The Duplication of the Square in Plato's *Meno*
(An Appendix to Glenn Rawson's translation)

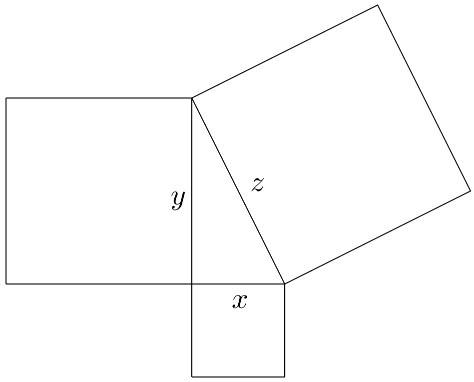
23 July 2006, Laurence Barker, Bilkent University, barker@fen.bilkent.edu.tr

Shortly before beginning his questioning of the slave-boy, Socrates reports an opinion which he considers to be "Something true, it seems to me, and beautiful". The reported opinion concludes "For inquiry and learning as a whole is recollection". To illustrate the conclusion, Plato has Socrates enter into a detailed mathematical discussion. Nowhere else does Plato present mathematics in any detail. Presumably, there was a reason for doing so on this occasion. To what extent did Plato use mathematics as a model for his philosophy? When he speaks of learning as recollection, is this an expression of a metaphysical belief (something true) or is it mystical insight (something beautiful) or would Plato not have seen much of a distinction between the two? In this appendix, we shall not be discussing such questions. We shall merely be presenting some prerequisites that are needed by anyone wishing to engage in the controversies. It is likely that the slave-boy passage is drawn from a core piece of mathematics that would have been recognized, in Plato's time, by educated readers. It is likely that, in a mundane sense, those readers would indeed have been recollecting the material. To put ourselves in their position, we must try to see the mathematics as they might have seen it.

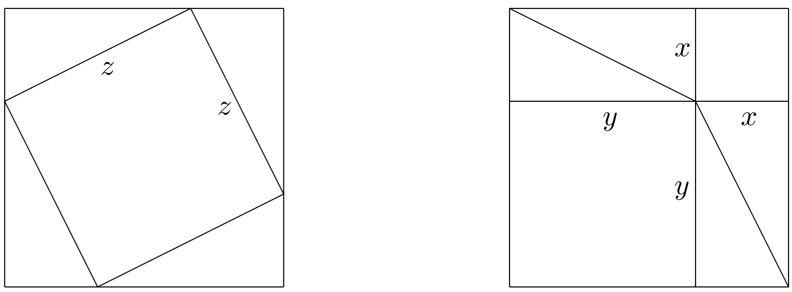
One style of teaching mathematics is first to state a problem, then to present a solution, and then to prove that the solution is correct. The problem: given a square, how can we construct a new square such that the area of the new square is double the area of the original square? The solution: a diagonal of the original square is to be an edge of the new square. The two squares are depicted in the left-hand part of the following diagram. The proof: extending two of the lines, as shown in the right-hand part of the diagram, we obtain five triangles all with the same area. The original square is made up of two of the triangles. The new square is made up of four of the triangles. So the new square does indeed have twice the area of the original square.

Another style is dialogue. Through the teacher's questions and hints, the student is led towards a rediscovery of the material. Of course, Plato employs this style throughout his work (as if he were deliberately emulating a teacher of mathematics). One disadvantage of the style is that it tends to make simple things seem complicated. Dull students dislike being confused by false trails. So Plato emphasizes the benefits: by making mistakes and pursuing misconceptions, the student may achieve a "state of perplexed lacking", and thence, "starting from this wanting and perplexity", the student may be encouraged to "discover while inquiring".

From the perspective of modern school-child mathematics, the slave-boy passage is merely exasperating. Quite simply, the required conclusion follows immediately from Pythagoras' Theorem: given a right-angled triangle, and constructing three squares as shown in the next diagram, then the area of the largest square is equal to the sum of the areas of the two smaller squares. The theorem is nowadays more often stated in the following form. Recall that a square with edge-length z has area z^2. So the theorem says that, letting z be the length of the longest edge of the triangle, and letting x and y be the lengths of the other two edges, then $x^2 + y^2 = z^2$.



