An Introduction to Discrete Mathematics: how to make mistakes, how to get it right

Laurence Barker, Bilkent University
version: 11 May 2014

These notes in this file are associated with MATH 132.

They are also in an early state of preparation for a chapter of a book.

Style warning: This draft is incomplete and unpolished. Earlier drafts had some terrible overkill on pedantic commentary. The present draft still has some overkill, probably a lot of it. Finding the right balance is tricky, because I am deliberately trying to explicitly discuss material which successful students normally catch onto only through oral communication or by reading between the lines.

The chapters and their associated files are:

Some of those files can be found, under the heading MATH 132, on my homepage at: http://www.fen.bilkent.edu.tr/∼barker/fenis2.html .

Chapter 1: Mathematical induction

Contents of Chapter 1

1.1: The Königsburg Bridge Problem

Mathematical induction is a technique for proving general results by reducing each case to a smaller case, except for those cases which are too small to be reduced and which have to be dealt with separately. The main purpose of this chapter is to introduce that technique.

Of course, mathematics does not proceed through the development of theory first, applications later. Rather, theory and applications develop together, each stimulating the other. It is not required that the applications should actually be useful outside of mathematics. Sometimes, the sole motive for considering a scenario of application is simply that it helps towards stimulating some interesting theory. Sure enough, the whole enterprise would be just an indulgence if mathematical theory were never to have external applications. That, however, is hardly the case. Mathematics does, of course, have prolific applications to other disciplines. However, genuinely useful external applications of mathematics nearly always involve many different areas of mathematical theory. Sometimes, as a way of learning to walk before one learns to run, one has to focus on the mathematics itself, without any distracting discussion of relevance to other domains of study.

For motivation, let us raise the following problem. The problem is of scant direct practical utility. We are not pretending to convey any real contribution towards urban planning. We are considering the problem only because of its mathematical interest.

Problem: *Consider a city that spans across both banks of the river and several large islands within the river. The regions of land covered by the city are connected to each other by bridges over the water. Is there a straightforward criterion for determining whether it is possible to make a tour of the city, crossing the water via the bridges, using each bridge exactly once?*

In Section 1.2, we shall state a complete resolution of the problem.

As we mentioned in the Introduction, a very clear deductive explanation is called a *mathematical proof*. Let us comment briefly on the three characteristics:

• *Explanation:* Every mathematical proof is something that can be communicated from one person to another. It is not an occult ritual, nor a computer program, nor a formal game. When writing a proof, one must think carefully about what the reader may be presumed to know already. If you use a symbol X, say, then you must think about whether or not your reader has been informed about what X denotes. If you use a theorem, then you must to think about whether or not your reader can be presumed to know the theorem. Thus, the validity of a proof depends on the intended audience or readership.

• *Deductive:* Every mathematical proof is a deductive argument. Let us give an example of deduction: granted the assumption that all ravens are black, granted also the assumption that the birds in the tree are ravens, then we can deduce that the birds in the tree are black. Typically, in a deductive argument, a big leap is broken down into

little steps, each little step being obvious. Deduction is the most reliable of all forms of reasoning. It does not include argument by example, argument by most parsimonius guess, argument by analogy.

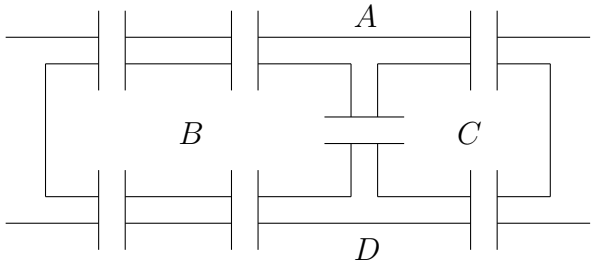
• *Very clear:* Every mathematical proof is very clear. It is merely very clear. One must accept this human limitation. No mathematical proof is perfectly clear. When one has done the best one can to cover everything genuinely helpful to the reader, one must stop. Excessive or irrelevant detail is counter-productive, making it harder for the reader to follow the argument. If, when writing a proof, you try to break down an obvious little step into other obvious little steps, then that will just give you and your reader a headache, whereupon clarity will be lost.

The primary purpose of a proof of an assertion is to supply a very reliable justification of the assertion. Another major purpose, often no less important, is to illuminate the underlying ideas. Sometimes, when a theorem is applied to a problem, it is not the statement of the theorem that is applied, but rather, the main ideas that appear in the proof. So, to apply the theorem, it is not always enough just to know the statement. One must also understand why the theorem holds.

Let us emphasize that mathematical proofs are not absolutely reliable. Professional mathematicians, aware of their own capacity for making mistakes, frequently test their deductive conclusions in special cases and frequently check that their conclusions are consistent with each other. Novices, too, need not to be too surprised when some of their arguments, apparently clear as crystal, sometimes suddenly shatter.

The following famous problem is a special case of the problem stated at the beginning of this section. To resolve this special case, we shall be making use of the notions of even numbers and odd numbers. Recall, the **even numbers** are the numbers 0, 2, −2, 4, −4, 6, −6 and so on. The **odd numbers** are the numbers 1, −1, 3, −3, 5, −5 and so on.

The Königsburg Bridge Problem: *The diagram below is a schematic map of the city of Königsburg, as it was during the middle of the 18th century. A river runs through the city. Part of the city is on one side of the river, region A, part of the city is on the other side, region D, and two other parts of the city are on the islands B and C in the middle of the river. The four regions A, B, C, D are connected to each other by seven bridges. Is it possible to make a tour of the city using each bridge exactly once?*



Resolution: No. Such a tour of the city is impossible. Suppose, to the contrary, that such a tour is possible. Let X be any one of the regions A, B, C, D. If the tour does not start or finish at X, then the number of bridges used to enter X must be equal to the number of bridges used to exit X. So the number of bridges at X must be even. To put it another way, we have shown that, if the number of bridges at X is odd, then X must be the starting region or the finishing region of the tour. There is only one starting region and only one finishing region, yet all four of the regions A, B, C, D have an odd number of bridges. It is now clear that such a tour cannot exist. $\square$

The symbol $\square$ means: the very clear deductive explanation is complete.

$\clubsuit$ $\heartsuit$ $\diamondsuit$ $\spadesuit$

Before developing some of the ideas that appeared in the previous subsection, it will be convenient to devote a little subsection to terminology.

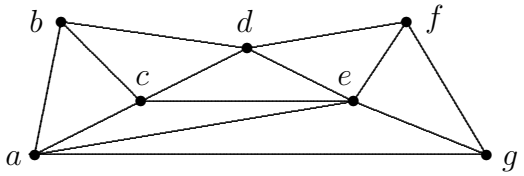
Consider mathematical objects x and y. When x and y are one and the same object, we say that x and y are **equal** to each other and we write $x = y$. Otherwise, we say that x and y are **distict** from each other and we write $x \neq y$. For example, $2 + 2 = 4$ and $2 + 2 \neq 5$.

A precise definition of the notion of a graph will appear in Section 2.2. For now, let us be content with a rough intuitive description: we understand a **graph** to be something which can be represented by a picture consisting of dots and lines, each line connecting two distinct dots to each other, any two distinct dots being connected to each other by at most one line. The dots are called **vertices**. The lines are called **edges**.

For emphasis, let us reiterate some of the conditions. Given a graph G and any two vertices x and y of G, then either there is exactly one edge connecting x and y or else there are no edges connecting x and y. Furthermore, if there is an edge connecting x and y, then $x \neq y$.

A graph is said to be **finite** provided there are only finitely many vertices. Perforce, obviously, a graph has only finitely many edges. In all our discussions, we shall be considering only finite graphs. Whenever we speak of a graph, it is to be understood that the graph is finite.

The depicted graph has 7 vertices, labelled a, b, c, d, e, f, g. It has 13 edges. One of the edges connects a and b. There is no edge connecting a and d.



We understand a **path** in a graph to be a path along the edges, starting at a vertex and finishing at a vertex. Let us give a more precise definition, just to make sure that there is no ambiguity. A **path** starting at vertex x and finishing at vertex y is defined to be a finite sequence of vertices $(z_0, z_1, z_2, ..., z_n)$ such that $x = z_0$ and $y = z_n$ and, for all integers integers i in the range $1 \leq i \leq n$, there is an edge between z_{i-1} and z_i. We

call $(z_0, z_1, ..., z_n)$ a path of **length** n from x to y. For the graph depicted above, the sequence (c, d, b, a, c, d, f) is a path from c to d with length 6.

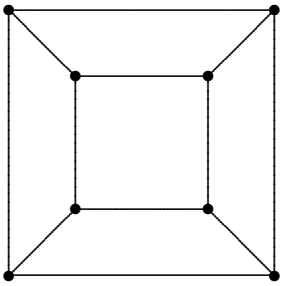
The **degree** of a vertex is defined to be the number of edges at that vertex. For the example above, the degrees of the vertices a, b, c, d, e, f, g are, in respective order, 4, 3, 4, 4, 5, 3, 3.

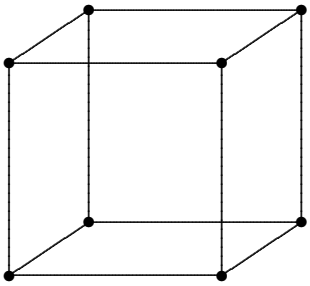
Below, for graphs in general, we shall be concerned with the number of vertices such that the degree is odd. For this example, the vertices with odd degree are b and e and f and g; the number of vertices with odd degree is 4.

♣ ♡ ♢ ♠

An exercise, in a mathematical text, is an invitation to develop mathematical skill by doing mathematics. Some exercises state conclusions and request proofs. Some exercises raise questions. In the latter case, answers are to be supplied and, if the answers are not obvious, proofs of the answers are also to be supplied.

