this new sort of song.

Monophonic music requires only that all singers keep time with one another; it is not essentially written. Polyphonic music, because it involves different singers singing various notes with different durations, requires that time itself be measured, that is, that a musical notation be developed in which time is tracked independently of the notes that are to be sung. Polyphonic music needs, in other words, a time standard, a unit that is the measure of a duration, and with it a sign for a "rest," that is, the temporal duration within which a singer is to remain silent. Indeed, "the symbols of rests are nothing if not instructions to measure time intervals independently of anything else" (Szamosi 1986, 105). (The symbol for zero is essentially similar, an instruction to conceive numbers independently of counting and things counted—something no one in Europe seemed able to do before the sixteenth century.)

We are so familiar with the idea that time can itself be measured, that it marches on in measured steps, that it can be hard for us to imagine what it was like to live in a world that knows no measured time. We do not in the same way need to use our imaginations to know how things seemed before we achieved the idea of measured space. We have early maps and pictures to show us.

Measuring spaces, the lengths, areas, and volumes of things is as natural as counting, and is common already in various ancient cultures. But this is not yet the

[12] I here follow Szamosi (1986, ch. 5). But see also Crosby (1997, ch. 8).

idea of measuring space itself. So, for instance, we find that thirteenth-century European maps are more like pictures than maps, and peculiar pictures at that. They are "a non-quantificational, non-geometrical attempt to supply information about what was near and what was far—and what was important and what was unimportant" (Crosby 1997, 40; see also Harvey 1993). Pre-Renaissance paintings similarly depict what is more important, say, a deity or an emperor, literally as larger and more central than what is less important, for instance, a peasant. Things are not placed in space in such pictures; they are merely placed. To depict things in space requires a means of depicting the spatial analogue of a musical rest, that is empty space, and only in the fourteenth century did painters begin to consider how to do this. They began to paint not merely spatial things but things in space, that is, as near or far (see Crosby 1997, ch. 9). But the development of perspective painting requires much more than painterly insight. Like the development of a notation for polyphonic music, it requires both a theory (in this case, of optics) and a carefully articulated, step-wise method of depiction.

We say that perspective painting is more realistic than its predecessors, and this is often taken to mean that such paintings depict what one actually sees rather than, say, how what one sees is interpreted. The fact that such works require both a theory of optics and explicitly stated rules governing their production, together (often) with special apparatuses or mechanical aids to help one to see in this way, indicate that this is a mistake. What is true is that such paintings, like fifteenth-century maps, are geometrically accurate; they encode, often with great accuracy, quantitative information about the relative placement and size of things. Much as a fifteenth-century map, unlike earlier maps, accurately depicts relative size and location as if seen from above, so a renaissance perspective painting accurately depicts relative size and location as if seen from in front. In both cases, what is depicted is quantitative information, measurable features of things.

With these developments in the arts came as well the "profoundly unmedieval idea" of progress in human affairs, and self-conscious efforts to cast off old ways in order to develop something new (Crosby 1997, 155). Copernicus's *De revolutionibus* (1543) was to have much the same effect in the sciences. If Copernicus was right, the scholastic Aristotelian conception of the world with its concentric crystalline spheres, at the center of which was the earth, was wrong; and it was shown to be wrong by careful observations, measurements, and reflection. This new account furthermore suggested that the relationship between what we observe and reality is more complex than one might naturally think, that one cannot merely read how things are directly off their appearance to one. By the sixteenth century, it was becoming clear that a general reform of all knowledge and all learning was needed, that a new beginning, a "great instauration," as Bacon would put it, had to be made, one that would "try the whole thing anew upon a better plan, and to commence a total reconstruction of sciences, arts, and all human knowledge raised upon the proper foundations" (Bacon

1620, 66). The idea of modernity, "decisively and *irreversibly* novel in comparison with the past in its entirety," was being born (Lachterman 1989, 126).[13]

But although the Copernican picture of the relative motions of the earth and the heavenly bodies provided, relative to the available data, an intellectually satisfying account, it also posed a serious problem: if the earth moves, why do we not feel it or see it in, say, the displacement of falling objects? Galileo (1564–1642) provided the answer, and thereby solved as well the ancient problem of the motion of projectiles. His use of diagrams in solving these problems is an important precursor to Descartes' use of them.

According to the Aristotelian conception, what we today call motion is only one instance, namely, local motion, of a much wider conception that includes all manner of change. For Aristotle, natural motions, which comprise all the ways a thing changes over time as the sort of thing it is, belong to the essence of a thing. The ripening of fruit, the education of youths, the rolling of a rock down a hill, all are conceived as motions appropriate to the natures involved. They belong to various sorts of things as the kinds of things they are; things move according to their natures. And all such motions naturally come to an end. But there are also unnatural motions, for instance, the motion of something pushed or pulled and the motion of projectiles. The first sort of unnatural motion was easily explained by appeal to the agent doing the pushing or pulling because as soon as the agent stops pushing or pulling the motion itself stops. Projectile motion was much more puzzling. Why do projectiles keep moving for as long as they do given that nothing seems to be making them move after they are released? Aristotle did come up with an account of sorts involving the air as the pusher but it was not very plausible, and in the fifth century John Philoponus suggested instead that when a projectile is set into motion it acquires some kind of motive force or impulse that keeps it moving, at least for a while. (Buridan would suggest something similar in the fourteenth century.) Galileo would finally solve the puzzle and thereby the problem of the moving earth as well.[14]

Careful observation and measurement of a ball rolling down a slightly inclined plane reveals that the ball accelerates at a uniform rate. A ball rolling up such a plane likewise decelerates at a constant rate. It stands to reason, then, that a ball rolling on a horizontal surface will neither accelerate nor decelerate, that is, that it will roll along at the same speed forever. But if so then Aristotle had been asking the wrong question. What needs to be explained is not why projectiles move but why they

[13] This may not be entirely true. According to Klein (1968, 120), the new science "conceives of itself as again taking up and further developing Greek science, i.e., as a recovery and elaboration of 'natural' cognition. It sees itself not only as the science of *nature*, but as '*natural*' science—in opposition to *school* science."

[14] Early intimations of the notion of inertia are found already in the *Mechanical Problems* of pseudo-Aristotle, which Winter (2007) ascribes to Archytas of Tarentum. See also Laird (2001) on the place of the *Mechanical Problems* in Galileo's thinking. Apparently Galileo lectured on the *Mechanical Problems* in Padua in 1598.

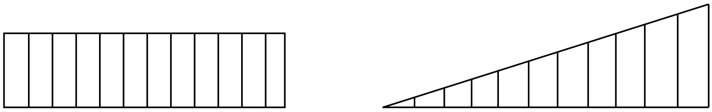


Figure 3.2 Oresme's time-speed graphs of uniform speed (left) and uniform acceleration from rest (right).

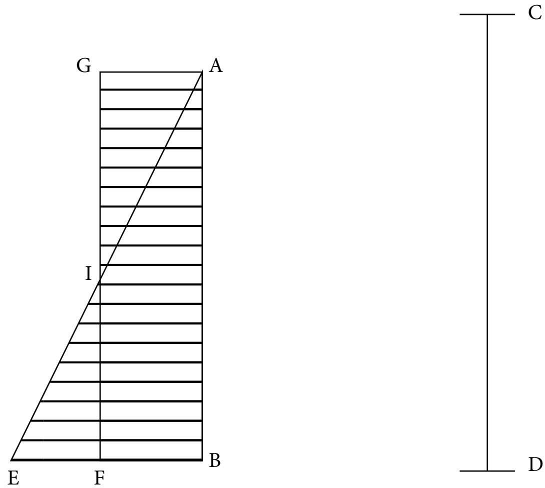
ever stop moving having once been set into motion. And as Galileo saw, if that is the right question to ask then a body must be indifferent to its motion. Motion does not belong to a thing in virtue of its nature, its being as the sort of thing it is; and a body can be moving even at a great speed without that motion being apparent to it. Only changes in motion (or rest now conceived as the limit of motion) need a cause, the force of which can be felt. Furthermore, because a body in motion is indifferent to that motion, different motions can be combined in one moving object, for instance, in a projectile, which combines vertical and horizontal motion. And even an object dropped from some height above the earth forms a kind of projectile: because it is already moving with the earth it continues that motion combined with its downward acceleration until it lands directly below the point at which it had been released—not somewhere behind that point as had seemed to be predicted by the Copernican idea of a moving earth.

Like Copernicus's discovery of the movement of the earth relative to the sun and other heavenly bodies, Galileo's discovery of the law of inertia combined careful observation and measurement, in this case of distance traveled per unit of time, with thoughtful reflection on what those observations and measurements revealed. His analysis of projectile motion involved in addition a very novel use of Euclidean principles. Already in the fourteenth century Oresme had taken the idea of a time-pitch graph from music and applied it to the case of motions. The progression of times was exhibited in a horizontal line and relative speed in a series of vertical lines along the horizontal. A depiction of uniform speed over time thus forms a rectangle while uniform acceleration from rest forms a triangle. (See Figure 3.2.) And other cases similarly generate other shapes. Galileo would combine this basic idea of mapping speed over time with that of measure to realize a radically new form of argument.

To show, for example, that the time taken to traverse a given distance by a body starting at rest and uniformly accelerating is the same as that in which the distance is traversed by a body traveling at a constant speed equal to half the maximum speed achieved by the accelerating body, Galileo appeals to the diagram shown in Figure 3.3.[15] AB represents the time the accelerating body takes to travel the distance

[15] Grosholz (2007, sec. 1.1) provides an insightful discussion of this and other examples from Galileo's *Discorsi*. As she notes (2007, 9), Galileo turns Oresme's time-speed graph on its side so that it can more directly be applied to the case of free fall.

Figure 3.3 Galileo's diagram for Theorem I, proposition I of his *Discorsi*, Third Day.



CD. BE represents the maximum speed achieved by that body, and the lines parallel to BE, that is, the lines drawn from AB to AE, represent the increasing speeds from rest at A to the maximum at BE. F is the midpoint of EB and FG is drawn parallel to AB. The parallelogram AGFB is constructed. Galileo then argues that because the area ABE, which represents the distance traveled by the accelerating body, equals the area AGFB, which represents the distance traveled by the body in uniform motion, the time taken by a body traveling at constant speed half the maximum of a uniformly accelerating one is the same as that taken by the accelerating body to come the same distance.[16] What is crucial here is the fact that although distance traveled is most intuitively represented by a line, such as CD above, Galileo sees that because distance is a function of time and speed at a time, it can also be represented as an area, for instance, as the area of a triangle or of a rectangle. Galileo uses lines and areas not iconically to formulate the contents of concepts as ancient Greek geometers did, but instead to represent or stand in for that which is measurable, such as time, speed, and distance traveled, and the diagram enables him to formulate in a visual display arithmetical relationships among these measurables. The notion of a dimension is thus loosened from its association with the three dimensions of experiential space and is well on its way to becoming, as Descartes thinks of it in the *Regulae*, as "a mode or aspect in respect of which some subject is considered

[16] Notice that in this argument lines serve indifferently as sides of geometrical figures and as representations of arbitrary numbers. Both uses of lines are found already in Euclid's *Elements*, but in Euclid they are scrupulously distinguished and never combined. In Euclid, geometrical lines are utterly different from numbers, despite the fact that numbers can be expressed in lines, as in Books VII–IX. In Euclid, lines (in the relevant demonstrations) do not merely represent or stand for numbers. They *are* numbers, collections of units, albeit arbitrary ones insofar as the unit of measure is not given.

measurable. Thus length, breadth and depth are not the only dimensions of a body: weight too is a dimension—the dimension in terms of which objects are weighed. Speed is a dimension—the dimension of motion; and there are countless other instances of this sort" (CSM I 62; AT X 447[17]).

The idea of tracing in a visual display a phenomenon that is not itself visual is manifested already in written natural language. In the simplest musical notation that same idea recurs but in a form that is essentially two-dimensional and iconic. Instead of a series of arbitrary symbols standing for sounds, musical notation uses spatial position to track not only order but also pitch. Beginning with the simplest time-pitch graphs of Gregorian chants, through those involving a time measure that are needed for writing polyphonic music and Oresme's time-speed graphs, Galileo comes to use Euclidean diagrams and principles together with careful observation and measurement to analyze motion itself. This new form of representation is like a picture, map, or traditional Euclidean diagram, and unlike written natural language, insofar as it signifies a thing (or in the case of Euclid, a content) directly, not as mediated by the sounds we make in talking about it. But like written language, and unlike a picture or map, it depicts something that is not intrinsically visual. We can, of course, see things move, and often we can see as well the traces that can be left by moving things, but motion itself is not visual in the sense that, say, the shape of a thing (that is, parts in spatial relation) is visual. It cannot be pictured and yet, as Galileo realized following Oresme, it can be depicted in a two-dimensional array. Even more important, one can use such arrays to discover relations among the various notions—speed, distance, and time—that are involved in any analysis of motion.