However, to Plato, and to the classical Greek mathematicians (and to pure mathematicians ever since), correct beliefs cannot be properly grasped (paragraph 97e) "until one ties them down by working out the reason. And that is recollection... When they are tied down, they become knowledge... That is why knowledge is more valuable than correct belief." One might believe a theorem, and the theorem might be correct, but one does not know the theorem unless one has a proof of it. Curiously enough, the essential mathematical content of the slave-boy passage is a simplification of what is nowadays the most famous proof of Pythagoras' Theorem: the theorem becomes clear by inspecting the following diagram.



A cryptic hint of that proof has been detected in the smug gibberish of the anonymous Chinese work *Chou-pei Suan-ching*, probably written less than two hundred years after the "burning of the books" in 213 BC. The oldest surviving record of the proof — presented as knowledge in Plato's sense — is by Bhaskara, 12th century AC. To the mathematicians of Plato's time, Pythagoras' Theorem was already ancient, but we know next to nothing about how they proved it. We can only speculate that Plato and his intended readers did have the above proof in mind. (Euclid's proof of the theorem, as Proposition 47 in *Elements I*, has been described as a mousetrap; it seems to ramble on, but then it snaps to the conclusion in an unexpected way. Maybe, rather than presenting a proof that all his intended readers would have been shown during childhood, Euclid decided instead to have some fun.)

To empathize with the mathematics of the past, one must first locate the anachronisms in the way one thinks, and then one must erase them from mind. The equation $x^2 + y^2 = z^2$ involves addition and multiplication of a certain kind of number called a positive real number.

The positive real numbers are the numbers that we use to represent magnitudes. In the equation, the positive real numbers x and y and z represent lengths, and the positive real numbers x^2 and y^2 and z^2 represent areas. But, in classical Greek mathematical literature, the only things that could be freely added and multiplied together were the positive integers, that is to say, the numbers 1, 2, 3, 4 and so on. Magnitudes of the same kind could be added together, but not multiplied. (One can add volumes without reference to any notion of a number: a small soup-tin and a medium wine-bottle together have the same capacity as a large milk-carton. One can test this by pouring the soup into the carton, and then adding the wine. But one cannot multiply the soup by the wine.) Comparisons of magnitudes could be multiplied together but not added: the comparative scale of an elephant to a mouse is the scale of an elephant to a cat compounded with the scale of a cat to a mouse. A comparison of magnitudes was called a $\lambda o \gamma o \zeta$. Etymological studies of its usage in 5th and 4th century sources indicate that, in mathematical contexts, $\lambda o \gamma o \zeta$ was to be understood as *word* or *term*, but with the connotation of *something that can be expressed* or *something that is subject to reason.* The usual English translation is *ratio.* It appears that the classical Greeks understood a ratio to be something rather abstract; not a magnitude, and not a number, but some kind of word-thing that could validly be mentioned in mathematical reasoning.

The style of classical Greek mathematical literature is very similar to the style of modern pure mathematical literature. (By comparison, 18th century pure mathematical literature is quite alien.) The modern discipline allows liberal use of imprecise and invalid concepts in oral communication, but it has a tacit ban on mentioning such concepts in the written record. So it is reasonable to surmise that the classical Greek mathematicians were in possession of heuristic concepts which were to be used only while work was in progress. If we find no evidence for any recognition of those concepts in the manuscript sources, that may be because the ancient writers made every effort to remove the evidence; just as, for instance, modern authors of analysis texts usually take great pains to remove all trace of the use of infinitesimals. For reasons that we shall touch upon below, it appears that the concept of a fractional number was considered to be heuristic. In *Republic VII* (525e), Plato remarks “For you are doubtless aware that experts in this field, if anyone attempts to cut up the pure one in argument, they laugh at him and refuse to allow it, but if you really do cut it up, they multiply it, always on guard lest the one should appear to be not one but a multiplicity of parts.” Likewise, modern pure mathematicians laugh at the practical but epistemologically incoherent mathematics of engineers and natural scientists. Why is mathematics sometimes so pernickety? More particularly, why was the mathematics of Plato’s milieu so pernickety? We shall return to that question.

Very often, a mathematical argument begins as a heuristic sketch, and then the mistakes and weak steps are corrected and polished. In the process, for the sake of deductive rigour, the material may have to be rearranged, and the driving intuitive ideas may unfortunately have to be obscured. We shall begin with an intuitive heuristic sketch of the material in the slave-boy passage, and then we shall polish the weak steps. As we perfect the argument, shall find that we need to change the order in which the squares are introduced. Eventually, we shall arrive at Plato’s version of the argument. Of course, we cannot know whether or not Plato himself went through a similar train of thought, but our speculative reconstruction will at least reveal that certain awkward features of his discussion are there for good reasons.