Exercise 1.1.A: *The following two diagrams are two different depictions of the same graph. The graph has* 8 *vertices and* 12 *edges. Show that there is no way of adding two more edges so as to produce a graph with a path that uses every edge exactly once.*





Response: Let G be any graph obtained from the given graph by adding two edges. Suppose that G has a path using every edge exactly once. For each vertex x, let $d_G(x)$ denote the degree of x as a vertex of G. We mean to say, $d_G(x)$ is the number of edges of G that connect x to another vertex. If x is not the start or the finish of the path, then the number of times the path enters x is equal to the number of times the path exits x. Hence, $d_G(x)$ must be even. It follows that there cannot be more than 2 vertices such that $d_G(x)$ is odd. But all 8 of the vertices of the depicted graph have degree 3, which is odd. Adding two edges changes the degree of at most 4 of the vertices. So, for at least 4 of the vertices x, the degree $d_G(x)$ is odd. We have shown that, no matter how the two edges are added, the graph G cannot have a path as specified. □

For comparison, let us consider another possible response to the exercise.

Weak response to Exercise 1.1.A: If vertex not start or finish, number of entries is number of exits. At least four vertices odd degree. Impossible. ⊠

Certainly, the second response is much better than nothing. It could, quite possibly, help a reader to catch onto the key idea behind the argument.

In one sense, the second response is actually better than the first one. It captures the essence succinctly. For some readers, clever enough to see the resolution as soon as two quick clues are supplied, the second response might communicate the whole argument faster than the first response did.

The weakness of the second response, though, is that its vagueness protects it from detailed criticism. If someone wished to have a go at picking holes in the first sentence, then a dialogue with the writer might be needed to confirm interpretations. "If vertex not start or finish...". Some particular vertex, or any vertex? Start or finish of what? Arguments that resist criticism by failing to define their terms tend to be unconvincing.

The strength of the first response is that every assertion is, potentially, immediately open to criticism. We were not afraid to write, very clearly, "For each vertex x..." and "If x is not the start or the finish of the path, then the number of times the path enters x is equal to the number of times the path exits x. Hence, $d_G(x)$ must be even." We had no fear of being as clear as possible because we knew that, if anyone were to raise an objection, then we would be right, the opposition would be wrong, and we would win the debate. For a mathematical argument to be convincing, every deduction must be expressed so clearly that, were it wrong, then it would be immediately vulnerable to refutation.

To express mathematical arguments with adequate clarity, telegraphic notes rarely suffice. One should write in complete sentences that are grammatically correct or, at least, grammatically sound enough to be understandable and unambiguous.

We now make a generalization. The following proposition captures, in general form, the idea that we made use of above. The proposition was given by Leonhard Euler in 1736. A stronger result, the Euler Path Theorem, will be stated in the next section.

Proposition 1.1.1: (Euler's Half of the Euler Path Theorem.) *Let G be a graph, let s and t be vertices in G, and suppose there exists an path from s to t such that the path uses every edge of G exactly once. Let r be the number of vertices x of G such that the degree of x is odd.*
(1) *If $s = t$, then $r = 0$.*
(2) *If $s \neq t$, then $r = 2$.*

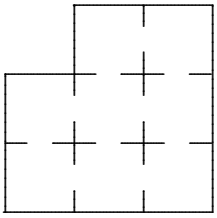
Proof: We first deal with part (2). Consider an Euler path from s to t. Let x be a vertex distinct from x and y. The number of times the path enters x is equal to the number of times the path exists x. So the degree of x must be even. On the other hand, the number of times the path enters s is one less than the number of times the path exits s. So the degree of s is odd. Similarly, the degree of t is odd. We have shown that there are exactly 2 vertices with odd degree, namely, s and t.

The argument for part (1) is similar, but let us run through the details again. For every vertex x, including the case where $x = s = t$, the number of entries into x equals the number of exits from x, hence the degree of x is even. □

The advantage of the generality is that, when dealing with suitable problems, we can simply refer to the proposition instead of repeating a discussion of the idea behind it.

Quicker Response to Exercise 1.1.A: After adding two edges, the number of vertices with odd degree is at least 4. But the Euler Path Theorem tells us that there cannot exist a path of the specified kind unless the number of vertices with odd degree is 0 or 2. □

Exercise 1.1.B: *A prisoner is trapped in a dungeon consisting of* 8 *cells and* 10 *open doorways, as depicted. She wishes to make a tour of the dungeon, passing through each of the ten doorways exactly once, finishing in the cell where she started. Is this possible?*



Response 1: No, there is no such tour. If such a tour exists then, for each cell, the number of times the tour enters the cell would be equal to the number of times the tour exists the cell. That is impossible, because two of the cells have an odd number of doorways. □

Response 2: The answer is: no. Let us reformulate the problem in terms of graph theory. Replacing each cell with a vertex and replacing each doorway with an edge, as indicated in the left-hand diagram, we obtain the graph shown on the right. We must show that there is no path starting and finishing at the same vertex and using every edge exactly once. By Euler's Half of the Euler Path Theorem, for any graph with such a path, all of the vertices have even degree. But, for the graph in question, two of the vertices have degree 3. □

Applying the Euler Path Theorem to the Königsburg Bridge problem is fairly easy too, but a bit of care is needed. We shall return to it in the next section.

♣ ♡ ♢ ♠

Pedantic footnote: One awkward feature of the practical usage of the word *proof* is that, after an uncorrectable mistake has been discovered in an argument previously considered to have been a proof, the argument is then deemed not to be a proof and never to have been a proof. Even though it was previously considered to be very clear, it is now deemed to be not very clear and never to have been very clear. Thus, mathematical proof would seem to be subject to the anachronizing power of the historical pen.

As a closely related source of trouble with the concept of proof, there is a widespread view that the concept is culturally dependent and has changed over time. Certainly, many mathematical arguments that were considered to have been proofs in the 18th and early 19th centuries are no longer considered to be proofs.

Without getting deeply into the debate, let us say a few words to counter the surmise that, at any given moment in time, a mathematical proof is nothing more nor less than

an argument that would, at that particular moment in time, be awarded a stamp of approval by a consensus of authoritative mathematicians.

Judgements about how much detail to give, in a proof, are influenced by experience of recognizing mistakes. The collective experience of the mathematical community, at the present time, is very different from its collective experience two-hundred or three-hundred years ago. Styles of judgement have not changed in a steady gradual way. Mathematics underwent a revolution during the final three decades of the 19th century. Feelings about how much detail might be appropriate, in various different kinds of context, have remained fairly static since the early 20th century.

As a reflection in microcosm, similar differences in judgement may arise in discussions between experienced and inexperienced individuals. A teacher and a novice student may frequently disagree about whether the supplied detail in an argument is insufficient or satisfactory or excessive. Often, that may be because the teacher has much more experience about what kinds of argument frequently collapse and what kinds of argument rarely collapse. Skill at making such calls of judgement can come only from lots of practise. Professional mathematicians do not always arrive at similar verdicts on this matter. It is not very uncommon for one referee of a submitted paper to demand more detailed proofs, another referee to demand abridgement.

Sometimes, when a teacher is objecting to a particular step in an argument produced by a student, the teacher may be able to supply a clinching illustration of how that kind of argument can be dodgy. However, in cases where the student remains unconvinced, the student need not surmise that she is having trouble with the concept of proof. The overriding principle is that, for each individual, the sole legitimate arbiter of proof is oneself.

An examiner may mark a good exam script by subtracting marks for mistakes instead of adding marks for positive contributions. Thus, to be assessed by mistakes does tend to be a good sign. Getting things right is okay too. What is not okay is that anxious state of mathematical paralysis which renders some novices incapable of doing anything mathematical at all.

Familiarity with pandemic kinds of mistake is important, and some degree of partial immunity does have to be acquired if one is later to progress to more advanced theory. Besides, an individual mathematician who keeps making and then eventually recognizing mistakes is, in some sense, recapitulating the process by which mathematics has been developing for more than two-dozen centuries.

Mathematics, like swimming or basket weaving, is an activity that requires skill. To acqure proficiency at mathematics, one must develop skill though active exercise. Passively reading, listening and watching would not suffice if one wished to become good at swimming or basket weaving, nor does it suffice if one wishes to become good at mathematics.

In the research literature, exercises abound, but they are usually implicit, written between the lines. Research mathematicians know how to ask themselves good questions. Besides, when a research mathematician is reading somehing of interest to her, she

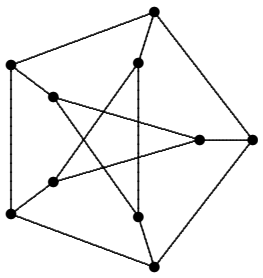
probably has at least some vague musing about how she might use the material in her own work. Exercises related to problems already on her mind will tend to arise inevitably. She is likely to spend much more time following her own thoughts, rather than reading what the author of the text has written.

In introductory books on mathematics, it is usually presumed that the student has not yet aquired a strong inclination to follow her own lines of thought. Partly for that reason — and also with the aim of training the student for success in exams — exercises are usually supplied at the ends of each section or, at least, at the end of each chapter. You should spend much more time doing mathematics, rather than just reading about it. That is to say, you should spend much more time on exercises than on studying text.

When you find an exercise easy, you might learn just a little bit from it. When you struggle for a long time with an exercise that you find difficult, you are very likely to learn a lot. Getting frustrated, possibly confused too, is good for you, because the resolution, when it eventually comes, will be memorable. If, after much work, you really feel that you cannot do an exericse, then you may look at the response supplied at the end of the chapter. What you must not do, though, is to give up on an exercise before you have endured great suffering. If you give up too quickly, then you will not properly appreciate the crux of the problem, you will not properly appreciate the resolution, the ideas will not etch themselves deeply into your memory, and you will not recognize the essence of the problem when you later encounter it in a superficially different form.