We have seen that for Aristotle the paradigm of a substance, of what is, is a living thing, an organic unity with a characteristic form of life that has the form of a narrative or story with a beginning, a middle, and an end. For the Aristotelian scholastics, reality itself unfolds in a narrative; it takes the form of a story that can be illustrated in pictures but is itself something to be told—spoken and heard. Galileo's work on motion typifies the new sensibility that had developed by the sixteenth century, a sensibility according to which reality is conceived not narratively, as a story to be told, but visually, as something to be mapped in two-dimensional space. To discover how things are, it was coming increasingly to seem, is to turn one's back on the stories we tell, to confront nature directly through careful observation and measurement. Nature, it was coming to seem, does not address us in our natural languages; it does not speak. Rather it is, Galileo suggests, "written in the language of mathematics, and its characters are triangles, circles, and other geometrical figures, without which it is humanly impossible to understand a single word of it; without

[17] All references to Descartes' works, save for the *Geometry*, are to the three volume edition of Cottingham, Stoothoff, and Murdoch (CSM) together with reference to the standard twelve-volume *Oeuvres de Descartes*, edited by Ch. Adam and P. Tannery (AT), revised edition, (Paris: Vrin/C.N.R.S., 1964–76). All references are by volume and page.

these, one is wandering about in a dark labyrinth" (Galilei 1623, 184). Nature is written; it is a book. But unlike any ordinary book, it is not written to communicate something to someone in a natural, primarily aural and narrative, language. It is written in the constitutively visual language of (Euclidean) mathematics.[18] Mathematics, which for the ancient Greeks was irrelevant to the science of nature insofar as mathematics concerns what is unchanging and the science of nature that which changes, has become nature's own language.

The idea that reality is to be mapped and measured in diagrams and graphs rather than told in stories furthermore helps to explain why ancient atomism came now to grip the imagination of not only Galileo but also, for instance, Hobbes, Mersenne, Gassendi, and even, at least for a time, Descartes. What for Aristotle had seemed to be merely two sorts of sensory entities, the proper sensibles such as color, taste, odor, and sound, which can be sensed by only one sense organ, and the common sensibles such as shape, number, and motion, which can be both seen and felt, came now to be conceived as two radically different aspects of reality. And they did so because only Aristotle's common sensibles can be mapped and measured, written in the language of mathematics. On this new conception, the natural world, as contrasted with our experience of it, contains no colors, odors, tastes, or sounds, but only that which has shape, size, number, and motion.

For a sentient animal making its way through its environment, the sensory features of things, their colors, odors, taste, and sounds, are crucial guides to what is to be pursued and avoided. Nor in our own everyday experience of things can such features be separated from their shapes, sizes, number, and motions, except in an act of thought. And yet Galileo suggests that his mind "feels no compulsion" to take the former qualities to be essential to things in the way the latter are. As he writes in *The Assayer*:

> upon conceiving of a material or corporeal substance, I immediately feel the need to conceive simultaneously that it is bounded and has this or that shape; that it is in this place or that at any given time; that it moves or stays still; that it does or does not touch another body; and that it is one, few, or many. I cannot separate it from these conditions by any stretch of my imagination. But that it must be white or red, bitter or sweet, noisy or silent, of sweet or foul odor, my mind feels no compulsion to understand as necessary accompaniments. Indeed, without the senses to guide us, reason or imagination alone would perhaps never arrive at such qualities. For that reason I think that tastes, odors, colors, and so forth are no more than mere names so far as pertains to the subject wherein they reside, and that they have their habitation only in the sensorium. Thus, if the living creature (*l'animale*) were removed, all these qualities would be removed and annihilated. (Galilei 1623, 309)

I do not believe that for exciting in us tastes, odors, and sounds there are required in external bodies anything but sizes, shapes, numbers, and slow or fast movements; and I think that if

[18] Thus early modern thought is sometimes described as distinctively visual, a matter of a new way of seeing. According to Crosby (1997, 132), for example, "the shift to the visual" is the catalyst of the quantification of reality.

ears, tongues, and noses were taken away, shapes and numbers and motions would remain but not odors or tastes or sounds. (Galilei 1623, 311)

The proper sensibles, the color, taste, odor, and sound of a thing, are inconceivable independent of our experiences of them, of what it is like to see, taste, smell, and hear them. This is not true in the same way of the common sensibles insofar as they can be both seen and felt. The common sensibles, that is, the shape, size, number, and motions of things, thus can come to seem to belong to things independent of our perceptual experiences of those things. Only in these cases is there a resemblance between the properties of things and our experience of these properties. Our understanding of mechanical contrivances such as clockworks further reinforces the idea. The workings of a mechanism can be fully understood by appeal only to the shapes, sizes, number, and motions of its parts. Colors, sounds, tastes, and odors have no role to play in an account of the workings of a mechanism. Hence, it was coming to seem, they have no place in nature either but only in our experience of nature. Natural changes were one and all to be reduced to the motions and impacts of atoms.

The new mathematical physics that was being developed by Copernicus, Kepler, Galileo, and others provided an intoxicating vision of precise mathematical descriptions of all the phenomena of nature. But it was as yet only a vision. The idea of modernity, of science raised upon its proper foundation, of a new form of explanation to replace that of Aristotelian substantial forms, remained to be fully realized. The idea of nature as a clockwork, a mere mechanism, needed to be combined with that of a properly symbolic language.

3.2 Viète's Analytical Art

Between the thirteenth and the sixteenth centuries, developments in music, painting, astronomy, and physics had begun radically to alter our understanding of nature. But mathematics too was changing. By the thirteenth century, the algebra of Al Khwarizmi (d. *c.*850) and Omar Khayyam (*c.*1040–1123), which dealt with linear, quadratic, and cubic equations (expressed in natural language), was widely known in Europe; and by the sixteenth century this work had been extended to quartic equations.[19] Various abbreviations and symbols were introduced for powers and other mathematical operations culminating in the literal notation, developed by François Viète (1540–1603), in which two sorts of letters, one for the unknown and the other for the known parameters of a problem, are employed in its solution. Viète is often regarded

[19] It is often assumed that "a single line of development led from the Latin presentations of the subject [that is, of Arabic *al-jabr*] (the translations of al-Khwārizmī and the last part of Fibonacci's *Liber abbaci*) to Luca Pacioli, Cardano, and Tartaglia" (Høyrup 2006, 5). Høyrup (2006) argues, on textual grounds, that a more complex account is needed.

as the first modern mathematician for just this reason.[20] In fact, we will see, Viète's most fundamental orientation remains essentially pre-modern. Only with Descartes is early modern mathematical practice fully realized.

According to a familiar account due to G. H. F. Nesselmann (*Versuch einer kritischen Geschichte der Algebra. Vol. 1: Die Algebra der Griechen* first published in 1842), algebra was at first rhetorical, that is, written in natural language, then syncopated insofar as it began to involve the use of signs or ligatures in place of some words, and finally properly symbolic, all words having been replaced by written signs of some sort. But as Heeffer (2009) argues, there are a number of difficulties, both historical and conceptual, with the account. Most important for our purposes, Nesselmann's account fails to distinguish in any significant way between *symbols* and shorthand notation, that is, abbreviations of various sorts (Heeffer 2009, 5). But as we will see, even this distinction between the use of signs as abbreviations and the use of signs as symbols does not cut finely enough. Viète does not use signs as abbreviations for words but nor, it will be argued, does he use them as symbols in the sense that Descartes does, in accordance with a properly symbolic way of thinking.

We have already noted that although paper-and-pencil calculations using Arabic numeration existed alongside the use of a counting board throughout the first half of the second millennium, they seem to have been understood merely instrumentally. This, I suggested, is unsurprising if numbers were understood as collections of units, and hence as properly expressed not in the positional decimal notation of Arabic numeration but instead in an additive notation such as Roman numeration. Nevertheless, it is remarkable that one can by paper-and-pencil manipulations *work out* the solution to an arithmetical problem that is otherwise worked out either "in one's head," by mental arithmetic, or by moving counters on a counting board. By the fifteenth century, the contrast between "symbolical" (*figuratamente*) and "rhetorical" (*per scrittura*) methods for solving problems had been made explicit (Heeffer 2009, 5–6). This distinction is one we have already appealed to (in the introductory section of Chapter 1). Leaving aside the technology of the counting board, which seems clearly to be neither symbolical nor rhetorical in the relevant sense, one can solve a problem in arithmetic either by way of a paper-and-pencil calculation (*figuratamente*) or by working it out through a chain of reasoning in mental arithmetic, that is, by reflecting on the relevant mathematical ideas and what they entail. Such a chain of reasoning can then be reported in natural language, whether written or spoken (*per scrittura*).

As Heeffer (2009) notes, a sign such as, for example, the familiar '+' for addition can function either as a mere abbreviation for a word of natural language or as a symbol properly speaking, and only the context of use can determine which way it is being used.

[20] See for example, Bos (2001, 154), Mahoney (1980, 144), Bashmakova and Smirnova (1999, 261), and Klein (1968, 183).

When printed in an early fifteenth century arithmetic book, the + sign in '3 + 5 makes 8' would be interpreted as a shorthand for 'and', meaning the addition of five to three. Here, 'plus' describes an operation, a mental or even physical action. There is some temporal element present in the description '3 + 5 makes 8'. First you have three; after adding five, you find out that you have eight. The + sign in this context is thus a direct representation of the action of adding things together. (Heeffer 2009, 11)

In this context the sign is merely shorthand for a description of something to be done on numbers conceived as collections of units that can be combined or put together. Correspondingly, Heeffer claims, properly symbolic thinking is possible not only in symbols but also in words, as, for example, in Pacioli's *Summa de arithmetica geometria proportioni* of 1494 where we find in words various rules of signs, for instance, these: Dividing a positive by a negative produces a negative (*a partire pui per meno neven meno*) and dividing a negative by a positive leads to a negative (*a partire meno per pui neven meno*) (Heeffer 2009, 14). As Heeffer (2009, 15) notes, "the formulation of the rules does not refer to any sorts of quantities, integers, irrational binomials or cossic numbers. The rules only refer to 'the negative' and 'the positive'." Hence, he concludes, "despite the absence of any symbolism, we consider this an early instance of symbolic reasoning." But as in the case of Arabic numeration, we need to ask whether the reasoning involved is conceived merely instrumentally or whether it is taken somehow to embody properly mathematical content.

This question is further complicated, for us, by the fact that we inherit a nineteenth-century tradition that is explicitly formalist in algebra. We read, for example, in an essay by George Peacock published in 1833, that we must distinguish between arithmetical and symbolic algebra.[21] As Peacock explains:

The first of these sciences would be, properly speaking, *universal arithmetic*: its general symbols would represent numbers only; its fundamental operations, and the signs used to denote them, would have the same meaning as in common arithmetic; it would reject the *independent* use of the signs + and −, though it would recognize the common rules of their incorporation, when they were preceded by other quantities or symbols: the operation of subtraction would be *impossible* when the subtrahend was greater than the quantity from which it was required to be taken, and there the proper *impossible* quantities of such a science would be the *negative* quantities of *symbolical* algebra; it would reject also the consideration of the multiple values of simple roots, as well as of the negative and impossible roots of the second and higher degree: it is this species of algebra which alone can be legitimately founded upon arithmetic as its basis.[22]

On this account, arithmetical algebra is a perfectly meaningful science "concerning which questions of truth and falsity are significant"; symbolical algebra is not

[21] According to Heeffer (2009, 8), Peacock's *A Treatise on Algebra* (1830) marks the first use of the term 'symbolic algebra'.