We are given a square, and we are to construct a new square with double the area. The left-hand part of the next diagram shows that, if the edge-length of the new square is double the edge-length of the original square, then the area of the new square is four times the area of the original square. By how much should we increase the edge-length so as to double the area? If we were to increase the edge-length by a factor of one-and-a-half, would the new square have double the area of the original one? Alas not. The area would be increased by a factor of two-and-a-quarter. Indeed, this is clear from the calculation $(3/2)^2 = 9/4$. But now let us insert the four diagonal lines, as indicated in the right-hand part of the diagram. Eight triangles are formed, all with the same area as each other. Consider the central square whose edges are the four diagonal lines. The central square is made up of four of the triangles, while the original square is made up of two of the triangles. So the central square has twice the area of the original square, as required.

That completes our heuristic sketch of the essential mathematical content of the slave-boy passage. To satisfy the likes of Euclid, we must now do some polishing. Actually, the only unsatisfactory part of the argument is in the use of fractional numbers to justify the assertion that, if the edge of a square is increased by a factor of 3/2, then the area is increased by a factor of 9/4.

Let us rephrase the assertion in the language of ratios. A ratio is a comparison of the magnitudes of two objects. The magnitude of an edge is its length. The magnitude of a square is its area. The assertion to be proved is that, if an edge of the original square and an edge of a new square are in the ratio 3 to 2, then the original square and the new square are in the ratio 9 to 4. Observe that the original square can be cut up into 4 small squares as shown in the left-hand part of the next diagram, and the new square can be cut up into 9 small squares as shown in the right-hand half of the diagram. The edge-length of the small squares is half of the edge-length of the original square, and it is a third of the edge-length of the new square. All of the small squares have the same area as each other. Therefore, the original square and the new square are in the ratio 9 to 4.

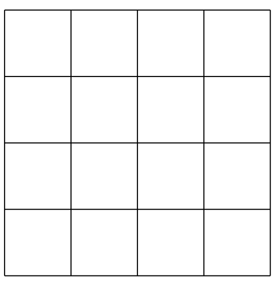
Euclid would still not have been satisfied. From his point of view, the snag in our argument is this: how do we know that any given square can be cut up into 4 squares of equal size, as depicted? To cut up the given square in that way, we must cut all the edges in half. In other words, we must find the mid-point of each edge. But how do we know that the midpoint of an edge exists?

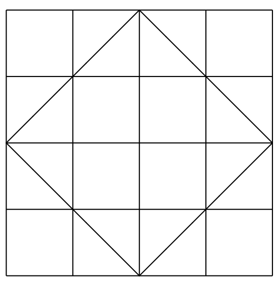
It is not a daft question. A length can be added to itself, and in this way, it can be doubled. Similarly, a length can be tripled or, in fact, multiplied by any positive integer. But halving a length is another matter. After all, a rod with an odd number of atoms cannot be cut exactly in half. The classical Greeks certainly did take such questions seriously. They worked with two basic kinds of mathematical propositions: those which asserted properties of given objects, and those which asserted that an object with given properties can be constructed. The proposition proved in the slave-boy passage is of the latter kind: a new square with double the area of a given square can be constructed. Proposition 1 in Euclid's *Elements I* is another proposition of the same kind: the mid-point of a given edge can be constructed. We can interpret the construction propositions as emphatic assertions that objects with given properties do exist. A new square with double the area of a given square can be constructed, so such a new square *does* exist. The midpoint of a given edge can be constructed, so the midpoint *does* exist.

To rescue our claim that the original square can be cut up into 4 squares above, we could invoke Euclid's Proposition 1. But then we would have to present a proof of the proposition. Euclid's proof is non-trivial. In fact, his proof is flawed, because it relies on the unstated assumption that two circles intersect if each passes though the centre of the other.

Instead, Plato cheats. He cuts up the original square into 4 squares but, in almost the same breath, he proposes that the original square is a two-foot-by-two-foot square. The very notion of a two-by-two square involves the perception that it is made up of 4 one-by-one squares. Thus, the so-called "original" square is not *any* given square. It is a *particular* square that has implicitly been constructed by joining 4 small squares. Perforce, the 4 small squares manifestly do exist. Perhaps the cheating is just for brevity, or perhaps Plato is using the best expedient that the mathematics of his time can supply. The trick does conform to the principle we quoted earlier, "... but if you really do cut it up, they multiply it, always on guard lest the one should appear to be... a multiplicity of parts." We really do cut the edges of the original square into halves but, to play the game properly, we must reorganize the argument, starting with those halves and then multiplying them to obtain the other squares.