Exercise 1.1.C: A chessboad and 32 dominoes pieces are of such a size that each dominoes piece exactly covers two squares of the chessboard. It is not hard to see that that the 64 squares of the chessboard can be covered using the dominoes pieces. Now suppose that two opposite corners of the chessboard are removed. Is it possible to cover the remaining 62 squares using 31 dominoes pieces? (Hint: Consider the usual colouring of the squares of a chessboard, alternatively black and white.)

Exercise 1.1.D: The following diagram depicts a graph called the **Peterson graph**.



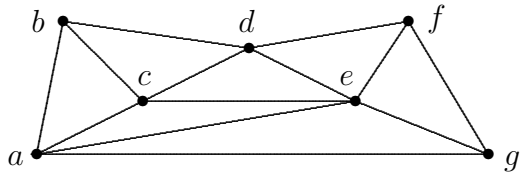
(1) Show that there is no way of adding 3 edges to this graph so as to obtain a graph with a path that uses every edge exactly once.
(2) Show that there is a way of adding 4 edges so as to obtain a path that uses every edge exactly once.
(3) Show that there is no way of adding 4 edges so as to obtain a path such that the start is the same as the finish and every edge is used exactly once.

Exercise 1.1.E: Let G be a graph and let r be the number of vertices of G that have odd degree. Show that r is even.

1.2: Statement of the Euler Path Theorem

Having stated and proved half of the Euler Path Theorem in the previous section, let us not keep up the suspense about what the whole theorem says. We shall state the theorem in this section. Its proof, though, is quite substantial, and we postpone it to Section 3.FISH of Chapter 3.

Some further notation will be useful. Recall that, given a vertex x of a graph, we define the **degree** of x to be the number of edges at x. We denote the degree of x by the expression $d(x)$. Returning to one of the examples we introduced in the previous section, consider the following graph. The degrees of the vertices are $d(b) = d(f) = d(g) = 3$ and $d(a) = d(c) = d(d) = 4$ and $d(e) = 5$.



A path in G is called an **Euler path** provided the path uses every edge of G exactly once. Obviously, if G has an Euler path then G is connected.

When the start of a path is the same as the finish, we call the path a **circuit**. That is to say, a path $(z_0, z_1, ..., z_n)$ is a circuit if and only if $z_0 = z_n$. For example, for the graph drawn just above, the path (a, e, d, b, a) is a circuit. The path (a, e, d, b, c) is not a circuit.

An Euler path that is also a circuit is called an **Euler circuit**. The next result is just a restatement of Proposition 1.1.1.

Proposition 1.2.1: (Euler's Half of the Euler Path Theorem.) *Let G be a graph.*
(1) *If G has an Euler circuit, then every vertex of G has even degree.*
(2) *If G has an Euler path that is not a circuit, then G has exactly two vertices with odd degree. Moreover, the path starts at one of those two vertices and finishes at the other.*

In particular, for the latest graph drawn above, there are exactly 4 vertices with odd degree. Therefore, the graph does not have an Euler path.

A graph G is said to be **connected** provided, for all vertices x and y of G, there exists a path from x to y in G. A graph that is not connected is said to be **disconnected**. The latest graph drawn above is an example of a connected graph. If we were to remove the vertices a and d and e, also removing all the edges associated with those two vertices, then we would obtain a disconnected graph with 4 vertices and 2 edges.

Obviously, every graph with an Euler path is connected. So, when discussing Euler paths, we may as well confine our attention to connected graphs.

The following theorem was stated by Euler in an 1873 paper that was chatty and recreational in style. Proof of the harder half of the theorem was omitted. Perhaps Euler did see a proof of that half, his omission deliberate, for the sake of a wide audience. A proof, communicated to Christian Wiener by Carl Hierholzer, was published in an 1873 paper posthumously authored by Hierholzer, actually written by Wiener.

Theorem 1.2.2: (Euler Path Theorem.) *Let G be a connected graph. Let r be the number of vertices of G that have odd degree. Then:*
(1) *We have $r = 0$ if and only if G has an Euler circuit.*
(2) *We have $r = 2$ if and only if G has an Euler path that is not a circuit. Every such path starts at one of the two vertices with odd degree and finishes at the other.*

In particular, a connected graph has an Euler path if and only if the number of vertices with odd degree is 0 or 2. We have already established that result in one direction. If G has an Euler path, then $r = 0$ or $r = 2$.

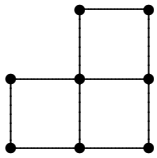
The implication in the other direction will require much more work. It is not at all easy to prove that, if $r = 0$ or $r = 2$ then there exists an Euler path. Heiholzer's proof, presented in Section 3.FISH, will make use of a technique called *argument by mathematical induction*.

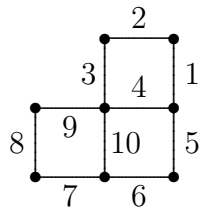
♣ ♡ ♢ ♠

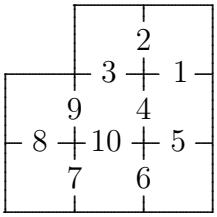
Let us give two little examples to illustrate the more difficult half of the theorem.

In the scenario of Exercise 1.1.B, suppose that the prisoner lowers her expectations, still wishing to make a tour of the dungeon passing through each of the ten doorways exactly once, but now no longer insisting that the starting cell must be the same as the finishing cell. Glancing at the graph we associated with that exercise, drawn again on the left below, we see that there are exactly two vertices of odd degree, whereupon part (2) of the Euler Path Theorem immediately tells us that the graph has an Euler path. In other words, the prisoner can take herself on a guided tour of the dungeon, recounting the history of each doorway as she passes though it, no doorway left out of the tour and no doorway repeated.

For this little graph, fining an Euler path is very easy. One such path is specified in the middle diagram, the edges numbered according to their order of appearence in the path. The corresponding tour of the dungeon is indicated in the right-hand diagram.





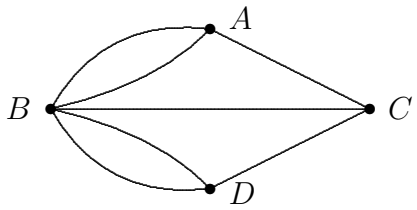


Another little application arises in Exercise 1.1.D. Recall that parts (1) and (3) of that exercise immediately succumbed to Euler's half of the Euler Path Theorem. But, to do part (2), some work was required, indeed, we had to go to the trouble of actually finding an Euler path. When we have proved the whole of that theorem, though, we will be in a stronger position: we shall be able to deduce the existence of an Euler path without any need to specify one.

♣ ♡ ♢ ♠

We resolved the Königsburg Bridge Problem in Section 1.1. Now that we have some graph theory behind us, let us see if we can simplify the argument by invoking the Euler Path Theorem. The following argument has a tiny mistake.

Slightly flawed resolution of the Königsburg Bridge Problem, this time using graph theory. If there is a tour of the city such that each bridge is used exactly once, then the graph depicted in the following diagram has an Euler path. But that is impossible by Euler's half of the Euler Path Theorem, since all four of the vertices have odd degree. ⊠



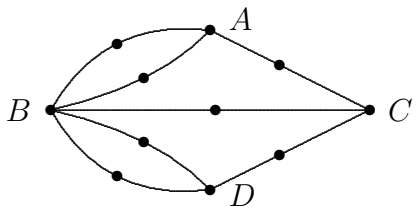
In an exam script, the red pen of the examiner draws blood from mistakes. Generally, criticism focusses on weaknesses. However, criticism can take place only after the ideas have been introduced. Without ideas, there is nothing. Mistakes, very often, can be detected and corrected later, during the polishing phase. The argument we have just presented does, pretty much, resolve the Königsburg Bridge Problem. It is good mathematics.

Nevertheless, the phrase "the graph depicted in the following diagram" is potentially confusing. A reader, struggling to follow the argument, can validly raise the objection that the diagram does not depict a graph. In the diagram, there are two edges connecting A and B, two edges connecting B and D. But, in the definition of a graph, one of the conditions is that there is at most one edge between any two vertices. The Euler Path Theorem, as stated above, pertains only to graphs.

It is a very minor mistake, though. Let us give two corrected versions of the argument.

Resolution 1 of the Königsburg Bridge Problem using graph theory. Consider the diagram above, where each spot represents a region of the city and each line represents a bridge. The diagram does not depict a graph. Nevertheless, bearing in mind that all four spots have an odd number of edges attached to them, the proof of Proposition 1.1.1 can easily be adapted to show that there does not exist any tour as specified. □

Resolution 2 of the Königsburg Bridge problem using graph theory: Consider the graph depicted below, each vertex representing either a region or a bridge, each edge representing contact between a region and a bridge. Plainly, there exists a tour using each bridge exactly once if and only if the graph has an Euler path. But 4 of the vertices have odd degree. So by Euler's half of the Euler Path Theorem, no such tour exists. □



Of course, the abstract notion of a graph is not needed for the Königsburg Bridge Problem. After all, we did manage to resolve that problem, in the previous section, without making use of any graph theory. But the abstraction does help us to generalize. Let us return to the problem presented at the beginning of this chapter.