[22] "Report on the Recent Progress and Present State of Certain Branches of Analysis," *Report of the British Association for the Advancement of Science* 3 (1833), p. 189; quoted in Nagel (1935, 180).

meaningful—"questions of truth are meaningless"—but instead merely formal (Nagel 1935, 180). On this "fundamental point", Nagel (1935, 182) furthermore holds, Peacock "saw clearly and truly: the possibility of interpreting extrasystemically the symbols of an algebra [that is, the symbols of a symbolical algebra] in no way affects the validity of equivalences established by means of the formal rules of the combination of symbols which characterize that algebra".[23] But this, I will argue, is not a historically adequate account insofar as in the history of mathematics, and in particular in mathematics as it was developed by Descartes in the seventeenth century, the symbolism of algebra is not merely formal but serves rather to introduce a *new* subject matter for mathematics. Although, as we will soon see, algebraic symbolism is for Viète merely formal in something very like Peacock's (nineteenth-century) sense of formality insofar as it is an uninterpreted system of signs, algebraic symbolism is not at all formal/uninterpreted in Descartes' employment of it.[24]

Viète's Analytical Art comprises three stages. At the first stage, *zetetics*, a problem, whether of arithmetic or geometry, is translated into Viète's newly created symbol system or *logistice speciosa* in the form of an equation. At the second stage, *poristics*, the equation is transformed according to rules until it achieves canonical form. At the third and last stage, *exegetics*, a solution to the problem is found on the basis of the derived equation. As Viète himself emphasizes, at this third stage the analyst turns either geometer, "by executing a true construction," or arithmetician, "solving numerically whatever powers, whether pure or affected, are exhibited" (Viéte 1646, 29). That is, one begins with a problem, either arithmetical or geometrical, that is then formulated in Viète's *logistice speciosa* and operated on according to rules; at the third and final stage, one returns either to arithmetic or to geometry to complete the problem. Viète teaches the Art in eight essays first published individually between 1591 and 1631, then brought together in a single volume, the *Opera Mathematica*, in 1646.[25]

Both in the opening paragraph of the Introduction to *The Analytical Art* and in the Dedication that precedes it, Viète emphasizes the close connections between his art

[23] Nagel immediately continues: "It is this basic doctrine which is at the heart of modern logical theories, and which is the despair of their hostile critics." And this may well be right insofar as both Peacock's work and "modern logical theories" follow in the wake of Kant. But Kant, we need to remember, is responding to a *philosophical* difficulty raised by early modernity, and doing so with a philosophical *theory*. We will not be able so much as to understand the philosophical difficulty to which he is responding if we read back into developments in mathematics *before* Kant philosophical views that Kant developed in response to those difficulties. We will examine in Chapter 6 some of the sources of the formalism of "modern logical theories."

[24] Becoming clearer about this history, and in particular the transformed understanding of the nature and subject matter of mathematical practice that Descartes achieves, may help to explain why, for instance, the identity sign first introduced by Robert Recorde in 1557 "was not universally accepted for another century" (Heeffer 2009, 22). As long as the most fundamental orientation of mathematicians was that of the ancient Greeks, the theory of equations independent of any reference to numbers that might be equated had inevitably to be treated as somewhat anomalous.

[25] The *Opera Mathematica*, edited with notes by Frans van Schooten, was published in Leyden. A facsimile reprint has been issued by Georg Olms Verlag, Hildesheim (1970).

and the work of the Greeks.[26] Three themes from the then newly rediscovered Greek mathematical tradition are especially relevant: Pappus's conception of the analytic method in geometry as outlined in the seventh book of his *Mathematical Collection*,[27] Diophantus's treatment of arithmetical problems using letters for the unknown and for powers of the unknown in his *Arithmetica*,[28] and Eudoxus's general theory of proportions as set out in the fifth book of Euclid's *Elements*.[29] In the spirit of Eudoxus's general theory, Viète aimed to provide a general method for the solution both of the sorts of arithmetical problems Diophantus had considered and of the sorts of geometrical problems Pappus discusses. The method itself was that of analysis, a method Pappus describes as making "the passage from the thing sought, as if it were admitted, through the things which follow in order [from it], to something admitted as the result of synthesis."[30] Diophantus's treatment of arithmetical problems provides an instructive illustration.

Diophantus's *Arithmetica* is a collection of arithmetical problems involving determinate and indeterminate equations together with their (reasoned) solutions. The text is remarkable along a number of dimensions. First, unlike earlier Babylonian and Egyptian texts dealing with similar sorts of problems, Diophantus's *Arithmetica* refers not to numbers of cattle, or sheep, or bushels of grain, but to numbers of pure monads (or unknown numbers of monads, or powers of unknown numbers of monads); and it aims to provide not merely rules for the solution of problems but a demonstration, of a sort, to show why the rule is a good one. The *Arithmetica*, in other words, is a scientific or theoretical work at least as much as it is a practical manual in the art of solving problems. It (or possibly only a later copy[31]) is also remarkable in employing abbreviations—for the unknown and for powers of the unknown (up to the sixth), for the monad, and so on—all of which are explicitly introduced at the beginning of the work, and in providing explicit rules for the transformation of equations (by adding equal terms to both sides and reducing like terms).[32] The letter 'ς' (from *arithmos*, number) is used to signify the unknown, 'Δ^{υ}' (from *dunamis*, power or square) is employed for the square of the unknown, and

[26] As Heeffer (2009, 7) notes, "the humanist project of reviving ancient Greek science and mathematics played a crucial role in the creation of an identity for the European intellectual tradition. Beginning with Regiomontnus' 1464 lecture at Padua, humanist writers distanced themselves from 'barbaric' influences and created the myth that all mathematics, including algebra, descended from the ancient Greeks." As has already been noted, Klein (1968, 120) makes a related point for the case of *natural* science as contrasted with the school science of the Scholastics.

[27] A Latin translation of the *Collection* (fourth century A.D.) appeared in 1588/9. It is likely that Viète had access to the recovered original before then. See Klein (1968, 259, n. 214).

[28] The first six books of the *Arithmetica* (third century A.D.) were rediscovered in 1462 and published in Latin around 1560.

[29] The *Elements* was first printed (in Latin) in 1482.

[30] I here follow Mahoney's (1968, 322) translation.

[31] Heeffer (2009, 3) provides evidence to suggest that this use of signs in the extant text of Diophantus dates back only to the ninth century and not to Diophantus himself in the third. But see also Netz (2012).

[32] See Bashmakova (1997).

'K^{υ}' (from *cubos*) for the cube of the unknown.[33] The fourth power is a square-square, the fifth a square-cube, and the sixth a cube-cube. The monad (unit) is abbreviated 'M^{o}'. Negative numbers are conceived in terms of missing or lacking and are signaled by a special sign (an inverted 'ψ').[34] Diophantus uses Greek alphabetic numerals, and he indicates addition by concatenation.

A simple problem illustrating his analytic method is to divide a given number into two numbers with a given difference. Diophantus turns immediately to a particular instance: the given number is assumed to be, say, one hundred, and the difference forty, units.

Let the less be taken as $\varsigma\alpha$ [one unknown]. Then the greater will be $\varsigma\alpha M^{o}\mu$ [one unknown and forty units]. Then both together become $\varsigma\beta M^{o}\mu$ [two unknowns and forty units]. But they have been given as $M^{o}\rho$ [one hundred units]. $M^{o}\rho$ [one hundred units], then, are equal to $\varsigma\beta M^{o}\mu$ [two unknowns and forty units]. And, taking like things from like: I take $M^{o}\mu$ [forty units] from the ρ [one hundred] and likewise μ [forty] from the β [two] numbers and μ [forty] units. The $\varsigma\beta$ [two unknowns] are left equal to $M^{o}\xi$ [sixty units]. Then, each ς [unknown] becomes $M^{o}\lambda$ [thirty units]. As to the actual numbers required: the less will be $M^{o}\lambda$ [thirty units] and the greater $M^{o}o$ [seventy units], and the proof is clear.[35]

Rather than merely telling us what to do to find the desired answer as his predecessors had done (say: take forty from one hundred to give sixty, then divide by two to give thirty; the two numbers, then, are thirty and thirty plus forty, or seventy), the results of which can then be checked against the original parameters of the problem, Diophantus works the problem out arithmetically. Because he has a sign for the unknown, Diophantus can treat it as if it were known and proceed, through a series of familiar operations, to the answer that is sought. This analytical treatment of arithmetical problems through the introduction of signs for the unknown and its powers, together with Eudoxus's treatment of proportions—where a proportion, according to Viète (1646, 15), is "that from which an equation is composed"—provides Viète with the crucial clues to his Analytical Art.

Although Diophantus's method for solving a problem is clearly meant to be a general one for problems of the relevant type, his solutions are always of particular numerical problems. He has a general method but no means of expressing it in its full generality. Viète resolves the difficulty by appeal to the distinction between vowels and consonants: unknown magnitudes are to be designated by uppercase vowels and given terms by uppercase consonants, all of which are to be operated on as Diophantus operates on his signs for the unknown and its powers.[36] Viète also greatly

[33] As Klein (1968, 146) notes, these are likely merely word abbreviations and ligatures.

[34] On the question of the notion of negative number in play in Diophantus, see Bashmakova (1997, 5–6).

[35] Quoted in Klein (1968, 330–1, n. 22).

[36] Where this difference does not need to be marked, as in setting out the rules governing addition, subtraction, multiplication, and division of "species," Viète uses the two sorts of letters indifferently.

simplifies matters by designating the unknown and its powers not by using a variety of different signs (as Diophantus, and the cossists, had) but by using one sign, a vowel, for the unknown, followed by a word ('*quadratum*', '*cubum*', and so on) to indicate the power of the unknown. More significantly, he also generalizes Diophantus's method to apply not only to the sorts of arithmetical problems Diophantus considers but also to well-known geometrical problems. It is this dimension of generalization, we will see, that provides the key to an adequate understanding of Viète's *logistice speciosa*, his symbolic language or algebra.

Perhaps the first person to recognize the fundamental connection between Euclidean geometry and the new algebra, or art of the coss, was Petrus Ramus (1515–1572), the influential French pedagogue and author of textbooks of mathematics. It was Ramus who first gave the sort of algebraic reading of the *Elements*, in particular, of Books II and VI, that would become standard with Zeuthen and Tannery.[37] But it was Viète who would realize Ramus's ambitions, both mathematical and pedagogical, by showing that algebra, or as he preferred to call it, analysis, provides a general method for the solution of problems whether geometrical or arithmetical. The aim of Viète's Analytical Art, following Ramus, is to teach this method in a pedagogically effective fashion, that is, in a way that will enable students systematically to solve mathematical problems. (See Mahoney 1973, 32–3.)

As already noted, the *logistice speciosa* that Viète introduces in the Analytical Art uses two different sorts of uppercase letters—vowels and consonants—for unknown and known parameters of a problem. The various species (or powers) of unknowns are designated by a vowel followed by a word marking the power to which it is raised, for example, 'A *cubum*' (sometimes 'A *cub.*') or 'E *quadratum*' ('E *quad.*'). It is clear that these expressions are comparable to our 'x^3' and 'y^2' at least in this regard: in Viète's system, if one multiplies, say, A *quadratum* by A, the result is A *cubum*. In such expressions 'A' designates not the value but only the root. Signs for Viète's known parameters, although they too take the form of a letter followed by a word indicating the species (e.g., 'B *plano*' or 'Z *solido*'), do not function in the same way. In a sign such as 'B *plano*' of the *logistice speciosa*, it is the sign 'B' alone that designates the known parameter; '*plano*' merely annotates the letter.[38] As required by the law of homogeneity according to which "homogeneous terms must be compared with homogeneous terms," it serves to remind the analyst that if, at the last stage in solving a problem, he turns geometer (rather than arithmetician), he must put for 'B' something of the appropriate "scale."[39] If the problem is arithmetical, any number can be put for B because among numbers there is no difference in scale,

[37] See Mahoney (1980, 148).

[38] This is widely, if at times only implicitly, recognized. See, for example, Bos (2001, 151), Mahoney (1973, 149), Boyer (1989, 305), and Smith (1925, 449 and 465).

[39] Viète introduces the law of homogeneity in chapter III of Viète (1646) claiming that "much of the fogginess and obscurity of the old analysts is due to their not having been attentive" to it and its consequences.

all being measured by the unit; but if the problem is geometrical then only a plane figure (for instance, a square or a rectangle) can meaningfully be assigned to B. Where Viète wishes to indicate the known parameter raised to a power, say the second, he writes 'B *quad.*'; if he wishes to indicate that a root raised to a power (say, the second) is planar, he writes 'E *plani-quad*'.[40]

Viète's two different sorts of letters, uppercase vowels for unknowns and uppercase consonants for known parameters of a problem, function in his symbolic language in essentially different ways. Vowels signify roots the powers of which are then indicated by the word that follows the vowel. Consonants signify the known parameter itself. The word that follows the consonant (e.g., '*plano*' or '*solido*') serves only to indicate the sort of figure that can be put for the letter at the last stage of the art in the case in which the problem is geometrical. The *logistice speciosa* serves in this way as a symbolic language that can be applied to both arithmetical and geometrical problems. It is, in this regard, quite like Eudoxos's general theory of proportions as developed in Book V of the *Elements*—though, we will see, with one essential difference.

Eudoxos's theory is general in the sense of applying generally to numbers and geometrical figures. It concerns itself not specifically with ratios of numbers or ratios of geometrical figures but more generally with ratios of any sorts of entities that can stand in the relevant relationships; it concerns numbers but not qua numbers because it applies equally to figures and motions, and it concerns figures and motions but not qua figures or motions because it applies equally to numbers. The theory is concerned with such objects insofar as they fall under a "higher universal," one that "has no name" (Aristotle, *Posterior Analytics* I.5). That is, it applies to such objects insofar as they belong to some genus, which has no name, of which *number* and *geometrical figure* are species much as a theory of mammals applies to cats and cows (among other things) insofar as they belong to a genus—one that does have a name, *mammal*—of which *cat* and *cow* are species.