Thus, we arrive at Plato's version of the argument. First, he observes that the original two-by-two square consists of 4 of the small one-by-one squares. Since the original square has 4 times the area of each small square, the problem is to construct a square with 8 times the area of the small square. Doubling the edge-length of the original square, he obtains a four-by-four square as in the left-hand part of the next diagram. Does the four-by-four square have the required property? No, it has 16 times the area of the small square. Does the three-by-three square have the required property? No, it has 9 times the area of the small square. He then constructs the central square made up of the diagonal lines, as in the right-hand part of the diagram. To show that the central square does have the required property, he supplies the argument that we presented above: the central square consists of 4 triangular halves of the original square, so the central square has twice the area of the original square. He then gives an alternative argument: by inspecting the diagram, we can see directly that the central square has 8 times the area of the small square, as required.





We have now offered some detailed rationales for the particular technical manoeuvres that occur in the slave-boy passage. However, thus far, we have done nothing at all to explain, in the first place, why Plato bothered to reason in such a neurotically scrupulous manner. Why was classical Greek mathematics so obsessed with rigour? In particular, why was it so concerned with construction and mathematical existence? These questions are vitally relevant to Plato's philosophy as a whole, because mathematics appears to have acquired those traits while he was living. He was witness to what can justifiably be called a revolution in mathematics, and since he was anything but a recluse, we may safely presume that he was caught up in the spirit of the time.

Let us return to the point in the dialogue where we have found that, if the edges of two squares are in the ratio 3 to 2, then the squares are in the ratio 9 to 4. A natural idea would be to seek positive integers a and b such that, if the edges of two squares are in the ratio a to b, then the squares are in the ratio 2 to 1; in other words, the larger square has double the area of the smaller one.

We need to make some observations concerning the compounding of ratios. Given positive integers c, d, e, f, then the ratio c to d compounded with the ratio e to f is the ratio ce to df. The duplicate of the ratio c to d is defined to be the ratio c to d compounded with itself. That is to say, the duplicate is the ratio c^2 to d^2. It is easy to see that if the edges of two squares are in the ratio a to b, then the squares are in the duplicate ratio a^2 to b^2. (We observed this above in the special case $a = 3$ and $b = 2$. The general case can be established in a similar way.) Now imposing the assumption that this duplicate ratio is the same as the ratio 2 to 1, we deduce that $a^2 = 2b^2$.

It may be helpful to recast those observations into a modern idiom. Of course, c/d multiplied by e/f is ce/df. In particular, the square of the number c/d is the number $(c/d)^2 = c^2/d^2$. Recall that a square with edge-length z has area z^2. So, if the edge-length of a square is increased by a factor of a/b, then the area is increased by a factor of a^2/b^2. To double the area, we must increase the edge-length by a factor of a/b such that $(a/b)^2 = 2$, in other words, a/b is a square root of 2. The equation $(a/b)^2 = 2$ can be rewritten as $a^2 = 2b^2$.

Now, assuming there exist positive integers a and b such that $a^2 = 2b^2$, then we can cancel out any common factors, so the smallest possible a and b do not have any common factor. Let us take a and b to be as small as possible. Recall that a positive integer is said to be even when it is divisible by 2, otherwise it is said to be odd. An odd positive integer times an odd positive integer is an odd positive integer, an even times an odd is even, and an even times an even is divisible by 4. Since a and b are as small as possible, they do not have a common factor and, in particular, they cannot both be even. But the equality $a^2 = 2b^2$ implies that a^2 is even, so a is even, so a^2 is divisible by 4, so b^2 is even, so b is even. We have deduced that both a and b are even. But, earlier, we deduced that they cannot both be even. We have contradicted ourselves. The only possible explanation for this absurdity is that our initial assumption is false. We conclude that there do not exist positive integers a and b such that $a^2 = 2b^2$. In other words, there do not exist positive integers a and b such that, if an edge of the original square and an edge of the new square are in the ratio a to b, then the new square has double the area of the original square.

Apparently, then, there does not exist a ratio whose duplicate is the ratio of 2 to 1. But that would suggest that there do not exist two squares, one of them having double the area of the other. On the other hand, the slave-boy passage shows how to construct two such squares. According to the later Hellenic commentators, who may have been reporting hearsay, these perplexing observations originated with the Pythagoreans. At any rate, an accumulation of indirect evidence does strongly indicate that the conundrum — the doubling of the area of the square and the above little wrangle with even and odd numbers — was recognized before the 4th century.