Exercise 1.2.A: *Consider a city consisting of separate regions together with some bridges, each bridge affording passage between two distinct regions. Let r be the number of regions that have an odd number of bridges. Suppose that it is possible to get from any region to any other region using the bridges. Using the full version of the Euler Path Theorem, stated as Theorem 1.2.2, show that there exists a tour of the city using each bridge exactly once if and only if $r = 0$ or $r = 2$.*

Response: Consider the graph such that each vertex corresponds either to a region or to a bridge, two vertices having an edge between them if and only if one of the vertices corresponds to a region and the other vertex corresponds to a bridge affording passage to and from that region. Clearly, this graph has an Euler path if and only if the city can be toured in the specified way. The Euler Path Theorem — the full version of it, stated as Theorem 1.2.2 — now yields the required conclusion. □

But we have just now done something pedagogically very dangerous. We have broken an important rule of undergraduate mathematics: one may apply a piece of theory only if one knows that it is correct. Criminally, we applied the full version of the Euler Path Theorem, *but we have not yet proved that version of the theorem!*

The rule applies only to undergraduates, not to research students or to professional mathematicians. There is a very good reason for the double-standard.

In research, most mathematicians do make use of theorems without having personally checked the proofs. Research mathematicians with no need to do that are, for the most part, specialists in backwater little topics that have few connections with other areas of mathematics.

However, all competent professional mathematicians have a clear awareness of the epistemological status of their items of knowledge. In their own branch, at least, they have clear mental pictures of the networks of deductive relationships. Mentally, they have a red asterisk, or something equivalent, associated with every result that depends on material they have taken on trust. Such scruples are needed because knowledge is fluid: as new material is learned, details of material learned in the past are fogotten. The corpus of knowledge is constantly being checked and reviewed as mathematicians explain things to themselves and to each other. But mathematicians have this ability only because, at some stage in their education, they were constrained by the strict rule whereby the pronouncements of authority are deemed to be of zero value, nothing to be accepted unless personally checked by oneself.

Thus, we have a loose end to tie up. We are not yet in possession of a proof of the full version of the Euler Path Theorem. So we have not yet properly conquered Exercise 1.2.A. As we mentioned above, a proof of the full version of the Euler Path Theorem, making use of the technique of mathematical induction, will appear in Section 3.FISH. When we have proved that theorem, we shall also have Exercise 1.2.A.

1.3: Argument by contradiction and the Well-Ordering Property

Argument by contradiction, traditionally called *reductio ad absurdum* or reduction to absurdity, is a fairly straighforward technique of proof. Another good name for it might be *argument by vandalism*. When using the technique, instead of working our way directly towards the desired conclusion, we start from the statement asserting that the desired conclusion is false, and then we deduce our way towards a contradiction or an impossibility.

Let us make a comment on the language that tends to be used in connection with the technique. In an argument by contradiction, we *suppose* that the desired conclusion is false, we *assume* that the desired conclusion is false. This does not mean, of course, that we actually believe the desired conclusion to be false. After all, when trying to prove something, we often do already believe it to be true, or at least, we do usually expect it to be true. Rather, the idea behind argument by contradiction is to examine the hypothesis that the desired conclusion is false, and then to destroy the hypothesis.

We have already been making use of the technique. In Section 1.1, in our resolution of the Königsburg Bridge Problem, we were to show that no tour of the specified kind exists. Our method was to assume that, on the contrary, such a tour does exist. We then destroyed that assumption by using it to deduce an assertion that is plainly impossible, namely, that each of the four regions of the city must be a start or a finish of the tour.

In this section, we shall be making use of the technique in connection with a property of the natural numbers called the Well-Ordering Property.

Let us agree on some terminology. The numbers appearing in the infinite sequence $(0, 1, -1, 2, -2, 3, -3, ...)$ are called the **integers**, those in $(1, 2, 3, 4, 5, ...)$, the **positive integers**, those in $(-1, -2, -3, -4, ...)$, the **negative integers**, those in $(0, 1, 2, 3, ...)$, the **natural numbers**. Note that, under our conventions, zero is a natural number, and zero is neither positive nor negative. (Some writers adopt other conventions about this, but they are nowadays probably in a minority.)

For many purposes, and for all our purposes in this book, we can understand a **set** to be a collection of things. The things in a set S are called the **elements** of S. Thus, any set is the collection consisting of all its elements. We admit the notion of a set with no elements, called the **empty set**, denoted $\emptyset$.

As examples: the number -1 is an element of the set of integers, but it is not an element of the set of natural numbers. The number 0 is an element of the set of natural numbers, but it is not an element of the set of positive integers.

We distinguish between **infinite sets**, which have infinitely many objects, and **finite sets**, which have only finitely many objects. For a finite set S, the number of elements in S, denoted $|S|$, is called the **size** of S. The natural numbers are the sizes of the finite sets, including the number 0 which is the size of the empty set $\emptyset$. Thus, $|\emptyset| = 0$. The set consisting of all the natural numbers is infinite.

Let us mention that, for some technical purposes, in mathematics at a much more advanced level, our above definition of the notion of a set is unsatisfactory. Trouble can

arise in connection with infinite sets. We shall make some further comments about this later in this section.

However, discrete mathematics rarely engages in very sophisticated constructions involving infinite sets. In areas of mathematics that fit comfortably under the heading "discrete mathematics" or "combinatorics", the problems that arise often focus on finite sets and often involve some form of counting. Although infinite sets do frequently come into consideration, exotic trouble with the infinite rarely arises.

As we mentioned in the Introduction, the mathematical term *discrete* was introduced by Henry Billingsley, in 1570. In his characterization of the positive integers as *discrete*, in the sense of *distinct and separate*, he was following a long tradition of commentating on Euclid using clues gleaned from studies of Aristotle and Plato.

Aristotle's take on the difference between the discrete and the continuous will be of no practical use to us. Nevertheless, it is interesting, so let us say a few words about it. In various works, most clearly in *Categories*, Aristotle proposed that there are two kinds of quantity: discrete and continuous. Discrete quantities are sizes of collections of distinct separate things, for example, a pack of dogs. The dogs are distinct and separate from each other. We can number the dogs, say, as dog number 1, dog number 2, and so on. If the last of them is dog number 10, then we say that there are 10 dogs. The size of the pack is 10. For continuous quantities too, the size of the whole is the sum of the sizes of the parts but, as Aristotle saw it, things with continuous sizes do not have distinct and separate parts, rather, the parts merge together and share common boundaries. The length of a piece of string is the length of the left-hand part plus the length of the right-hand part.

Of course, the distinction between "discrete" counted quantities and "continuous" measured quantities far predates Aristotle. It is deeply coded into the grammar of most or all spoken languages: how *many* eggs? How *much* omlette? But, for the Classical Greeks, the distinction was not just a banal observation about grammar. The following theorem, certainly known by the early 4th century BC, probably known long before that, can be interpreted, very roughly, as saying that the continuous is much more subtle than the discrete. We shall comment further on that interpretation in the next subsection.

We shall not be making use of the theorem in this book, but the proof is worth presenting here because it is a paradigm of the technique of argument by contradiction. We mean to say, it is a famous standard example.

Theorem 1.3.1: *There do not exist positive integers a and b such that $(a/b)^2 = 2$.*

Proof, expressed using the conditional tense: If such a and b were to exist, then we could choose a and b such that b would be small as possible. Obviously, the product of two odd integers is odd. We would have $a^2 = 2b^2$, which would be even, so a would be even. That is to say, we would have $a = 2\alpha$ for some positive integer α. Hence, $(2\alpha)^2 = 2b^2$. Dividing by 2, we would obtain $2\alpha^2 = b^2$, which would be even, hence b would be even, in other words, $b = 2\beta$ for some positive integer β. Hence, $\alpha^2 = 2\beta^2$. But α and β would

be positive integers, and β would be smaller than b. But that is preposterous, because b was to be as small as possible. $\square$

The argument becomes easier to read when presented without using the conditional tense, as below. Note, however, that the conditional mood is still to be understood. When trying to follow a proof that makes use of the technique, it might help to imagine that it is to be spoken in a sarcastic tone.

Proof, expressed more conventionally: Suppose, for a contradiction, that such a and b exist. Then we may choose a and b such that b is as small as possible. Obviously, the product of two odd integers is odd. But $a^2 = 2b^2$, which is even, so a must be even. That is to say, $a = 2\alpha$ for some positive integer α. We have $(2\alpha)^2 = 2b^2$. Dividing by 2, we obtain $2\alpha^2 = b^2$, which is even, hence b is even, in other words, $b = 2\beta$ for some positive integer β. Hence, $\alpha^2 = 2\beta^2$. But α and β are positive integers, and β is smaller than b. This contradicts the assumption that b is as small as possible. $\square$

In both versions of the proof, we implicitly made use of the fact that, among all the positive integers b for which there exists a positive integer a satisfying $(a/b)^2 = 2$, there is a smallest such b. Let us generalize that observation. Every non-empty set of positive integers has a smallest element.

The following assertion expresses the same idea. It does not really matter whether we state it for the positive integers or for the natural numbers. The idea is essentially the same either way.

The Well-Ordering Property of the Natural Numbers: *Every non-empty set of natural numbers has a smallest element.*

The Well-Ordering Property is obvious. As with all obvious assertions, there is no need to prove it and, in fact, it cannot be proved. There is no point in associating an obvious assertion with a fake imitation deductive argument consisting of a list of other obvious assertions.

We mention that, as an alternative way of organizing the material in this section, we could have stated the Well-Ordering Principle before discussing Theorem 1.3.1. Then, when proving Theorem 1.3.1, instead of writing "Then we may choose a and b such that b is as small as possible", we could have written "By the Well-Ordering Property of the Natural Numbers, we may choose a and b such that b is as small as possible". Maybe that would have made the proof a bit easier to follow. Or maybe it would have cluttered the proof — starting from the obvious fact that a smallest such b exists, abstracting an obvious general principle, then invoking the obvious principle to get back to an obvious fact of the existence of b — thus making the proof a bit harder to follow. Anyway, it does not matter. Proofs, even short ones, usually admit much variety of satisfactory presentations.