Viète's *logistice speciosa* is not general in the way that Eudoxos's theory of proportions is general. Though it does in a way apply generally to both numbers and geometrical figures, Viète's *logistice speciosa* also "generalizes" over two very different sorts of operations. Whereas in Eudoxos's theory, the notions of ratio and proportion are univocal—precisely the same thing is meant whether it is a ratio or proportion of numbers or of geometrical figures that is being considered—in the Analytical Art, the notions of addition, multiplication, and so on, are not univocal: the arithmetical operations that are applied to numbers in the Analytical Art are essentially different from those applied to geometrical figures. In arithmetic, as Viète understand arithmetic, one calculates with numbers, each calculation taking numbers to yield numbers; in geometry (again, as Viète understand it) one constructs

[40] See, for example, the first of the "Two Treatises on the Understanding and Amendment of Equations," chapter 12. (In Theorem II, 'B *plani-quad*' should be 'E *plani-quad*'; the error is corrected in Witmer's translation (Viète 1646, 192).)

using figures, and in the cases of multiplication and division, and in the geometrical analogue of the taking of roots, the result of a construction is a different sort of figure from that with which one began (or even a different sort of entity altogether, namely, a ratio). Furthermore, in arithmetic the result of an operation can be merely determinable, as it is in the case of the root extraction of, say, two; in geometry, all results are fully determinate. There is in Viète's Analytical Art, by contrast with what we find in Eudoxos's theory, no genus to which numbers and figures belong such that they can be, for instance, added, multiplied, or squared. How, then, are we to read an expression of Viète's *logistice speciosa* such as 'A *quadratum* + B *plano*' given that there is no genus relative to which the mathematical operations (here, addition) can coherently be applied?

For Viète, as for the ancient Greeks, relations depend essentially on the objects that are their relata; there are no relations independent of the objects they relate. It follows that Viète can have no generic notion of an arithmetical operation, say, addition, that serves as the genus, as it were, of which arithmetical and geometrical addition are species. A sign such as '+' in Viète's *logistice speciosa* cannot signify either arithmetical addition or geometrical addition to the exclusion of the other; it cannot be merely equivocal or ambiguous; and there is no genus that might be signified instead. The only plausible reading of Viète's *logistice speciosa* is, then, a reading of it as a formal theory or uninterpreted calculus. The first stage of the Analytical Art, *zetetic*, takes one out of a particular domain of inquiry, either arithmetical or geometrical, into a purely formal system of uninterpreted signs that are to be manipulated according to rules laid out in advance, and only at the last stage, *exegetics*, are the signs again provided an interpretation, either arithmetical or geometrical. As Mahoney (1973, 39) explains,

> the elevation of algebra from a subdiscipline of arithmetic to the art of analysis deprives it of its content at the same time that it extends its applicability. Viète's specious logistic, the system of symbolic expression set forth in the Introduction is, to use modern terms, a language of uninterpreted symbols.

Bos (2001, 148) makes essentially the same point:

> While considering abstract magnitudes Viète could obviously not specify how a multiplication (or any other operation) was actually performed but only how it was symbolically represented. Thereby the "specious" part of the new algebra was indeed a fully abstract formal system implicitly defined by basic assumptions about magnitudes, dimensions, and scales . . . and by axioms concerning the operations . . . [of] addition, subtraction, multiplication, division, root extraction, and the formation of ratios.

Viète's *logistice speciosa* is not, then, a language properly speaking at all. It is an uninterpreted calculus, a tool that is useful for finding solutions to problems but within which (that is, independent of any interpretation that might be given to it) neither problems nor their solutions can be stated. Indeed, its usefulness is a direct function of its being an uninterpreted calculus. Because the *logistice speciosa* has no meaning or content of its own, the results that are derivable in it can be interpreted either arithmetically or geometrically. It is in just this way that, as Viète (1646, 32) proudly announces, "the Analytical Art claims for itself the greatest problem of all, which is

To solve every problem."

Descartes' symbolic language, we will see, is not such an uninterpreted formal system. Despite its superficial similarity to Viète's *logistice speciosa*, Descartes' symbolic language functions in an essentially different way. It is a fully meaningful language in its own right.

3.3 *Mathesis Universalis*

We know that Descartes was inspired to begin working on mathematical and physico-mathematical problems by the Dutchman Isaac Beeckman, whom he first met late in 1618 when Descartes was twenty-two. Already within a few months of that meeting Descartes had the idea of a new science.[41] As Viète had envisaged his Analytical Art, so Descartes envisaged his science as one that "would provide a general solution of all possible equations involving any sort of quantity, whether continuous or discrete, each according to its nature" (CSM III 2; AT X 157). But as indicated already in the unfinished *Regulae*, Descartes has something very different from Viète's Analytical Art in mind. Because, as he has come to think, "the exclusive concern of mathematics is with questions of order and measure and . . . it is irrelevant whether the measure involves numbers, shapes, stars, sounds, or any other object whatever," what he envisages is "a general science which explains all the points that can be raised concerning order and measure irrespective of the subject-matter," a *mathesis universalis* (CSM I 19; AT X 377–8). Whereas traditional mathematics, up to and including Viète's, concerns objects, that is, geometrical figures, kinds of numbers (conceived as collections of units), and so on, Descartes' new mathematics is to be a science of order and measure. What Descartes means, we will see, is that mathematics is now to be conceived as a science not of things, objects, but of the relations and patterns that objects can exhibit, of "the various relations or proportions that hold between these objects" (CSM I 120; AT VI 20).

We saw in section 3.1 that Galileo uses diagrams to exhibit the relationships between time, speed, and distance traveled that are involved in various kinds of motion. Descartes similarly exhibits relationships in diagrams, only now what are exhibited are arithmetical relationships among arbitrary quantities. In fact, what

[41] In a letter dated March 26, 1619, Descartes tells Beeckman that he has just, in the past six days, "discovered four remarkable and completely new demonstrations" and talks of producing "a completely new science" (CSM III 2; AT X 154).

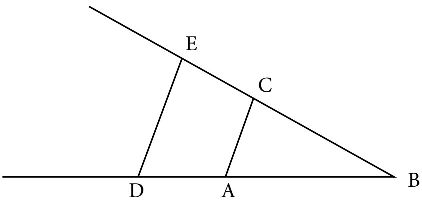


Figure 3.4 A graphic display of the relationship between two numbers and their product.

Descartes does is to employ two very different means of representing the subject matter of this new universal mathematics. In cases in which relations among arbitrary quantities are to be considered separately, they are to be taken, Descartes tells us, "to hold between lines, because I did not find anything simpler, nor anything I could represent more distinctly to my imagination and senses." Where they are to be considered together they are to be designated "by the briefest possible symbols." "In this way," he explains, "I would take over all that is best in geometrical analysis and in algebra, using the one to correct all the defects of the other" (CSM I 121; AT VI 20).[42]

We want, for example, to represent the relationship between two numbers and their product, that is, not two numbers and their product in a relation (of equality) but the relationship itself that holds generally between pairs of numbers and the number that is their product. This can be achieved graphically, as a relation among line lengths, where AB is to be understood as the unit length and DE is drawn parallel to AC (see Figure 3.4).

Because BE is to BD as BC is to BA (because angle B is common to both triangles and the lengths DE and AC are parallel), it follows that BE is the product of BD and BC: if BE:BD = BC:BA, that is, BE/BD = BC/BA, and BA is the unit length, then BE $\times$ 1 = BC $\times$ BD; that is, BE is the product of BC and BD. But, as Descartes goes on, this same relationship can also be expressed symbolically: where a is the length of BD and b that of BC, the product of the two lengths can be given as ab. Descartes claims, in other words, that the geometrical relationship between the line lengths that is presented in the above diagram is not merely *analogous to* but an *alternative expression of* that which is expressed symbolically. It is the relationship itself, which is common to the two means of expression, that is displayed both in the diagram and in symbols. And the same is true of the graphic and symbolic representations in Descartes' geometry of a sum, of the difference between two lengths, of the division of one length by another, and of a square root. Both the graphic display

[42] According to Descartes, the analysis of the ancients "is so closely tied to the examination of figures that it cannot exercise the intellect without greatly tiring the imagination," and the algebra of the moderns (perhaps Viète's in particular?) is "so confined to certain rules and symbols that the end result is a confused and obscure art which encumbers the mind, rather than a science which cultivates it" (CSM I 119–20; AT VI 17–18).

with lines and the symbolic display are a means of mapping or "making visual" the relations and patterns that are of concern in Descartes' universal mathematics.

In Euclid's geometrical practice, we have seen, diagrams formulate content in a way that is mathematically tractable. A drawn circle, for example, iconically presents the relation of parts that is constitutive of a circle. Just the same is true of a drawn triangle, or any drawn figure in Euclid. What is drawn, and what is seen in the drawing, is the content, conceived as parts in relation, of the concept of some sort of geometrical figure. Already in Galileo we find an essentially different conception of a drawing of, say, a triangle. Galileo uses drawn triangles to formulate not the content of the concept of a triangle but instead a relationship among measurable "dimensions" of space, time, and motion. Here we see Descartes doing something similar. Relying on fundamental geometrical features of triangles, and appealing to a given unit, he is able to formulate in a diagram the relationship that holds between two numbers, any two numbers, and their product. Exactly that same relationship is to be expressed symbolically, as *ab*, in the language of arithmetic and algebra as Descartes intends it to be read. What is exhibited both in the diagram and in the symbolic language is not an object (that is, a number or geometrical figure) but a relation that objects can stand in.[43]

In Viète's *logistice speciosa*, one abstracts from differences between geometrical and arithmetical objects. An expression such as 'A *quadratum* + B *plano*' can be interpreted either geometrically (as involving figures classically conceived) or arithmetically (as involving numbers classically conceived as collections of units). Independent of any interpretation, the expression is a mere form. Viéte's *logistice speciosa* functions as an uninterpreted calculus, one that can be interpreted either geometrically or arithmetically. What we have just seen is that in Descartes' geometry an essentially new mathematical notion is introduced, that of a geometrically expressible quantity that is, as numbers are, dimension-free. Just as operations on numbers yield numbers in turn, so operations on line segments in Descartes' geometry yield line segments in turn. That is why there is in Descartes' symbolism nothing corresponding to Viète's annotations '*plano*', '*solido*', and so on. In Descartes' *Geometry* a letter such as 'a' or a combination of signs such as '$(a + b)^2$' is not an uninterpreted expression that can be interpreted either geometrically or arithmetically; it is a representation of an indeterminate line length that in the case of a complex sign displays that length via a display of the relations that it bears to other lengths. Descartes' symbolic language is, then, always already interpreted. As Descartes employs them, letters and combinations of them, signify something in particular, namely, relations among line lengths (themselves conceived as representing arbitrary quantities), either those that are given or those that are sought.

[43] Chapter 1 of Grosholz (1991) also concerns itself with the method of Descartes' geometry, and its relationship to Euclid's geometry. More than is done here, she highlights limitations of Descartes' method.

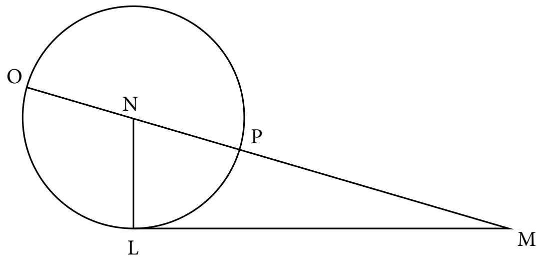


Figure 3.5 Diagram for finding the root of $z^2 = az + b^2$.

We have seen how Descartes represents both diagrammatically and symbolically the relationship of two numbers to their product. Now we need to see how this new form of inquiry focused on such relations can yield new results, for instance, the positive roots of quadratic forms. Suppose, for example, that $z^2 = az + b^2$. (As we do still today, Descartes uses letters from the beginning of the alphabet to represent given indeterminate quantities and letters from the end of the alphabet for the unknowns.) To find the root z we construct a right triangle NLM with LM equal to b and LN equal to $1/2a$, and then prolong MN to O so that NO is equal to NL (see Figure 3.5).

Because OM is to LM as LM is to PM, OM $\times$ PM = LM^2, that is, $z(z - a) = b^2$, or $z^2 - az = b^2$. The diagram thus formulates the relation that is expressed symbolically in '$z^2 = az + b^2$'. The desired root, then, is the length OM, that is, ON + NM, or $1/2a + \sqrt{(a^2 + b^2)}$. As Descartes soon discovered, this result can be generalized using an ingenious compass of Descartes' own devising.

Soon after his first meeting with Beeckman, Descartes began building compasses that would enable a mechanical means of computing the solutions to various problems. One such compass is that shown in Figure 3.6. In the *Geometry* Descartes describes its behavior—that is, the pattern of motions it generates—as follows.