In modern terminology, a positive rational number is a number that can be expressed in the form a/b where a and b are positive integers. We have proved that no positive rational number is a square root of 2. This is the famous theorem called the Irrationality of the Square Root of 2. The modern resolution of the conundrum is to introduce the positive real numbers, which include all the positive rational numbers but which also include other numbers associated with magnitude, such as the square root of 2. It is worth making some notes on some comparatively recent history, because that will help us to appreciate the classical Greek resolution of the conundrum. In the late 16th century, when the differential calculus was seen as mysterious and perplexing, its two greatest pioneers, Newton and Leibniz, both took care over the ontology of the real numbers (Newton taking after Archimedes; Leibniz after Aristotle). By the end of the 18th century however, the techniques had become routine, the sense of mystery had faded, and no longer was any attention paid to queries about what the real numbers actually are. But, during the 19th century, new conceptual troubles arose in connection with Fourier analysis, complex analysis and Riemannian geometry. In the face of various counter-intuitive features of those topics, the casual disregard for the underlying metaphysics became increasingly untenable. And then, suddenly, between 1869 and 1871, three different definitions of the real numbers were proposed, one by Cantor, one by Weierstrauss, one by Dedekind. To repeat Plato's words again: the fundamental questions were considered interesting only when mathematicians were "in a state of perplexed lacking"; the fundamental discoveries were made "starting from this wanting and perplexity".

We do not know whether the 4th century mathematicians saw the conundrum as something essentially numerical concerning a hypothetical square root of 2. Ratios, as we have said, were treated neither as magnitudes nor as numbers. Nowhere, in the surviving literature, are two ratios ever added together in any very direct way. The term *compounding* is, in the context of ratios, the conventional English translation of the Greek term $\sigma\mu\nu\theta\epsilon\sigma\iota\zeta$ — synthesis — which literally meant *combining* or *putting together*. They used a different term, $\pi o\lambda\lambda\alpha\pi\lambda\alpha\sigma\iota\alpha\sigma\mu\o\zeta$, for multiplication of positive integers. But, whether or not the conundrum was intuitively perceived in a very numerical way, the fact remains that it raises a question about the existence of a ratio whose duplicate is the ratio of 2 to 1.

The classical Greek resolution of the conundrum was to develop an axiomatic theory of ratios. In *Elements V*, Euclid states his axiomatic assumptions about ratios, and all further assertions about ratios are to be deduced from those axioms. In this way, he accommodates ratios that cannot be expressed as ratios of positive integers. Above, we noted that if the edges of two squares are in the ratio a to b, where a and b are positive integers, then the squares are in the duplicate ratio a^2 to b^2. Proposition 20 in *Elements VI* tells us that the ratio of any two squares is the duplicate of the ratio of their edges. This is a substantial generalization, and the proof requires much work. It applies to the situation in the slave-boy passage, where an edge of one square is a diagonal of another square. In that case, the duplicate of the ratio of the edges is the ratio of 2 to 1. In particular, there does exist a ratio whose duplicate is the ratio of 2 to 1. Thus, in effect, the theory of ratios is a (somewhat incomplete) theory of the real numbers.

The theory of ratios certainly predates Euclid. All or most of it had been been developed by the time of Aristotle, who died in 322 BC. For the other side of the window in time, there is some circumstantial evidence to suggest that, to the eminent mathematician Democritus, of the late 5th or early 4th century, a ratio was always a ratio of positive integers. The Hellenic commentators attribute the theory of ratios to Eudoxus, an Athenian contemporary of Plato. There is some closely related work that can more reliably be attributed to Eudoxus: the Method of Exhaustion, which was an impractical but impeccably rigorous ritual for confirming results concerning lengths of curves, areas of figures and volumes of solids. (The results would first be found by heuristic means, and then they would be proved using the Method of Exhaustion.) So we can reasonably surmise that Eudoxus made some significant progress along the path that eventually led towards what is arguably the supreme achievement of the classical Greek civilization: the theory of ratios as presented in Euclid's *Elements*.

Plato's *Meno* perhaps predates both the theory of ratios and the Method of Exhaustion. Nevertheless, Plato (or the mathematicians who advised him when he composed the slave-boy passage) certainly had the outlook that lies behind those theories. Obviously, it was not that Plato was presciently conforming to the concerns of Euclid. Rather, Euclid was later to inherit concerns that had already been recognized when Plato wrote *Meno*.