Let us give another version of essentially the same proof. It may be a moot question as to whether, in this version, the Well-Ordering Property has any relevance at all.

Alternative version of the same proof of Theorem 1.3.1: Assuming, for a contradiction, that such a and b exist, then we can replace a and b with a/g and b/g, where g is the

largest positive integer that divides a and b. So it must be possible to choose a and b such that 1 is the only positive integer dividing a and b. Arguing as before, we deduce that 2 is a common divisor of a and b. Again, this is a contradiction, as required. □

♣ ♡ ♢ ♠

Let us say a few more words about an interpretation of the latest theorem.

Recall, we use the term **rational numbers** to refer to the numbers having the form a/b where a and b are integers and $b \neq 0$. We use the term **real numbers** to refer to the numbers, positive or zero or negative, that are used for expressing continuous magnitudes such as length, time, speed. Of course, when expressing such magnitudes, we make use of units of measurement: to say that a giraffe is x meters tall is to say that the height of the giraffe is x times the length of a one-meter rod.

It is helpful to think of the real numbers as the points on a straight line that runs infinitely far in both directions. The line is called the **real number line**. If one prefers not to think of a number as being exactly the same thing as a point then, instead of identifying the real numbers with the points of the real number line, one can understand that there is a correspondence, each real number corresponding to exactly one point on the line, each point on the line corresponding to exactly one real number.

Of course, every integer is a rational number. Every rational number is a real number. The real numbers that are not rational numbers are called the **irrational numbers**. As we shall explain in moment, irrational numbers do exist.

For convenience, let us imagine the real number line to be horizontal, placed in front of us, the positive real numbers to the right, the negative real numbers to the left. Consider real numbers x and y. When $x \neq y$ and x located leftwards of y, we say that x is **less than** y, we say that y is **greater than** x and we write $x < y$ and $y > x$. For example, $-3 < 0 < 2 < 10^2/7^2$. The negative real numbers are precisely the real numbers that are less than zero. The positive real numbers are precisely the real numbers that are greater than zero. When x is less than or equal to y, we write $x \leq y$ and $y \geq x$.

The question as to whether or not the next remark is obvious depends, very much, on what one understands the real numbers to be. In many areas of mathematics, it is helpful to have more than just one way of thinking about the real numbers. At the foundations of a branch of mathematics called *real analysis*, there are some ways of constructing the real numbers without making any use of the visually intuitive geometric conception of the real number line. In introductory courses on real analysis, much of the work is in confirming that some alternative definitions of the real numbers, based on constructions in terms of the rational numbers, do amount to the same thing as the geometrically conceived real numbers. If one starts with one of those alternative definitions, then the conclusion of the following remark requires a considerable amount of preliminary theory. On the other hand, if one understands the real numbers to be the points on the real number line, as intuitively conceived, then the remark is obvious. For any real number x, we call x^2 the **square** of x. We call x a **square root** of x^2. Of course, $-x$ is also a square root of x^2.

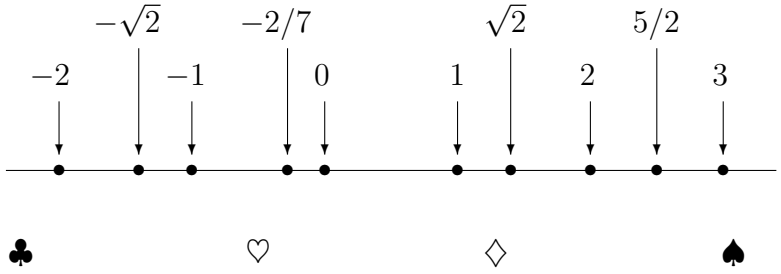
Remark 1.3.2: *Every non-negative real number has a unique non-negative square root. That is to say, given a real number x such that $x \geq 0$, then there exists one and only one real number w such that $w \geq 0$ and $w^2 = x$.*

For w and x as in the theorem, we write $w = \sqrt{x}$. We can now reinterpret Theorem 1.3.1 as follows.

Corollary 1.3.3: *The real number $\sqrt{2}$ is irrational.*

In practise, when indicating approximate lengths, we usually consider fractions of units of measurement. A typical length of a rabies virus is, in decimal notation, about 0.00000018 meters, in other words, 180 nanometers. However, given two lengths, we cannot always find a suitable unit such that both of the lengths are an integer multiple of that unit. The diagonals of a square are $\sqrt{2}$ times as long as the edges of the square. Corollary 1.3.3 says that there is no unit of length such that both the length of the diagonal and the length of the edge are counts of parts that are one unit long. Continuous quantities are not just counts of units. Continuous quantities are, as the Classical Greeks understood, fundamentally different from discrete quantities.

The next diagram depicts part of the real number line. The diagram indicates the locations of (the points corresponding to) several of the integers, the locations of the rational numbers $5/2$ and $-3/7$, the locations of the irrational numbers numbers $\sqrt{2}$ and $-\sqrt{2}$.



In a moment, we shall be presenting another argument which draws from perceptions associated with the Well-Ordering Principle. First, let us set the scene with an easier result whose conclusion is similar in kind.

Exercise 1.3.A: *Show that $1 + 2 + 3 + ... + n = n(n + 1)/2$ for all positive integers n.*

Response 1: We have

$$2(1 + 2 + ... + (n-1) + n) = (1 + 2 + ... + (n-1) + n) + (n + (n-1) + ... + 2 + 1)$$

$$= (n+1) + (n+1) + ... + (n+1) + (n+1) = n(n+1) \; . \qquad \square$$

Response 2: This is a sum of n terms whose average is plainly $(n + 1)/2$. $\square$

We now turn to a more tricky exercise.

Exercise 1.3.B: *Show that, given a positive integer n, then*

$$1^2 + 2^2 + 3^2 + ... + (n-1)^2 + n^2 = n(n + 1)(2n + 1)/6 \; .$$

Response, for a reader not yet familiar with the mathematical induction technique: Let $S_n = 1 + 2 + ... + n$ and $T_n = n(n+1)(2n+1)/6$. We are to show that $S_n = T_n$ for all positive integers n. Assume, for a contradiction, that $S_n \neq T_n$ for some n. Let us take n to be as small as possible such that $S_n \neq T_n$. We have $S_1 = 1^2 = 1 = 1.2.3/6 = T_1$, so $n \geq 2$. But

$$T_n - T_{n-1} = n(n+1)(2n+1)/6 - (n-1)n(2n-1)/6 = \frac{n}{6}\big((n+1)(2n+1) - (n-1)(2n-1)\big)$$

$$= \frac{n}{6}\big((n^2+3n+1) - (n^2-3n+1)\big) = n^2 = S_n - S_{n-1}\ .$$

Combining that equality with the equality $S_{n-1} = T_{n-1}$, we deduce that $S_n = T_n$. This contradicts the assumption that $S_n \neq T_n$, as required. □

Let us admit that, because of the sting at the end of the proof, the reasoning is not very easy to follow. The argument does seem to be rather like a snake eating its own tail, or perhaps we should say, a scorpion eating its own tail. In the next section, we shall introduce a technique which allows such arguments to be presented in a smoother and more easily readable way.

Pedantic footnote: The material in this book can be followed well enough if we simply understand a set to be a collection of things. Unfortunately, that straightforward understanding of the concept is inadequate for most branches of mathematics at a more advanced level.

From its inception, set theory has always been grappling with metaphysical troubles. Surprisingly, the troubles have genuine relevance to technical arguments, and they not just a source of entertainment for dilettante philosophers. In almost all branches of pure mathematics, one cannot progress far without learning some set theory and coming to grips with the problems which set theory addresses. The only exception to that rule, perhaps, would be finite combinatorics.

Set theory arose, initially, from some work of Georg Cantor on infinite trigonometric sums. In 1871, while considering some constructions involving some infinite sets of real numbers, he invented a generalization of the technique that we shall be discussing in the next section: argument by mathematical induction. His generalized technique was based on a notion of an *ordinal number*. The ordinal numbers include the natural numbers but they also include some infinite numbers. A crucial feature of the ordinal numbers is that, like the natural numbers, they satisfy a version of the Well-Ordering Principle.

Nowadays, use of the ordinal numbers is very rare. Nearly always, problems that can be resolved using the ordinal numbers can be resolved more easily using a theorem called Zorn's Lemma. Nevertheless, applications of Zorn's Lemma are sometimes rather delicate and cannot be carried out reliably without a working knowledge of set theory.

Let us briefly indicate the nature of the troubles. What kind of things are to be studied by mathematicians? Granted that the Horsemen of the Apocalypse are War, Plague, Famine and Pestilence, does it make sense to speak of the Four Horsemen of the Apocalypse? Does it make sense to say that the set of Horsemen of the Apocalypse is a finite set with size four? We can avoid such questions by insisting that mathematics is the

study of mathematical objects. Mixing mathematics with other disciplines is sometimes necessary, indeed, it is often very interesting, but it is a venture that is always hopelessly subject to vagueness.

For the purposes of the present discussion, let us confine our attention to mathematics itself, viewed in isolation from other disciplines. The question still remains as to what the term *mathematical object* ought to mean. We shall not give a definitive answer to that question. Numbers, though, must surely be mathematical objects. Points and lines must be mathematical objects. Sets of numbers, sets of points and sets of lines must be mathematical objects. From there, it is not a great leap to propose that sets of sets must be mathematical objects. Hence we may consider sets whose elements are sets.