This instrument consists of several rulers hinged together in such a way that YZ being placed along the line AN the angle XYZ can be increased or decreased in size, and when its sides are together the points B, C, D, E, F, G, H all coincide with A; but as the size of the angle is increased, the ruler BC, fastened at right angles to XY at the point B, pushes towards Z the ruler CD which slides along YZ always at right angles. In like manner, CD pushes DE which slides along YX always parallel to BC; DE pushes EF; EF pushes FG; FG pushes GH, and so on. Thus we may imagine an infinity of rulers, each pushing another, half of them making right angles with YX and the rest with YZ. (Descartes 1637, 44, 47)

The compass is constructed so that the series of right triangles CYB, DYC, EYD, FYE, GYF, HYG, and so on are all similar: YB is to YC as YC is to YD, and YD is to YE, and YE is to YF, and so on. That is, YB/YC = YC/YD = YD/YE = YE/YF = YF/YG = YG/YH (and so on). If, now, we take YB to be the unit, and YC to be the unknown, then

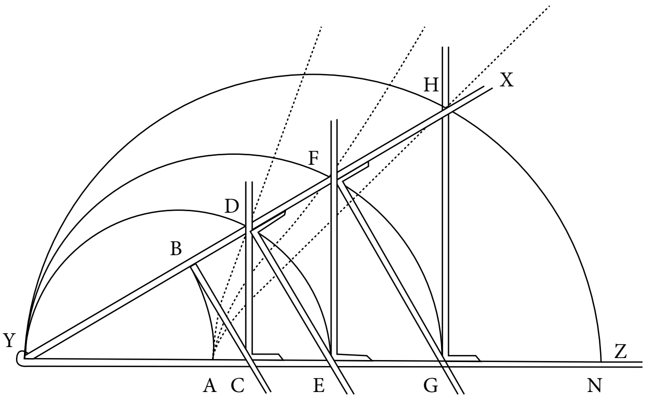


Figure 3.6 Descartes' compass.

this series is precisely the continuing proportional that Descartes discusses in Rule Sixteen of the *Regulae*, namely, $1{:}x = x{:}x^2 = x^2{:}x^3 = x^3{:}x^4 = \ldots$ Given that YE = YC + CE, that is, that $x^3 = x + \text{CE}$, we can use the compass to solve arbitrary equations of the form '$x^3 = x + a$': for a given value of a, if one opens the compass to the point at which the length CE is equal to a (the unit length having been given by YB), then the length YC is the root that is wanted.[44]

Descartes' use of his compass exhibits a number of remarkable features. First, and most obviously, it requires abandoning the traditional understanding of roots, squares, and cubes as, literally, figures. "Squaring" a line in Euclidean geometry literally means making a square, and cubing a line makes a cube. Higher powers make no geometrical sense at all—as is reflected in the fact that there are no names for such higher powers other than, say, "the fourth power" (or as it was sometimes put, unintelligibly, "the square-square"). Descartes' conception of powers in terms of continued proportions replaces the traditional conception in terms of geometrical objects. Even the terminology of squares and cubes is, Descartes thinks, "a source of conceptual confusion and ought to be abandoned completely.... We must note above all that the root, the square, the cube, etc. are nothing but magnitudes in continued proportion which, it is always supposed, are preceded by the arbitrary unit" (CSM I 68; AT X 456–7).[45]

Notice further the generality of the solution that is marked by the use of two different sorts of letters, 'x' for the unknown and 'a' for the known parameter of the problem. Descartes emphasizes the value of such general solutions in Rule Sixteen of the *Regulae*: by using "the letters a, b, c, etc. to express magnitudes already known"

[44] This is one of the cases Descartes mentions in his letter of March 26, 1619 to Beeckman.

[45] By the time he was writing the *Geometry*, Descartes had given up the idea of a reform of language; he there employs the familiar expressions 'square' and 'cube', while making it clear that what he means thereby are the proportional magnitudes.

one abstracts "from numbers . . . or from any matter whatever" (CSM I 67; AT X 455–6). For example, "if the problem is to find the hypotenuse of a right-angled triangle whose sides are 9 and 12, the arithmetician will say that it is $\sqrt{225}$ or 15. We on the other hand will substitute a and b for 9 and 12 and will find the hypotenuse to be $\sqrt{(a^2 + b^2)}$, which keeps distinct the two parts a^2 and b^2 which the numerical expression conflates" (CSM I 67–8; AT X 456). It is in just this way, through the formulation of an expression such as '$\sqrt{(a^2 + b^2)}$', that one can see not only what the answer is but how it depends on the given data. In the general expression of the answer, '$x = \sqrt{(a^2 + b^2)}$', one exhibits the *relationship* of that answer to the given data, the way it relates to what is given.

In his mathematical practice, we have seen, Descartes is not interested in any traditional subject matter but instead in order and measure, that is, in the patterns the objects forming the subject matters of various disciplines exhibit. One such pattern is that displayed by the length of the hypotenuse of a right triangle relative to the lengths of the other two sides, the pattern exhibited in the equation '$x = \sqrt{(a^2 + b^2)}$'. More generally, as is made explicit in Rule Eighteen of the *Regulae*, Descartes is interested not in particular sums and products but instead in the arithmetical operations themselves, addition, subtraction, multiplication, division, and root extraction. That is, he is not merely abstracting from any actual instances of objects of various kinds (as Aristotle suggests mathematicians do). Nor is he dealing with an object that is itself somehow more abstract (as Plato's mathematicals and Forms seem to be). Nor, finally, is he abstracting from content altogether, as Viète does, to consider only a form.[46] Instead the whole discussion is effectively moved up a level, from talk about, or better formulation of, relata, that is, objects such as numbers and figures, to talk about, the formulation of, relations, the patterns such objects can be seen to exhibit. The function of the letters 'a' and 'b' in the equation '$x = \sqrt{(a^2 + b^2)}$' is to enable one to exhibit the precise mathematical relation that holds between the length of the hypotenuse of a right triangle and the lengths of the other two sides. The Cartesian geometer is in this way directed not on independently existing objects, whether sensible or intelligible, but instead on the relations, proportions, and patterns that such objects can display.

Perhaps it will be objected that relations are and must be empty, mere forms, independent of any reference to objects standing in those relations, that relations cannot be conceived antecedently to, and independent of, objects standing in those relations. In one sense this is right. Whatever Descartes' own view of the matter, we cannot *begin* with the conception of symbolic language that Descartes achieves, one

[46] Descartes' mathematics is sometimes read formalistically. See, for example, Gaukroger (1992). But this is an anachronism. As Hatfield (1986, 64) notes, although "the twentieth century mind" is tempted to read Descartes' notation purely formally, "Descartes scorned attempts to make words or symbols and formal rules for manipulating them primary; these are merely arbitrary sensory reminders for the content manifest in thought itself." "Descartes . . . despised formalism" (Hacking 1980, 170).

according to which the signs of the language—letters such as '*x*' and '*y*', and '*a*' and '*b*'. and symbols such as '+' and '='—function together to exhibit relations. Such a language is essentially late, possible at all only through a radical transformation in our primordial understanding as embodied in natural language. But although it is impossible to understand a symbolic language such as Descartes' within which to exhibit relations independent of any objects except against the background of natural language and the understanding of things that it enables, Descartes' language is, in its way, autonomous. That is, it is—or at least can, and did, become—a fully-fledged language embodying an understanding of the world, one that enables, we will see, a radically new mode of consciousness and intentional directedness on reality.

We have seen that between the thirteenth and sixteenth centuries a kind of a grid came to be laid over all reality, in particular, that both space and time came to be conceived as measurable dimensions. We also saw that Galileo had the further idea of using a spatial display not only to exhibit relations among, say, speeds over time (as Oresme had already done) but also to establish truths about the quantities standing in those relations. Descartes completes the thought by taking a visual display of lines to exhibit the relation itself. Descartes sees a drawing of, say, a right triangle, not as an *object* (even one that is somehow essentially general, or arbitrary) but instead as a presentation of one way, an especially interesting and revealing way, that measurable quantities can be *related* to one another. And having achieved this insight, he is able to understand the symbolic language of algebra as an alternative way of exhibiting this and other relations.

A Euclidean diagram appropriately used can reveal the relationship between, say, the squares on the sides of a right triangle. That same relationship, suitably reconceived as a relation among line lengths (themselves conceived as representative of arbitrary magnitudes), is exhibited symbolically as $a^2 + b^2 = c^2$, where c is the length of the hypotenuse. In his mathematical practice Descartes employs four (and only four) such basic reconceptualizations of traditional geometrical relations, all of which are utilized in his solution of the following problem in the *Geometry*.[47] Given the square AD and the line BN, the task is to prolong the side AC to E, so that EF, laid off from E on EB, shall be equal to NB (Descartes 1637, 188). Heraclides' solution, as given in Pappus's *Collection*, is the diagram displayed in Figure 3.7. That is, BD is extended to G, where DG = DN. Taking now the circle whose diameter is BG, it can be shown (by a chain of reasoning that need not concern us) that the point E that is wanted is at the intersection of that circle and AC extended as needed. As Descartes remarks regarding this solution, "those not familiar with this construction would not be likely to discover it."[48] His own approach is

[47] I am indebted to Manders (unpublished) for pointing this out.

[48] Already in the *Regulae*, Descartes complains that the ancient geometers' practice of demonstrating using diagrams, "did not seem to make it sufficiently clear to my mind why these things should be so and how they were discovered" (CSM I 18; AT X 375).

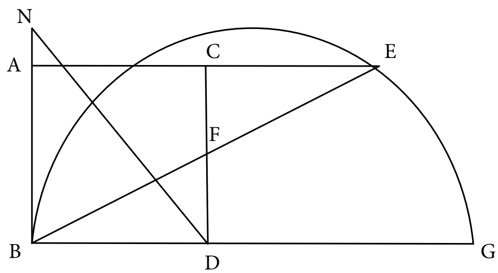


Figure 3.7 Heraclides' solution to the problem that given the square AD and line BN, one is to prolong the side AC to E so that EF, laid off from E on EB, shall be equal to NB.

methodical, a stepwise construction that eventuates in an equation the root of which solves the problem.

We know that BD equals CD because they are sides of one and the same square. Call that length, which is given by the terms of the problem, a. Let c be the length EF, and take the length DF as that which is sought, the unknown x. CF, then, is equal to $a - x$ because, by compositionality, CF + FD = CD. Because the triangles BDF and ECF are similar, we know that BF:DF :: EF:CF, that is, that $a - x{:}c :: x{:}$BF. So, transforming this proportion into an equation, we know that BF = $cx/(a - x)$. But because BDF is right at D, we also know that $BF^2 = BD^2 + FD^2$; that is, $BF^2 = a^2 + x^2$. Combining the two equations thus gives us $(cx/(a - x))^2 = a^2 + x^2$, or $x^4 - 2ax^3 + (2a^2 - c^2)x^2 - 2a^3x + a^4 = 0$. The root of this equation can then be constructed diagrammatically.

In solving this problem, Descartes appeals to four relationships that can be expressed both graphically in a diagram and symbolically: equality (of sides of a square), compositionality (of lines from their parts), proportionality (of triangles), and the relationship that is expressed in the Pythagorean theorem. And as we have just seen, the symbolic expression of these relations enables one to combine, by putting equals for equals, the given information systematically into a single equation and thereby to solve the problem without any need for the additional constructions that are involved in the Euclidean demonstration. In Descartes' geometry, one does not need to discover the diagram that is the medium of a Euclidean demonstration but only to express the given information in symbolic form and then combine it all (by putting equals for equals) into a single, soluble equation. It is just this that enables Descartes to claim that he has a *method* for the discovery of truths in mathematics. Although there is no method by which to discover the diagram that is needed in a Euclidean demonstration, once we conceive the problem symbolically—the diagram not as an iconic display of objects of various sorts in relations but as a presentation of relations and proportions that can equally well be expressed algebraically—Descartes can show us how to combine the given information symbolically in a single equation and thereby to solve the problem.