At last, we can address the question: why did mathematics (the theoretical mathematics of the aristocrats, not the practical mathematics of the artisans) become so pernickety during Plato's time? Above, we noted that attention was given to the foundations of real analysis during periods when mathematicians were having trouble making sense of counter-intuitive observations. The one and only counter-intuitive item in all of classical Greek mathematics is the discovery of the irrationals. According to Archimedes, (in his letter to Eratosthenes) the formula for the volume of a cone was found by Democritus but was first proved by Eudoxus. We might surmise that, with advances in techniques for determining lengths, areas and volumes, the queries about the irrationals came to a head and motivated Eudoxus to develop his Method of Exhaustion. Certainly, though, mathematics did undergo a kind of metamorphosis during the early to middle 4th century, mathematicians did start to occupy themselves with questions about mathematical existence, and the discovery of the irrationals did provide at least some of the stimulus. Thus, the direction of subsequent classical Greek thought seems to have been set. Just a little later, in *Physics IV* and *Metaphysics XIV* we have Aristotle agonizing even about the existence of the positive integers. And then, in *Elements VII*, we have Euclid answering Aristotle by defining the positive integers axiomatically.

We have now indicated something of the motives for the style of mathematics in the slave-boy passage (and, possibly, some motives for the style of philosophy in all of Plato's work). It remains for us to justify our assertion that the particular problem discussed in the passage constitutes a core piece of 4th century culture. The evidence lies in the history of another problem, called the Problem of Duplicating of the Cube: given a cube, how can we construct a new cube such that the volume of the new cube has twice the volume of the original cube? The Hellenic commentators associate the problem with colourful legends: doubling the size of the alter of a god, doubling the size of the tomb of a king, or whatever. Actually, the motive for the problem is that, if there exists a cube with double the volume of another cube, then there must exist a ratio whose triplicate is the ratio of 2 to 1 (or, in modern terms, there must exist a cube root of 2; a number t such that $t^3 = 2$). The above even-and-odd argument can easily be adapted to show that such a ratio cannot be a ratio of positive integers (the cube root of 2 cannot be a rational number).

At least two cube-duplication constructions were discovered during the first two-thirds of the 4th century BC. One of them, due to Archytas, involves the intersection of a cone, a cylinder and a torus. Another one, due to Menaechmus, involves the intersection of a plane and two cones. Although both of those solutions to the problem make use of solid geometry, they can be reformulated in terms of curves in the plane. Eutocius, a commentator writing in about 500 AC, mentions two other cube-duplications. He attributes one of then to Eudoxus, but he gives no details other than to inform us that the construction involves intersections of curves. He attributes the other one (implausibly) to Plato, and he does present the argument; it is an application of a so-called mechanical construction technique, which involves moving rods. We mention that Archytas is believed to have met Plato in Sicily; Eudoxus and Menaechmus are thought to have been members of an early-to-middle 4th century Athenian mathematical community which historians whimsically refer to as "Plato's Academy" (although the likelihood is that they had more influence on Plato, rather than the other way around). In the 3rd century BC, another cube-duplication was given by Eratosthenes, and yet another one by Apollonius. The list could go on, but all the other classical or Hellenic cube-duplication constructions on record came rather later. (To avoid misunderstanding, let us note that the famous Impossibility of Duplicating the Cube, proved by Galois in about 1829, asserts only that it is impossible to double the size of a cube using the construction methods implicit in *Elements*, which are called ruler-and-compass constructions.)

So, the mathematical problem discussed in *Meno* — the problem of doubling the area of a square — is one component of a combination of observations which together provided much of the impetus for abstract mathematics (and possibly, abstract thought in general). Since the problem and its solution can be explained to a complete novice in only a few minutes, the mathematical content of the slave-boy passage must surely have been familiar to anyone with a modicum of exposure to mathematics. It must have been a pedagogical set-piece.

Let us end with one last little question: Why does Plato not tell us, from the outset, that the aim is to double the area of a given square? Perhaps this little obfuscation is just for narrative effect. The active theme, in that part of the dialogue, is knowledge as recollection. Had Plato announced the problem right at the beginning of the slave-boy passage, then, as educated readers, we might have felt a wave of ennui: oh no, not again, the duplication of the square. By keeping up the suspense for a bit, Plato artfully delays the moment of recognition: ah yes, I remember this, the duplication of the square.