The following item of trouble was noted in a letter, dated 1902, from Bertrand Russell to Gottlob Frege. Another item of trouble, similar but a little more complicated, called the Burali-Forti Paradox, had been published by Cesare Burali-Forti in 1897. It seems likely that Cantor — highly sensitive to criticism and perhaps reluctant to communicate material that might have undermined his theory — could have been secretly aware of such troubles at a much earlier date.

Russell's Paradox: *Let us understand that every collection of mathematical objects is a set, every set is a collection of mathematical objects, every set is, itself, a mathematical object. Consider the set R whose elements are precisely those sets S such that S is not an element of S. Observe that, by the definition of R, if R is an element of R then R is not an element of R. Also observe that, again by the definition of R, if R is not an element of R, then R is an element of R.*

Very roughly, the established approach to dealing with such troubles is to modify the notion of a set. Instead of giving a definite criterion for telling whether or not an object is a set, some conditions, called axioms, are imposed on the collection of all sets. Although various systems of set-theoretic axioms are sometimes considered in exceptional circumstances, or in specialist work on mathematical logic, just one standard system, called the *ZFC system*, is normally used in ordinary pure mathematics. When doing mathematics, we assume that the collection of all sets satisfies the axioms. The effect of the axioms is to ensure that sets are very safe collections of mathematical objects. Unfortunately, confining attention to sets would disallow many mathematical objects that are of vital importance. So a broader notion has to be introduced. A class, roughly speaking, is a moderately safe collection of mathematical objects. As an example, under the ZFC system, it can be proved that the class of all sets is not a set.

♣ ♡ ♢ ♠

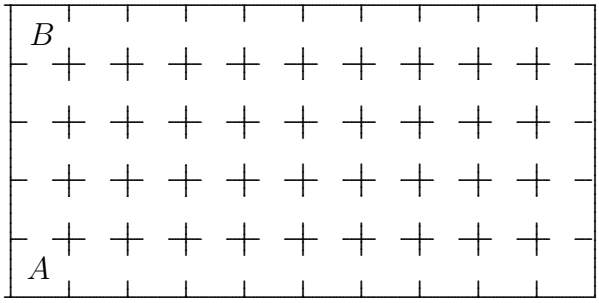
Exercise 1.3.C: Show that, if 100 rocks are to be distributed among 15 trolls, each rock given to exactly one troll, then two trolls will receive the same number of rocks.

Exercise 1.3.D: Show that $\sqrt{15}$ is irrational. Show that $\sqrt{3}+\sqrt{5}$ is irrational.

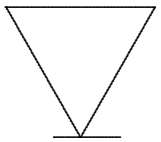
Exercises marked with the symbol ♠ may be especially difficult.

Exercise 1.3.E ♠: A prisoner is trapped in a dungeon consisting of 50 cells arranged in a 10 by 5 grid. She is free to pass from cell to cell via the doorways indicated in the

diagram. Can she make a tour of the dungeon, visiting each cell exactly once, starting in the bottom-left cell labelled A, finishing in the top-left cell labelled B?



Exercise 1.3.F ♠: Show that an equilateral triangle cannot be cut up into finitely many equilateral triangles all of different sizes. (Hint: among the triangles obtained, consider the smallest one with a vertex abutting against an edge as depicted.)



1.4: Some easy illustrations of mathematical induction

We shall introduce a technique of proof called *mathematical induction*. The technique tends not to be useful for finding new assertions. But, after a speculative assertion has been stated, mathematical induction can sometimes be used to prove it.

Let us agree that a **statement** is something that is either true or else false. It cannot be both true and false. It cannot be neither true nor false. When we say that a statement **holds**, we mean that it is true. When we say that it **fails**, we mean that it is false.

It is to be understood that, given statements P and Q, then the three statements "P implies Q" and "If P then Q" and "Either P fails or else P and Q both hold" all have exactly the same meaning as each other. If one of those three statements holds, then all three of them hold. For example, the statement "The sea is boiling hot" implies the statement "Pigs can fly". To put it another way, the statement "If the sea is boiling hot then pigs can fly" holds.

As an aside, we mention that, apocryphally, a student once challenged a lecturer to explain how the statement "$2+2=5$" implies the statement "the lecturer is the pope". The lecturer replied "Subtracting 3, we deduce that $1=2$. The pope and I are two. So the pope and I are one." Of course, that reply was a joke. Really, the statement "If $2+2=5$ then the lecturer is the pope" holds because the statements "$2+2=5$" and "the lecturer is the pope" both fail.

Principle of Mathematical Induction: *For each natural number n, let P_n be a statement. Suppose that P_0 holds and that each P_n implies P_{n+1}. Then P_n holds for all natural numbers n.*

To put it another way, given an infinite sequence of statements $(P_0, P_1, P_2, P_3, ...)$, supposing that P_0 is true, also supposing that P_0 implies P_1, that P_1 implies P_2, that P_2 implies P_3 and so on, then it follows that all the statements P_0, P_1, P_2, P_3, P_4, ... are true. Like the Well-Ordering Property, the Principle of Mathematical Induction is obvious and, in particular, it cannot be proved.

Let us apply it to Problem 1.3.B. Much of the following argument is repeated from the version of the argument that we presented in previous section, but the ending, in this version, will be more straightforward. Recall, we are to prove the following proposition.

Proposition 1.4.1: *For all positive integers n, we have*

$$1^2 + 2^2 + 3^2 + ... + (n-1)^2 + n^2 = n(n+1)(2n+1)/6 \, .$$

Proof: Let $S_n = 1^2 + 2^2 + ... + n^2$ and $T_n = n(n+1)(2n+1)/6$. We must show that $S_n = T_n$ for all n. We shall argue by induction. First note that $S_1 = 1 = 1.2.3/6 = T_1$. Now suppose that $n \geq 2$ and that $S_{n-1} = T_{n-1}$. Then

$$T_n - T_{n-1} = n(n+1)(2n+1)/6 - (n-1)n(2n-1)/6 = \frac{n}{6}\big((n+1)(2n+1) - (n-1)(2n-1)\big)$$

$$= \frac{n}{6}\big((n^2+3n+1) - (n^2-3n+1)\big) = n^2 = S_n - S_{n-1} \, .$$

Hence, via the equality $S_{n-1} = T_{n-1}$, we deduce that $S_n = T_n$. $\square$

The most important part of the argument was the sentence "We argue by induction". Generally, when you are arguing by induction, it is helpful to include the word *induction* or *inductive* or *inductively* in the text. That will put your reader roughly in the picture straight away.

Note that the latest argument does not fit exactly into the notation of the Principle of Mathematical Induction quite as we stated above. We might understand P_n to be the statement "$S_n = T_n$". Because we were working with the positive integers instead of the natural numbers, the statement we established at the beginning was P_1 instead of P_0. Later on in the argument, we showed that P_{n-1} implies P_n instead of showing that P_n implies P_{n+1}. But these differences are merely differences of notation, not essential content: we were still establishing that the first statement holds and that each statement in the sequence implies the next statement.

♣ ♡ ♢ ♠

Let us criticise some weaknesses of style which appear frequently in work by novices.

Incomplete proof of Proposition 1.4.1: $A_n - A_{n-1} = B_n - B_{n-1}$.
$n^2 = ((n+1)(2n+1) - (n-1)(2n-1)/6$.
$6n^2 = 2n^2 + 3n + 1 - (2n^2 - 3n + 1) = 6n^2$. True. ⊠

As a way of figuring out why the desired result is correct, it is quite natural to write something like the above. Experienced mathematicians often do something similar, but only on rough paper as private notes prior to composing proofs. Perhaps, if offered as a proof, the above note might be better than nothing. It might supply the reader with some helpful clues.

One weakness is that, when we wrote the above, we knew what we meant by A_n and B_n but we neglected to inform the reader. Perhaps, in this very simple case, the reader would be able to guess. In more complicated arguments, undefined symbols tend to be incomprehensible.

Slightly better incomplete proof: $A_n = 1^2 + ... + n^2$.
$B_n = n(n+1)(2n+1)/6$.
$A_n - A_{n-1} = B_n - B_{n-1}$.
$n^2 = ((n+1)(2n+1) - (n-1)(2n-1)/6$.
$6n^2 = 2n^2 + 3n + 1 - (2n^2 - 3n + 1) = 6n^2$. Yes. ⊠

The deductive flow, though, is still hard to follow because of the lack of signpost words. The kind of role played by each equation has not been indicated. For instance, it was perhaps not entirely clear that the equation $B_n = n(n+1)(2n+1)/6$ was the *definition* of B_n, not a *deduction* from something earlier.

With just two little guidelines of style, as follows, such omissions become much easier to avoid.

• Always write in grammatically correct or at least grammatically parsable sentences. The sentence should run smoothly when spoken out loud, with all but the shortest mathematical expressions pronounced as *blah*.

• Every sentence should begin with a word, not a mathematical symbol or expression. Distinct mathematical expressions should be separated from each other by words, not just by punctuation, except when they appear in a list.

It is okay to write "The four smallest natural numbers are 0, 1, 2, 3". That is a list. It reads aloud as "The four smallest natural numbers are blah", which parses well. It is not okay to write "For any integer n, if $n^2 = 1$, $n > 0$, $n = 1$." It reads aloud as "For any integer n, if blah", which parses badly enough to raise the alarm that something might be unclear. The statement "For any integer n, if $n^2 = 1$ and $n > 0$, then $n = 1$" is true. The statement "For any integer n, if $n^2 = 1$, then $n > 0$ and $n = 1$" is false.