Before Descartes, although there were mathematical geniuses able to discover the diagrams that are needed to demonstrate various theorems and problems, there was no general method by which to solve such problems. What Descartes discovered, through the use of his symbolic language, was exactly that, a method by which systematically to solve traditional problems of geometry, even some, such as the Pappus problem, that the ancients were unable to solve.[49] As Poincaré would later remark, "before Descartes, only luck or genius allowed one to solve a geometrical problem. After Descartes, one has infallible rules to obtain the result; to be a geometer, it suffices to be patient."[50] Little wonder, then, that Descartes saw in his method a model for all knowing.[51]

3.4 The Order of Things

Descartes' understanding of the sorts of problems Viète considers is essentially different from Viète's own understanding of them insofar as Viète remains oriented on objects, whether geometrical or arithmetical, and Descartes focuses directly on relations that objects can stand in. But there is a further difference between them as well. Whereas Viète systematically ignores "indeterminate" problems, that is, problems in two (or more) unknowns, Descartes does not. Already in his letter to Beeckman of March 26, 1619 he envisages being "able to demonstrate that certain problems involving continuous quantities can be solved only by straight lines or circles, while others can be solved only by means of curves produced by a single motion, such as the curves that can be drawn with the new compasses . . . and others still can be solved only by means of curves generated by distinct independent motions which are surely only imaginary" (CSM III 2–3; AT X 157).[52] But Descartes did not at that point know exactly how this was to go. Although he knew how to use his compass to find the roots of cubic equations, he did not yet know how to describe a curve, conceived in terms of the motion of a point, in an equation in two unknowns. He had the idea of conceiving spaces, geometrical objects, relationally, but not yet the idea of considering space itself as an antecedently given whole of relations. On November 10, 1619 he seems to have discovered precisely that, and thereby the foundations of his marvelous science. From conceiving space in terms of the (measurable) relative locations of objects, Descartes came, through a thoroughgoing figure/

[49] Descartes compares his method to the method of calculating in Arabic numeration in the *Discourse on Method*. Much as "if a child who has been taught arithmetic does a sum following the rules, he can be sure of having found everything the human mind can discover regarding the sum he was considering," so "the method which instructs us to follow the correct order, and to enumerate exactly all the relevant factors, contains everything that gives certainty to the rules of arithmetic" (CSM I 121; AT VI 21).

[50] Henri Poincaré, Preface to his *Oeuvre de Laguerre*, vol. I, quoted in Manders, (unpublished). In fact this is not true in the most interesting cases.

[51] As Klein (1968, 123) notes, although (as we saw) the ancient Greeks faced a problem of the "generality" of its objects in light of its methods, early modern mathematics "determines its objects *by reflecting on the way in which those objects become accessible through a general method*." As we will see in Chapter 4, this thought is fully explicit in Kant's Copernican turn.

[52] In the letter Descartes compares these three cases to arithmetical problems that can be solved by means of, respectively, rational, irrational, and "imaginary" numbers. In the last sort of case, he thinks, there is in fact no solution: we can imagine how to solve such problems but cannot actually solve them.

ground switch, to conceive space as an antecedently given irreducible whole of possible positions, and so as prior to and independent of any objects. It is this conception of space that is the key to understanding and expressing curves as equations.

As already noted in Chapter 1, we (unlike other animals) not only learn various routes through a terrain but also synthesize all the various routes we have learned into one unified conception of the layout of the land as a whole. We learn "to conceive of some very large spaces as integrated wholes rather than piecemeal as they are experienced" (Tversky 2003, 72). Such integrated wholes can be depicted graphically in maps that show various landmarks in their relative locations to one another as if seen from above. Furthermore, because it is achieved by synthesizing into one integrated whole one's procedural knowledge of routes from landmark to landmark, the conception of space that is exemplified in such a map is constitutively object based: "the things in space are fundamental" on such a conception (Tversky 2003, 67). By the seventeenth century, maps showing the relative locations of things were being widely produced for various purposes, and were increasingly being drawn to scale.

And now something remarkable can happen. One can make a gestalt shift, a global figure/ground switch that reveals space as an *antecedently given whole* of possible positions within which objects, landmarks, may but need not be placed, each independent of all the others. Such a view of space, which is not literally a *view* at all, is not object based. On this conception, not objects but space itself "is the foundation" (Tversky 2003, 67). Whereas on the first conception the map is a kind of picture or image of the layout of the land, a kind of "bird's eye" or bottom-up (because object-based) view of it, on this second, top-down, conception, the map records the locations of things in space (each independent of all the others) not as it is experienced in everyday sensory perception but as it is conceived or grasped in thought. Beginning with the piecemeal acquisition of routes, through the integration of those routes into a single whole of the relative locations of things, one in this way achieves finally (through a global figure/ground gestalt shift) a conception of space as an antecedently given whole of possible locations. Although not on the first conception of it, space on this second conception is intelligible prior to and independent of any reference to objects. It is a given whole within which objects can, but need not be, located. This is precisely the conception of space that Descartes achieved.[53]

[53] Høyrup (2004, 129–30) notes that Aristotle considers what Høyrup suggests is the modern view of space (and time) as prior to the things in it (what happens) but "he rejects them in the *Physics* because he cannot make philosophical sense of them" (2004, 130). Even the ancient Greeks, that is to say, could in a way make sense of the idea that space is prior to the things that are found in it (and similarly for time); for, after all, if I take things away, the place for them remains. But *pace* Høyrup, this is not yet the modern conception of space. On the modern conception, space as a whole is prior to and wholly independent of any objects that might be located in space. This conception cannot be achieved simply by imagining things being taken away but requires a thoroughgoing gestalt switch of one's view of a map of things in their relative locations.

Consider again Descartes' compass as shown in Figure 3.6. And imagine now that the angle XYZ is increased by raising the arm XY. In that case, as Descartes explains, "the point B describes the curve AB, which is a circle; while the intersections of the other rulers, namely, the points D, F, H describe other curves, AD, AF, AH, of which the latter are more complex than the first and this more complex than the circle" (1637, 47). The task is to find some means of expressing the curves so described in Descartes' new symbolic language, and it is precisely here that Descartes invokes his essentially new idea of space as an antecedently given whole within which the needed points can be plotted. He envisages his compass, and the movements it makes, in what we know of today as Cartesian space. Given space so conceived together with the (arbitrary) designation of certain imagined lines in that space as "principal lines" (as Descartes calls them) to which all others are to be referred, it is easy to find the equation governing the motion of the curve AB. This motion describes a circle whose radius is r = YA = YB. Dropping a perpendicular from the point B to the line AN (one of the principal lines), at a point we can label W, and letting that perpendicular BW = y and YW = x, the motion of the point B is along the path expressed in the equation $x^2 + y^2 = r^2$ in two unknowns. A circle, which had seemed to the ancients to be a certain plane figure, a particular sort of two-dimensional object with its characteristic nature, is now to be understood by appeal to the lawful relation holding between the two unknowns, x and y, as expressed in the equation that governs the motion of a point in Cartesian space.

Now we need an equation for the curve described by the movement of point D. Notice first that YC and CD together determine the point D because $YC^2 + CD^2 = YD^2$. So we let YC = x, and CD = y. What we want is an equation determining the relationship of the two values, y to x. Let YA = YB = a. Because we know that YD is to YC as YC is to YB, that is, that YD:YC = YC:YB, or $YD/x = x/a$, we know also that $YD = x^2/a$. But we know as well that $YD^2 = x^2 + y^2$ because triangle YCD is right at C and YC = x and CD = y. It then follows that $(x^2/a)^2 = x^2 + y^2$, from which it follows that $x^4 = a^2(x^2 + y^2)$. We have the desired equation for the curve AD, the equation expressing the relationship between the lengths x and y that determines the point D at any moment on its trajectory as the compass is opened.

To find the equation for the curve traced out by the moving point F is a little more complicated, but essentially similar. Here we let YE = x and EF = y. Again YA = YB = a. We know that YF:x = x:YD. So $YD = x^2/YF$. But we also know that x:YD = YD: YC. So by substitution, $x: x^2/YF = x^2/YF: YC$, or $YC = (x^2/YF)^2/x = x^3/YF^2$. But we also know that YD:YC = YC:a. In other words, $x^2/YF: x^3/YF^2 = x^3/YF^2: a$, or $ax^2/YF = (x^3/YF^2)^2$. This, by familiar algebraic operations, yields $YF = {}^3\sqrt{(x^4/a)}$.[54] But $YF^2 = x^2 + y^2$; so ${}^3\sqrt{(x^8/a^2)} = x^2 + y^2$, or $x^8 = a^2(x^2 + y^2)^3$. An exactly analogous chain of reasoning yields the equation for the path of point H, and any further points one

54 We begin with $ax^2/YF = (x^3/YF^2)^2 = x^6/YF^4$, from which it follows that $ax^2 = x^6/YF^3$. So $YF^3 = x^6/ax^2 = x^4/a$. So, $YF = {}^3\sqrt{(x^4/a)}$.

cares to consider as further rulers are added to the compass. In every case, as Descartes says (1637, 47), the description of the curve can be conceived "as clearly and distinctly as that of the circle, or at least that of the conic sections." Much as the Arabic numeration system, by contrast with (say) Roman numeration, enables one to solve arithmetical problems of arbitrary complexity simply by following rules in the proper order, so here again we see that Descartes' new way of approaching problems in geometry, by contrast with that of Euclidean geometers, enables one to solve geometrical problems of arbitrary complexity simply by following rules in the proper order. And the method clearly generalizes to problems in more than two unknowns. It provides just the understanding that enables Descartes to claim, in the opening sentence of the *Geometry* (1637, 2), that "any problem in geometry can easily be reduced to such terms that a knowledge of the length of certain straight lines is sufficient for its construction."

In ancient Greek mathematics, problems are classified according to the sorts of objects that need to be invoked to solve them. Plane problems are those that require recourse only to lines and circles, conceived as self-subsistent geometrical objects, in their solution. Next are solid problems, so called because their solutions require appeal also to "conics," that is, to hyperbolas, ellipses, and parabolas, which are understood as intersections of a cone (that is, a solid, a figure having length, breadth, and depth, the limit of which is a surface) and a plane. Conics, on this view, are plane figures and so in that respect like circles, but unlike circles they are intelligible only by reference to a solid figure, a cone, and are for this reason an essentially different sort of plane figure. The last sorts of problems are the so-called line problems that require appeal to "lines"—that is, loci of points to which no known figures, or boundaries of figures, correspond—in their solutions. Because such lines do not form the boundaries of any figures, they were regarded with deep suspicion by ancient geometers; they were taken to be mechanical rather than properly geometrical.

In Book III of his *Geometry*, Descartes continues this traditional line of thought; he classifies traditional determinate construction problems by the ancient criteria. In Book II, however, a new classification is given, a classification not of problems but of curves. According to this classificatory scheme, the simplest class of geometrical curves includes the circle, the parabola, the hyperbola, and the ellipse because all these curves are given by equations the highest term of which is either a product of the two unknowns or the square of one: they are one and all expressible in an equation of the form: $ax^2 + by^2 + cxy + dx + ey + f = 0$.[55] Whereas on the ancient view, conics are essentially different from circles, from Descartes' perspective in

[55] Unsurprisingly, Descartes also rejects the traditional distinction between mechanical and properly geometrical curves. According to him, mechanical curves are those that "must be conceived of as described by two separate movements whose relation does not admit of exact determination" (1637, 44). Because the relationship between the two separate movements cannot be exactly determined, Descartes thinks, the curves themselves cannot be exactly and completely known.

Book II, they are essentially alike. And they can be seen to be alike because a geometrical curve, from being conceived as the boundary of a figure, is now to be conceived instead as a curve all points of which "must bear a definite relation to all points of a straight line," where this relation in turn "must be expressed by means of a single equation" (Descartes 1637, 48). As Descartes conceives it, a curve is not an edge of a thing but instead the locus of points in space, where space is conceived in turn as an antecedently given whole of possible positions. Such a curve can be traced by a continuous motion but it is not constituted by the relative positions of points on it; it is constituted by the relationship expressed in an equation, that is, by the location of each point, independent of all the others, directly in space (relative to some arbitrarily given principal lines). From the perspective that Descartes provides, a problem such as that posed by the sort of quadratic equation that was the focus of Viète's interest, can be thought of as a limit case, with one unknown set equal to zero, of an indeterminate equation, for example, $x^2 + ax = b^2$ as the limit case of $x^2 + ax - b^2 = y$, that in which y is set equal to zero. As Descartes writes to Mersenne regarding the account of curves he develops in Book II of the *Geometry*, it is "as far removed from ordinary geometry, as the rhetoric of Cicero is from a child's ABC" (CSM III 78; AT I 479).

Descartes' new universal mathematics was to be a science of order and measure, and we saw in section 3.3 the ways Descartes is able to exhibit various relations of measure, for instance, that of two quantities to their product. But that is not yet to have a science of order. It is only in Descartes' account of curves conceived as the paths of points moving through Cartesian space, paths that can be expressed in equations in two unknowns, that we see how this new science is to be a science of order. Already in Rule Ten of the *Regulae*, Descartes enjoins us "methodically [to] survey even the most insignificant products of human skill, especially those that display or presuppose order" (CSM I 34–5; AT X 403). And as is explained in the ensuing discussion, "the simplest and least exalted arts . . . in which order prevails" are "weaving and carpet-making, or the more feminine arts of embroidery, in which threads are interwoven in an infinitely varied pattern." These arts, together with "number games and any games involving arithmetic," "present us in the most distinct way with innumerable instances of order, each one different from the other, yet all regular." "Human discernment," the discussion concludes, "consists almost entirely in the proper observance of such order." In the *Geometry* Descartes provides us the means to understand such order mathematically.

Something exhibits order, on Descartes' account, if it exhibits a pattern not of meaning, significance, or fittingness, but as the expression of a rule. As the pattern discernible in a woven cloth, by contrast with that discernible in, say, a free-hand drawing, is created by following a rule governing how at each stage the various threads are to be woven together, so order more generally is achieved by following a rule. The path traced by a moving point according to the rule expressed in an equation in two unknowns provides the paradigm of this conception of order.

The equation gives the rule that governs the movement of the point and thereby underlies the pattern that is exhibited in the curve that is described, or marked out, by the moving point.

Much as Descartes considers relations independent of any objects that might stand in those relations, so in his account of curves, he considers the laws governing the motions of points independent of those points and the visible traces they generate. Such a law is not *in* the objects it governs as a principle of their motion; and it is not a generalization about objects derived from some insight into those objects. The law is *independent* of the objects it governs, something in its own right that can be expressed, using two or more unknowns, in the symbolic language of algebra; and it is grasped in pure thought.[56] The law in this way both underlies and explains the appearance of, say, circles, their characteristic symmetries. Circles, that is to say, are not for Descartes, and for modern mathematics more generally, sensory objects, that is, objects with a characteristic look and feel, as Aristotle had claimed and ancient mathematicians assumed. They are instead (so Descartes thinks) *purely intelligible* objects, objects whose essences are given by equations that are grasped by the pure intellect independent of any images, and indeed of our sense organs generally. No matter what kind of body one has, even if one has no body at all, one can (Descartes holds) grasp the essence of a circle as Descartes conceives it by grasping the law that underlies (and explains) its visual representation.

As should be clear, this conception of order as what is subject to a law is radically different from the ancient conception of order. The ancient conception of order is in terms of what is fitting or appropriate, the relative significances of things; it is hierarchical and teleological. Furthermore, on the ancient view, to come to what Taylor calls self-presence just is to grasp this order. "On this [the ancient] view the notion of a subject coming to self-presence and clarity in the absence of any cosmic order, or in ignorance of and unrelated to the cosmic order, is utterly senseless: to rise out of dream, confusion, illusion, *is just* to see the order of things" (Taylor 1975, 6). Similarly, I think Descartes would say, to rise out of dream, confusion, illusion, *is just* to see the order of things, only what Descartes means by 'order' is not what the ancients meant. What Descartes means by 'order' is *law-governed*, according to a law, where a law is to be understood on the model of an algebraic equation in two (or more) unknowns that governs the motion of a point in Cartesian space. Descartes in this way realizes a radically new order of intelligibility, a conception of things unfolding not as an expression of their natures but under law. Descartes' new science of order is first and foremost a science of laws.

[56] Or so Descartes would claim. In fact, for reasons that will eventually become clear, the aspiration of early modern mathematics to be the work of the pure intellect will be fully realized only with the nineteenth-century revolution in mathematics that is the topic of Chapter 5. Only with that second revolution will the concepts of mathematics be stripped of *all* the sensory content that at first attaches to them.

3.5 Descartes' Metaphysical Turn

On November 10, 1619, Descartes had an epiphany, one that according to his biographer Baillet revealed to him "the foundations of a marvelous science" (CSM I 4, n. 1). I have suggested that what he realized that day (through a thorough-going figure/ground gestalt switch of our first, everyday conception of space in terms of the relative positions of objects) was what is now known as Cartesian space, space as a given whole of possible positions. Within a few months Descartes began writing his *Regulae*.[57] That work, however, was abandoned in 1628. The following year Descartes began working on *Le Monde*, and it was around this time that Descartes seems to have formed his distinctive conception of reality and our knowledge of it. Here we focus on four key elements of this new metaphysics: (1) Descartes' conception of mind as itself a substance, *res cogitans*, (2) his identification of body with extension, the subject matter of mathematics, (3) his account of the pure intellect as an autonomous faculty operating independently of the senses, and (4) his understanding of the eternal truths of mathematics as freely created by God. As we will see, although Descartes had all the essentials of his new mathematical practice already in November of 1619, it would take him a decade of work on the *Regulae* before he would come to conceive his new mathematical practice as a *purely* intellectual enterprise, one that in no way involves the senses, the imagination, or therefore, the body. Once he did have this conception of the pure intellect, the rest of his metaphysics quite naturally followed. Descartes' mathematics was in this way at once the foundation for and the catalyst of Descartes' radically new metaphysics.

The science of mathematics as Descartes understands it takes as its subject matter not objects but instead the relations and proportions that can obtain among objects, where these relations and proportions are exhibited in the symbolic language of arithmetic and algebra. Nevertheless, Descartes holds in the *Regulae*, the intellect should always be aided by the imagination; although expressible in the formula language of mathematics, mathematical entities must also be capable of graphic expression. Around 1628 Descartes discovered (so he thought) that imagination is not needed in mathematics, that the pure intellect is an autonomous faculty of knowing. Certainly we can imagine, that is, form images of, some mathematical entities—in Meditation Six, Descartes gives as examples, a triangle and a pentagon—but in other cases, say, that of a chiliagon, we cannot. And yet in every case, we can have mathematical understanding and knowledge as expressed in and mediated by the symbolic language of arithmetic and algebra. It follows that, contrary to the claims of the *Regulae*, imagination is not needed in mathematical practice but only the pure intellect.[58] And as Descartes eventually came to realize, this new insight changes everything. Descartes' new mathematical practice in this way ushers in a new metaphysics.

[57] I here follow Schuster (1980).

[58] Again, this actually is, at this moment in history, only an aspiration.

In the *Regulae* Descartes argues that mathematical practice should utilize images on the grounds that the intellect, reflecting in the absence of an image, can mistakenly separate what is in fact inseparable: "we generally do not recognize philosophical entities of the sort that are not genuinely imaginable . . . henceforth we shall not be undertaking anything without the aid of the imagination" (CSM I 59; AT X 442–3). As is made manifest in the subsequent discussion, the problem is that possibilities can be discovered by the intellect alone that are not real possibilities, that are shown to be unreal by an image: "even if the intellect attends solely and precisely to what the word denotes, the imagination nonetheless ought to form a real idea of the thing, so that the intellect, when required, can be directed towards the other features of the thing which are not conveyed by the term in question, so that it may never injudiciously take these features to be excluded" (CSM I 61; AT X 335). Though through the intellect alone one might discover what is logically possible, not all logical possibilities are real possibilities. Images, Descartes thinks at this point, help us to discover what is necessary despite not being logically necessary.

Descartes claims in the *Regulae* that by forming images of things one can avoid the error of taking to be separable what is in fact inseparable. For example, one might, using the intellect alone, determine that extension is not body and on that basis mistakenly conclude that there can be extension without body. Such a mistake is avoided, on the *Regulae* account, by one's forming a real idea of an extension in the imagination and discovering on that basis that it is impossible to form an image of extension that is not also an image of body. The distinction between extension and body is *only* a distinction of the intellect. There cannot actually be extension without body—which is why a vacuum is impossible according to Descartes.[59] That extension requires a body is thus an instance of the sort of judgment Kant would later describe as synthetic a priori, of a judgment that is necessary but not logically necessary, not analytic (by logic alone).

According to the *Regulae*, reason cannot derive the necessary truth that there is no extension without body from the concept of extension alone but must in effect go outside the concept of extension to an image in order to see the necessity of the connection between extension and body. But when Descartes realizes that his new mathematical practice does not require the use of images, when he realizes that his new system of written signs enables him to discover truths independently of his ability to form images, he needs some other way of constraining the intellect, and he needs this because not everything that is logically possible is really possible. This,

[59] "The impossibility of a vacuum, in the philosophical sense of that in which there is no substance whatsoever, is clear from the fact that there is no difference between the extension of a space, or internal place, and the extension of a body. For a body's being extended in length, breadth and depth in itself warrants the conclusion that it is a substance, since it is a complete contradiction that a particular extension should belong to nothing; and the same conclusion must be drawn with respect to a space that is supposed to be a vacuum, namely that since there is extension in it, there must necessarily be substance in it as well" (CSM I 229–30; AT VIII a 49).

then, provides grounds for thinking that God freely creates the eternal truths, the essences. These truths are not logically necessary. It is not, for example, analytic or logically necessary that extension should be the extension of a body; but Descartes thinks that it *is* nonetheless necessary that extension be the extension of some body. And, on his mature account, it is God who is the source of this non-logical necessity.[60] According to Descartes' mature account, we discover the necessary relationship between being extended and being a body using only the pure intellect, by careful examination of our clear and distinct idea of extension, an idea that is created by God and instilled in us as his creatures, and seeing its necessary, though not logically necessary, relationship to the idea of body.

Much the same is true, Descartes thinks, of other fundamental features of reality. We have innate ideas of various simple natures, for instance, substance, duration, order, thinking, doubt, and so on, on the basis of which we can discover, by thought alone, both the essence and fundamental laws of nature and the essence and fundamental laws of the mind, that is, how inquiry ought to be conducted. (See Marion 1992.) Henceforth, not experience, history, and tradition but pure thought is to be the primary and most fundamental means to knowledge.[61] The only authority is that of reason.[62]

Descartes came to his new conception of the science of mathematics through a fundamental and radical reorientation in his understanding not only of what is displayed in a diagram, a drawn figure, but of space itself. Descartes' conception of space is, furthermore, essentially late, possible at all only through a metamorphosis of an earlier, object-based conception. But Descartes takes this new orientation to reveal a cognitive capacity that is always already available to a thinker, something always already there to be discovered. And he explains the fact that this capacity had not before been recognized by appeal to the simple fact that we begin our lives as children

[60] What is logically necessary, on this account, is so even for God. It follows that some necessary truths are not created by God. That God exists, for instance, is presumably logically true, grounded on the (alleged) fact that God is logically possible and if logically possible then (by logic alone) also actual, hence necessary. Such a truth is not created by God. Similarly, it is not in God's power to make something that is a circle at the same time not be a circle because that is logically impossible. But it is in God's power to have made non-logical necessary truths about circles, such as that all radii are equal, be in fact false in much the way we think that it can be false (given certain assumptions) that the sum of the angles of a triangle equals two right angles. In a letter to Mersenne, May 27, 1630, Descartes writes that "God was free to make it not true that all the radii of the circle are equal—just as free as he was not to create the world" (CSM III 25; AT I 152). Whereas Suárez had claimed that God is the author of the existence of things but not also of their essences, Descartes now claims that God created both essences and existence. See Marion (1998, 273–5).

[61] See Garber et al (1998). As they note, "Descartes' explicit call for the complete rejection of learning and tradition had persuaded many Cartesians that they, at least, now possessed a permanent foundation for philosophy that no mere tradition could hope to supply. Even the tradition of Christian Epicureanism which Gassendi had established failed to offer such a guarantee. In the ensuing competition between Gassendists and Cartesians, not only would the future of atomism be affected, but the relevance of history and tradition to philosophical inquiry would also be decided" (1998, 587–8).

[62] But even reason will sometimes demand that experiments be done and observations made. See Larmore (1980).

lacking the full use of reason and so become accustomed to relying on our senses and traditional beliefs rather than on reason alone: “we were all children before being men and had to be governed for some time by our appetites and teachers” (CSM I 117; AT VI 13). Neither, Descartes thinks, can be trusted. Furthermore, in childhood the mind is “so closely tied to the body,” so completely immersed in it, that one comes to confuse the two, attributing to bodies that which can belong only to mind and to the mind that which can only belong to bodies (CSM I 218; AT VIII a 35; also CSM III 188; AT III 420). From the perspective Descartes has achieved according to which mind and body are wholly separate, radically distinct *sorts* of things, the scholastic Aristotelian philosophy that he was taught as a student is subject to exactly this error in supposing that there are substantial forms and real qualities attaching to corporeal substances “like so many little souls in their bodies” (CSM III 216; AT III 648).

Because in childhood one accepts many principles “without ever examining whether they were true” (CSM I 117; AT VI 14), and inevitably so given that we do not have the full use of reason as children, once in one’s life, Descartes thinks, one ought to submit all one’s beliefs to the strictest rational scrutiny so as to discover how knowledge is possible. The “greatest benefit” of the “extensive doubt” of the first meditation “lies in freeing us from all our preconceived opinions, and providing the easiest route by which the mind may be led away from the senses” (CSM II 9; AT VII 12). Hyperbolic doubt purifies the mind, so that the pure intellectual with its innate ideas might shine forth. It sweeps away all our childish beliefs, rids us of our prejudices about the world and the means by which we know it, and thereby clears the ground so that we may come to grasp through the pure intellect how things really are.