Rather better incomplete proof: Let $A_n = 1^2 + ... + n^2$ and $B_n = n(n+1)(2n+1)/6$. Do we have $A_n - A_{n-1} = B_n - B_{n-1}$? Do we have $n^2 = ((n+1)(2n+1) - (n-1)(2n-1)/6$? Do we have $6n^2 = 2n^2 + 3n + 1 - (2n^2 - 3n + 1) = 6n^2$? Yes, it is true. ⊠

Deductions are relationships between statements. As we noted earlier, a statement is something that is either true or else false. A question is not a statement. Expressing a deductive argument as a list of questions tends not to be very helpful.

Better still: Let $A_n = 1^2 + ... + n^2$ and $B_n = n(n+1)(2n+1)/6$. We must show that $A_n - A_{n-1} = B_n - B_{n-1}$, equivalently, $n^2 = ((n+1)(2n+1) - (n-1)(2n-1)/6$. But this is clear, because $6n^2 = 2n^2 + 3n + 1 - (2n^2 - 3n + 1)$. ⊠

Mistakes can easily arise when one works by manipulating a desired equality, testing it to see if it comes down to a triviality. To show that $2 + 2 = 5$, it suffices to show that $2 + 2 - 5 = 0$, in other words, $0 = 0(2 + 2 - 5) = 0$, which is clear.

In the next draft of the argument, we rearrange the order of the assertions so that the deductive flow becomes easier for the reader to pick up. We also helpfully inform the reader of the fact that we are appealing to the Principle of Mathematical Induction.

Even better: Let $A_n = 1^2 + ... + n^2$ and $B_n = n(n+1)(2n+1)/6$. Observing that $6n^2 = 2n^2 + 3n + 1 - (2n^2 - 3n + 1)$, we obtain $n^2 = ((n+1)(2n+1) - (n-1)(2n-1)/6$, hence $A_n - A_{n-1} = B_n - B_{n-1}$. By induction, we are now finished. ⊠

But one flaw remains. We failed to deal with the case $n = 1$. The latest attempt has exactly the same deductive structure as the next argument.

Attempted proof of the false assertion that $1^2 + ... + n^2 = (2n+3)(2n^2+1)/12$ *for all positive integers* n*:* Let $A_n = 1^2 + ... + n^2$ and $C_n = (2n+3)(2n^2+1)/12$. Then

$$12(C_n - C_{n-1}) = (2n+3)(2n^2+1) - (2(n-1)+3)(2(n-1)^2+1)$$

$$= 4n^3 + 6n^2 + 2n + 3 - (4n^3 - 6n^2 + 2n + 3) = 12n^2 = 12(A_n - A_{n-1}) .$$

Therefore $A_n - A_{n-1} = C_n - C_{n-1}$. By induction, we are now done. ⊠

To some of us, the criticisms we have just raised may seem to be just a pedagogical nitpicking about style, not content. That objection to the criticisms may be, perhaps, natural and reasonable. Experienced mathematicians dislike sloppy presentations of arguments. That is partly because they have seen, many times, how ambiguities attract

mistakes. But it is also partly because they know how much extra effort it takes to read though an argument where guesses have to be made about what the writer means.

Those of us who lack such experience may remain unconvinced for now. Let is not press a resolution to the objection. Time and experience will eventually resolve it for us.

♣ ♡ ♢ ♠

Let us give two more illustrations of the technique.

For each natural number n, we define $n!$ to be such that $0! = 1$ and, when n is positive, $n! = (n-1)!\,n$. The positive integer $n!$ is called the **factorial** of n, usually pronouced "n factorial".

The definition we have just presented is an example of what is called a **recursive definition**. Generally, in a recursive definition, each value is defined in terms of one or more earlier values, except for one or more initial conditions at the beginning. Mathematical induction is often useful for proving assertions about mathematical objects that have been defined recursively.

Directly from the definition,

$$1! = 1\ , \qquad 2! = 2\ , \qquad 3! = 2.3 = 6\ , \qquad 4! = 2.3.4 = 24\ , \qquad 5! = 2.3.4.5 = 120\ .$$

Generally, $n! = 1.2.3....(n-2)(n-1)n$, the formula intepreted suitably when $n = 0$.

Exercise 1.4.A: *Show that, for all positive integers n, we have*

$$1!\,1 + 2!\,2 + 3!\,3 + ... + (n-1)!\,(n-1) + n!\,n = (n+1)! - 1\ .$$

Response: Let $A_n = 1!\,1 + 2!\,2 + ... + n!\,n$ and $B_n = (n+1)! - n$. We shall show, by induction, that $A_n = B_n$. Observe that $A_1 = 1 = B_1$. Now suppose that $n \geq 2$ and that $A_{n-1} = B_{n-1}$. Then

$$A_n - A_{n-1} = n!\,n = n!(n+1-1) = (n+1)! - n! = B_n - B_{n-1}$$

and it follows that $A_n = B_n$. □

The **Fibonacci sequence** $(F_0, F_1, F_2, ...)$ is the infinite sequence defined recursively by the condition that,

$$F_{n+2} = F_{n+1} + F_n$$

for all natural numbers n, furthermore, $F_0 = 0$ and $F_1 = 1$. The integer F_n is called the **Fibonacci number** indexed by n. The next few terms of the sequence are

$$F_2 = 1\ , \qquad F_3 = 2\ , \qquad F_4 = 3\ , \qquad F_5 = 5 \qquad F_6 = 8\ , \qquad F_7 = 13\ , \qquad F_8 = 21\ .$$

The **Lucas sequence** $(L_0, L_1, L_2, ...)$ is defined by the same recursive equation

$$L_{n+2} = L_{n+1} + L_n$$

but with the conditions $L_0 = 2$ and $L_1 = 1$. The next few Lucas numbers L_n are

$$L_2 = 3\ , \qquad L_3 = 4\ , \qquad L_4 = 7\ , \qquad L_5 = 11 \qquad L_6 = 18\ , \qquad L_7 = 29\ , \qquad L_8 = 47\ .$$

Exercise 1.4.B: *Show that, for all natural numbers n, we have*

$$F_0 + F_1 + ... + F_{n-1} + F_n = F_{n+2} - 1 \ , \qquad L_0 + L_1 + ... + L_{n-1} + L_n = L_{n+2} - 1 \ .$$

Response: Let $S_n = F_0 + ... + F_n$. Clearly, $S_0 = F_0 = 0 = F_2 - 1$. For positive n, inductively supposing that $S_{n-1} = F_{n+1} - 1$, then

$$S_n - S_{n-1} = F_n = F_{n+2} - F_{n+1} = F_{n+2} - 1 - (F_{n+1} - 1) = F_{n+2} - 1 - S_{n-1} \ .$$

Cancelling, we deduce that $S_n = F_{n+2} - 1$. The argument for the Lucas numbers is similar, except that, pertaining to the case $n = 0$, we observe that $L_0 = 2 = L_2 - 1$. □

The final sentence of the latest response is not just a lazy summary of an argument that ought to be written out in full. It is a complete proof of the equality $L_0 + ... + L_n = L_{n+2} - 1$. Of course, it would have been possible to have written out a complete self-contained proof of that equality. But, in view of the context, it was unnecessary to do so. Any reader who has already read and understood the proof of the equality $F_0 + ... + F_n = F_{n+2} - 1$ will immediately see that most of the argument applies to the Lucas numbers too. Had we written out the main calculation again, but with L_n in place of F_n, competent reader would not have bothered to work through it.

A written proof being a communication to a reader, there is no point in writing anything that a reader ought to ignore. In the argument above, the natural way of grasping the second required equation is to recognize the similarity with the first required equation. So the best way of explaining the second equation, in a proof, is to draw attention to the similarity.

Phrases such as *the argument is similar* or *by similar argument* or *similarly* occur frequently in the literature. They are very useful as a device for avoiding repetition of an argument whose essence has already been communicated in a similar situation.

We could have used that device in the proof of Proposition 1.1.1. After having proved part (2) of that proposition, we could have dealt with part (1) very quickly, just by writing "The argument for part (1) is similar", ending it right there, without the unnecessary continuation "but let us run through the details again."

Of course, in an exam, few marks may be awarded to response of the form "the proof is very similar to an argument that was presented in lectures". Such a response might be quite correct, but it might be deemed inadequate because of a social complication: the examiner may be unsure about whether the candidate would be able to present the proof to someone who had not attended the lecture.

♣ ♡ ♢ ♠

Pedantic footnote: Almost always, when two mathematicians have a difference of opinion about whether some unambiguous mathematical statement is certainly true or certainly false or incompletely resolved, they can eventually reach agreement through discussion. Typically, in technical dialogues between two mathematicians, assertions are stated, queried, defended. Gradually, the two of them accumulate material upon which they do both agree.

Dialogue was, perhaps, the original form of mathematical discussion. The monologue feature of proof, perhaps, first arose when people started to make advance preparation for mathematical debate with others. When two debating mathematicians both agree that some given assertion is true, there may be no need for either one of them to keep trying to elucidate it. Thus, in oral communication, there is scant need to delineate a distinction between the obvious and the very easy.

In written mathematical communication, the distinction can become a bit awkward. Very frequently, in fact, almost constantly, the composer of a proof has to make judgements about whether some particular assertion would be accepted or queried by the reader. In borderline cases, it hardly matters which way the decision falls. What does matter is the treatment of the difficult steps.

The distinction between the difficult and the not difficult is of far greater importance than any pernickety separation of the completely banal from the almost but not quite completely banal. If one is reading a proof after having tried and failed to find a proof oneself, then one may be searching only for a key idea that will fill a gap which one was unable to fill oneself. Well-written proofs tend to make the key ideas stand out.

At an introductory undergraduate level, however, too much emphasis is sometimes placed on supplying explanations for very easy assertions. This is potentially harmful, because it may undermine the concept of proof at a stage where that concept has not yet been established. For the teacher, avoiding compromise may be exceedingly tricky, since material that makes sense to the weakest students may be trivial in the view of the strongest; exercises accessible to the weakest may not exercise the strongest at all. Thus, a student with strong mathematical insight, chronically failing to detect the non-obvious, may be led to the impression that mathematical proof is a meaningless ritual.