Descartes’ doubts are presented in the *Meditations* as self-standing. They are to serve to purify the intellect of the various prejudices accumulated in childhood and through one’s education. But it is hard to see how this is supposed to work insofar as Descartes’ radical doubts, in particular his doubts about the existence of the external world, are barely intelligible on the ancient understanding of being and the ancient conception of our intentional directedness on reality. They do not give someone who is not already a Cartesian compelling reason to doubt in the ways Descartes enjoins us to doubt. (See Broughton 2002.)[63]

Nor can we understand Descartes’ discovery of the pure intellect and all that it entails by appeal to his rejection of Aristotelian philosophy and its mode of explanation in terms of substantial forms.[64] First, as already noted, Descartes had rejected school philosophy long before he became a Cartesian for whom the mind, the pure

[63] Burnyeat (1982, 42) suggests that Descartes’ doubts are possible because his inquiry is strictly methodological but it is not at all clear that this really can explain what needs to be explained, namely, how it could be so much as intelligible that one might seriously doubt the existence of the external world while maintaining one’s apparent experience of it.

[64] Garber (1986) seems to suggest as much. No mention is made of the many non-Cartesians who also rejected the school philosophy.

intellect, is wholly separate from the body and an autonomous faculty of knowing. Along with many others, he was at first a mechanist without being a Cartesian. And others remained atomists and (quasi-Aristotelian) empiricists, even after Descartes himself had made his metaphysical turn. They could remain atomists because Descartes in fact had no good argument against atomism that did not depend in turn on his (anti-empiricist) conception of the pure intellect. The atomist could agree, for example, that there is no mathematical atom—that in mathematics quantity is infinitely divisible—because the atomist could distinguish, as, for instance, Gassendi does, though in a way Descartes' epistemology disallows, between mathematical atomism and natural or physical atomism. Nor could Descartes argue that atomism is false because God can divide even physical atoms; all the atomist needed to claim was that no *natural* force could divide the atom.[65] Rejecting Aristotelian hylomorphism in favor of a more mechanistic philosophy may be necessary for Descartes' conception of the pure intellect, but it is far from sufficient.

Nor, finally, can Descartes' discovery of the pure intellect be explained by appeal to the newly emerging anti-Aristotelian mode of explanation in terms of laws. The reason is simple: the relevant notion of a law—namely, that of a law of motion as expressed in a mathematical equation—is due to Descartes himself, due in particular to his advances in mathematics. Although both Descartes and many of his contemporaries had already rejected both the traditional hylomorphic conception of nature and the traditional conception of explanation by appeal to Aristotelian forms, only Descartes succeeded in developing the alternative to it that was needed.[66] The only viable explanation of the fact that Descartes became the first Cartesian philosopher is the fact that he was first a Cartesian mathematician. It was Descartes' new mathematical practice together with the new mode of intentionality it engendered that enabled Descartes to become a Cartesian metaphysician.

According to the classical scholastic Aristotelian view that Descartes aims to supplant, one's conceptions of things are inevitably conceptions of existing things because ultimately those conceptions are themselves inextricably tied to one's perceptual experiences of the relevant objects. Essence, on such a view, is not prior to existence. This conception is furthermore fundamental to natural language; our first language is and must be one that speaks of those things in the environment of which we are perceptually aware. And it is this everyday perceptual experience of things through the medium of natural language that is the model for all cognition in scholastic Aristotelian philosophy. Even the "pure intellect," for the scholastic Aristotelian, grasps

[65] See Garber (1992, 123–5).

[66] As Milton (1998, 686) writes, we find already in Bacon's writings that "the old theory of substantial forms had been at least officially abandoned; the intention was to replace it by a theory of natural laws, but the precise character of the explanations involved remained obscure even to Bacon himself.... It was that other reformer of the sciences, René Descartes, who made the decisive innovation, by formulating and bequeathing to his successors the vision of a science of moving bodies in which laws of nature, conceived quite specifically as laws of motion, were the most fundamental principles of explanation."

only what is mind independent; even an "intelligible species" is, as a sensible species is in the case of perception, a means by which the intellect understands what is outside the mind. All thought, on the traditional conception of it, is constitutively world directed and world involving. That the "external world" might not exist may have been conceivable as an abstract, academic possibility for the ancients. It was not in any sense a real possibility, and ancient skeptical arguments do not consider it.

But if, as Descartes' new mathematical practice seemed to show, the mind can make discoveries wholly independently of any relation to anything outside the mind, simply by reflecting on its own ideas without even the assistance of the imagination, then both the scholastic Aristotelians and the anti-Aristotelian atomistic empiricists had to be wrong to hold that knowledge of existence, which is achieved through sense experience, is prior to knowledge, if any, of essences. Instead, "according to the laws of true logic," Descartes came to think, "we must never ask about the existence of anything [never ask if it is, *an est*] until we first understand its essence [what it is, *quid est*]" (CSM II 78; AT VII 107–8).[67] Much as we have, Descartes thinks, an idea of space prior to and independent of any and all sensory experience, so essences generally are to be conceived to be prior to and independent of any objects that might instantiate those essences. Descartes' new mathematical practice seemed in this way to reveal not merely a method of discovery, but the mind itself as something in its own right, independent of and wholly different from any other thing—save for God.

On the ancient view, perceptual experience is the model for all intellection. Even pure thought is constitutively world-directed in the way that perception is; and it is possible at all only given our everyday perception of things. Descartes inverts this picture by beginning with the pure intellect as revealed in his new mathematical practice, and then modeling perceptual experience on that form of cognition. Even perceptual experience is not, on Descartes' new account as outlined in the second meditation, inherently object involving. Furthermore, because in Descartes' new mathematical practice the pure intellect is focused not on objects given to sense or to thought but instead on relations and patterns, its "objects" (that is, that on which it is trained) cannot be things outside the mind, that is, objects such as material spheres and cubes. Its "objects" are the patterns and relations that are now held to *underlie* the spherical or cubic appearance of some objects, that is, what it is to be a sphere or cube, an *essence*. The notion of an essence is thereby completely transformed. From being what is essential to some actually existing object, that which makes it to be what it is and so to be at all, essence has become a purely mental entity, a *meaning*, or as Descartes calls it, an *idea*, something that can be directly grasped by the pure intellect and is, on Descartes' new account of cognition, the means by which we understand anything at all. Much as time had come to be divorced from events in Galileo's

[67] This is a central theme in Secada (2000). See also Carriero (1986).

physics and space from our experience of objects in Descartes' mathematics, so "essence . . . is divorced from the object of reference" in Descartes' metaphysics and wedded, not to the word, as Quine (1951, 22) suggests, but to the mind, and becomes thereby a Cartesian idea.

A Cartesian idea, a paradigm of which is the relation expressed in the equation $x^2 + y^2 = r^2$, is nothing like an Aristotelian species, whether sensory or intelligible, through which one experiences or thinks of something outside the mind. Nor is it a universal abstracted from one's sensory experience of instances. Nor, finally, is it a Platonic mathematical or Form seen with the mind's eye. It is a purely mental entity. Although Plato called already for a turning of the soul away from the realm of the senses and becoming to the realm of intelligible things, of being, it is Descartes who realized such a turn, who achieves, in his mathematical practice with the formula language of algebra, the intellectual vision that Plato sought.[68] Unlike a sensory quality such as redness, or even an object such as a sphere on our everyday understanding of it as something with a characteristic look and feel, there is nothing that a properly Cartesian idea looks, feels, or smells like. Although, it can perhaps be pictured (as, for instance, a circle can), it is itself intelligible rather than sensory. And as intelligible it can come to have a kind of transparency, a clarity and distinctness to thought that sharply contrasts with the opacity, or brute givenness, of sensory experience.[69]

Thought on Descartes' mature account has content independent of the deliverances of the senses, independent even of there being anything without the mind at all—save, of course, for God. Thought is contentful not in virtue of being directed on something mind independent but simply by virtue of itself and its innate ideas, by virtue, that is, of the God-given created truths. And among these truths are, first and foremost, the truths of mathematics, the subject matter of which is extension. It follows, Descartes argues, that the essence of matter is extension. We know that the essence of matter is extension because we know that material things "possess all the properties which I clearly and distinctly understand, that is, all those which, viewed in general terms, are comprised within the subject-matter of pure mathematics" (CSM II 55; AT VI 80). Once Descartes has in place his radically new conception of

[68] That it is algebra that is critical for Descartes is made clear in a comment he makes to Burman: "to enable the intelligence to be developed, you need mathematical knowledge" and "mathematical knowledge must be acquired from algebra" (CSM III 351; AT V 176–7).

[69] The intelligibility of Cartesian ideas as contrasted with the sheer brute presence of sensory qualities is central to the third meditation proof for the existence of God. I do not have within me the idea of God as something sensory, as I have an experience of redness, or even of seeing the word 'God' written in (say) black ink on a piece of white paper. I do not experience but instead understand, grasp intellectually, my idea of God, and that is what is wholly inexplicable except, Descartes thinks, by appeal to God. How can I, a mere finite being, not merely contain within me a representation of the infinite but also understand or grasp, even imperfectly as I do, the idea of the infinite, know it as infinite? Only God, Descartes thinks, could bestow such understanding on me. See also Descartes' appeal to "the idea of a machine of a highly intricate design" in motivating the proof (CSM II 75; AT VII 103, also CSM II 97; AT VII 134–5).

pure intellection, his fundamentally transformed mode of intentional directedness, his understanding of matter as *res extensa* quite naturally follows.[70]

Descartes' new conception of curves, finally, provides the model for understanding the notion of a law of nature. We have seen that an equation in two unknowns in Descartes' symbolic language expresses a law governing the motion of a point in Cartesian space. A law of nature similarly is a rule governing matter, extended substance, in motion (CSM I 93; AT XI 37), one that is expressible in an algebraic equation (CSM I 97; AT XI 47). Such laws, like meanings or Cartesian ideas generally, are intelligible independent of the objects they govern. As already noted, a law of motion is not *in* the objects it governs as an internal principle of change, as the principle of motion of a thing is for Aristotle; and it is not a generalization about objects derived from one's experience of objects. It is instead to be conceived independently of the objects it governs, as something in its own right that is an object of knowledge as such. A law of nature, as Descartes comes to conceive it, is the (direct or immediate) expression of a pattern that objects can instantiate. As such a law, it furthermore grounds a new form of explanation in science. On the modern view that Descartes inaugurates, things happen not as the expression of a thing's nature or internal principle of changing and staying the same but as determined by the laws governing all matter, as instances of universal patterns. The central task of science as it is now to be understood is to discover these laws.

The formula language of mathematics that Descartes develops and uses to discover mathematical truths thus enables a radically new cognitive orientation, a new mode of intentionality or world-directedness. We do not now simply find ourselves with the world in view but must instead self-consciously judge, given sufficient reason, that things are thus and so. Sense experience, from providing us with immediate perceptual knowledge of things is now to be understood as something caused in us by the impacts of mere matter. Cognition more generally is to be understood to be directed on things without the mind not intrinsically, as on the classical view, but only through an act of will. The modern subject correlatively comes to understand itself as distinctively free, where to be free is, as Descartes explains in Meditation Four, to act for reasons. Both mind and nature are, then, most fundamentally to be understood in terms of laws, either laws of nature (motion) or laws of freedom (reason). To be, on the modern conception that is enabled by Descartes' new mathematical practice, is to be subject to laws. The order of things, that in terms of which they are intelligible at all, has become the order of law.

3.6 Conclusion

I have traced Descartes' mature philosophical views to his new mathematical practice, a practice that is enabled in turn by a metamorphosis in Descartes' understanding of

[70] See Hatfield (1993).

space. This new practice, we have seen, is focused not on objects but on relations; and as Descartes shows, such relations can be expressed in the formula language of elementary algebra. Pure intellection has thus become (at least in intention) an actuality, not as Descartes himself thought, by stripping away the accretions of youth and education to reveal the true and immutable natures created and implanted in us by God, but by a radical transformation in one's cognitive orientation as mediated by this new sort of language. The world, from being manifest in perceptual experience, is now to be represented in thoughts that are expressible not in the sensory, narrative language of everyday life but instead in a symbolic language, a language that is non-sensory, non-narrative, essentially written, and primarily a vehicle of thought.

This new sort of language engenders at the same time a new understanding of the being of beings. What had been an essential unity of (substantial) form and matter is now to be conceived as split into, on the one hand, autonomous mind with its innate ideas, and on the other, matter the motions of which are governed by discoverable laws of nature. The narrative order of nature, progressively unfolding in an organic process that actualizes it according to its own nature, gives way to the order of law, of reason and freedom on the side of the mind and of nature and causes on the side of matter. Science is henceforth to be reductive and mechanistic; to understand what a thing *is* one need only consider what it *does*. We have achieved precisely the conception of being that underlies and shapes a project such as Brandom's in *Making It Explicit*, the project of making explicit what something would have to be able to do in order to count as rational at all. We have achieved the sideways-on view. We have become modern.