But the compromise arises only because mathematical proof is barely applicable to very easy problems. Proof comes into its own in application to difficult problems.

The following interesting case sometimes crops up in the introductory literature on mathematical induction. In some introductory texts, the natural numbers are treated in an intuitive way, much as in Section 1.3, yet an attempt is made either to deduce the Principle of Mathematical Induction from the Well-Ordering Property of the Natural Numbers, or else to deduce the Well-Ordering Property from the Principle of Induction. But both the Well-Ordering Property and the Principle of Induction are obvious. It does not make any sense to try to deduce one of those observations from the other.

Let us offer a likely explanation as to how that blunder may have arisen. The explanation involves a bit of history.

In 1888, Richard Dedekind described the natural numbers in a very fundamental way. The description was developed in 1889 by Giuseppe Peano. Their aim was to clarify the foundations of mathematics and, in particular, the foundations of arithmetic. Dedekind's tactic was to find another way of defining the term *finite set*, then to define the natural numbers to be, in effect, the sizes of the finite sets. Starting from that definition, he deduced various fundamental properties of the natural numbers. Peano turned that material around, selecting some properties which, taken together, completely characterize the natural numbers. He then took that characterization as his definition of the natural numbers. The characterizing conditions which Peano selected are nowadays

called the Peano Axioms. The last of the Peano Axioms is the Principle of Mathematical Induction.

Thus, for Peano, the Principle of Induction was part of the definition of the natural numbers. We mention that some work is needed to confirm that the natural numbers, defined in that way, really do amount to the same things as the natural numbers of our primitive intuition.

It can be shown that, assuming all the other Peano Axioms, then the last Peano Axiom becomes equivalent to the Well-Ordering Property; each implies the other. So Peano could equally as well have specified the Well-Ordering Property as the last of his axioms, instead of the Principle of Induction. This equivalence result is not really about the natural numbers. It is a result about any mathematical structure satisfying all the other Peano Axioms. It asserts that, for any such structure, the Principle of Induction is equivalent to the Well-Ordering Property.

Thus, when an introductory textbook, working simply from an intuitive description of the natural numbers, makes an attempt to establish a deductive relationship between the Principle of Induction and the Well-Ordering Property, the material does have some kind of cultural resonance, at least in the perception of those who are aware of the history. However, if the Peano Axioms are not discussed, then nothing remains that can stand up by itself. Then a student who may already be struggling with mixed messages about the concept of proof may be further misdirected.

Of course, all lectures and texts are potentially misleading. Ultimately, the only solid ground in mathematics is your own judgement, fallible as it is. If you abandon that, then there can be no such thing as mathematical proof anyway.

♣ ♡ ♢ ♠

Exercise 1.4.C: Let $(x_0, x_1, ...)$ be an infinite sequence such that $x_0 = 3$ and $x_{n+1} = 3x_n$ for all natural numbers n. Evaluate x_n.

Exercise 1.4.D: Show that $1 + z + z^2 + ... + z^{n-1} = (z^n - 1)/(z - 1)$ for all real numbers z with $z \neq 1$.

Exercise 1.4.E: For which natural numbers n do we have $n! \geq 2^n$?

Exercise 1.4.F: Show that $\frac{1}{2} \cdot \frac{3}{4} ... \frac{n+1}{2n} < \frac{1}{\sqrt{2n+1}}$ for all positive integers n.

Exercise 1.4.G: Show that, for all integers n with $n \geq 2$, we have

$$\left(1 - \frac{1}{2^2}\right)\left(1 - \frac{1}{3^2}\right) ... \left(1 - \frac{1}{n^2}\right) = \frac{n+1}{2n} .$$

Exercise 1.4.H: Show that, for all positive integers n, we have

$$1^3 + 2^3 + ... + n^3 = n^2(n+1)^2/4 = (1 + 2 + ... + n)^2 .$$

Exercise 1.4.I: For each natural number n, let F_n and L_n denote, respectively, the Fibonacci number and the Lucas number indexed by n. Show that

$$F_0^2 + F_1^2 + F_2^2 + ... + F_n^2 = F_n F_{n+1} , \qquad L_0^2 + L_1^2 + L_2^2 + ... + L_n^2 = L_n L_{n+1} + 2 .$$

1.5: Some harder applications of mathematical induction

We shall state some other versions of the Principle of Mathematical Induction, and we shall present some illustrative applications.

Another version of the Principle of Mathematical Induction: *For each natural number n, let P_n be a statement. Suppose that P_0 and P_1 hold and that, for each n, if P_n and P_{n+1} hold, then P_{n+2} holds. Then P_n holds for all natural numbers n.*

In other words, if P_0 and P_1 hold, if P_0 and P_1 together imply P_2, if P_1 and P_2 together imply P_3, if P_2 and P_3 together imply P_4, and so on, then the statements P_0, P_1, P_2, P_3, and so on, all hold. Of course, this version of the Principle of Mathematical Induction, again, is obvious. Being already obvious, it would not make sense to try to deduce it from the version of the principle that we stated in the previous section.

Exercise 1.5.A: *Show that the Fibonacci and Lucas numbers satisfy $5F_n = L_{n+1} + L_{n-1}$ and $L_n = F_{n+1} + F_{n-1}$ for all positive integers n.*

Response: Recall that $F_1 = F_2 = 1$ and that $L_0 = 2$ and $L_1 = 1$ and $L_2 = 3$ and $L_3 = 4$. Hence $5F_1 = 5 = L_2 + L_0$ and $5F_2 = 5 = L_3 + L_1$. Now suppose, inductively, that $n \geq 3$ and that $5F_{n-1} = L_n + L_{n-2}$ and $5F_{n-2} = L_{n-1} + L_{n-3}$. Then, as required,

$$5F_n = 5F_{n-1} + 5F_{n-2} = L_n + L_{n-1} + L_{n-2} + L_{n-3} = L_{n+1} + L_{n-1} \ .$$

Recall that $F_0 = 0$ and $F_3 = 2$. We have $L_1 = 1 = F_2 + F_0$ and $L_2 = 3 = F_3 + F_1$. By an inductive argument similar to the one in the previous paragraph, $L_n = F_{n+1} + F_{n-1}$ for all positive n. □

♣ ♡ ♢ ♠

The next version of the Principle of Mathematical Induction, sometimes referred to as the Strong Form, appears quite frequently. This version of the principle will be needed in the proof of the Euler Path Theorem in Chapter 3.

Strong Form of the Principle of Mathematical Induction: *For each natural number n, let P_n be a statement. Suppose that P_0 holds and that, for each positive n, if P_0, P_1, ..., P_{n-1} all hold, then P_n holds. Then P_n holds for all natural numbers n.*

In the game of nim, there several piles of matchs on a table. Two players take turns to remove matches. When a player takes her turn, she must remove at least one match, and all the matches she removes must come from the same pile. There are two versions of the game. In **normal nim**, the player who removes the last match wins. In **misère nim**, the player who removes the last match loses.

The mathematical theory of nim dates back to a paper by Charles Bouton, *Nim, a game with a complete mathematical theory*, published in 1901. In that paper, he gave descriptions and proofs of the winning strategies for both normal nim and misère nim. Bouton stated that the misère version was the one usually played. He also mentioned that the winning strategy for the misère version had been communicated to him, without proof, in 1899. The rules of a somewhat similar game were indicated in a manuscript by

Luca Pacioli on puzzles and magic tricks. The manuscript, unpublished until modern times, was completed in 1508 or earlier.

The origin of the name *nim* is unknown. In 17th century English slang, *to nim* means to *to steal, to pilfer*. Some cognates of that verb, in Germanic languages, mean *to take*. The terms *normal* and *misère* demand explanation too. They probably do not predate the rise of combinatorial game theory in the 1960s. Nim belongs to a class of two-player games for which the so-called *normal* version, with the last move winning, tends to be much easier to study than the so-called *misère* version, with the last move losing. The French noun *misère* means *destitution, wretchedness*.

Exercise 1.5.B: *Show that, in the game of nim, when there are exactly two piles of matches and the piles are of different sizes, the player with the first move will win if she plays correctly.*

Response: Let us include the case where one of the piles is empty, we mean to say, the case where one of the piles has zero matches. We argue by induction on the number of matches n in the smaller pile. The assertion holds in the case $n = 0$ because, in that case, the first player can win immediately by taking the whole of the non-empty pile. Now assume, inductively, that $n \neq 0$ and that the assertion holds whenever the smaller pile has less than n matches. Then the first player can take matches from the larger pile in such a way as to make two piles equal in size. After the second player has moved, the two piles will again be unequal in size. Since the smaller pile will now have less than n matches, the inductive assumption implies that the first player wins. □

♣ ♡ ♢ ♠

Under a continuous assessment regime, students are, in effect, discouraged from responding to exercises in their own way, their own style. Divergence from recommended methods and approaches tends to increase the liklihood of mistakes. In particular, continuous assessment may interfere with efforts to properly learn the technique of mathematical induction. Very soon after first encountering the technique, students may be required to apply it in assessed work. There might not be enough time for the student to find out why, when using the technique, definitions and assumptions have to be specified as well as conclusions.

For very good reasons, experienced mathematicians prefer definitions and assumptions to be clearly stated. Their experience informs them that, when definitions and assumptions are unstated or indicated only vaguely, the account is likely to be hard to follow, furthermore, the resulting conclusions are likely to be incorrect. But a beginner, especially a beginner undergoing training in disposal of easy routine questions, might nor yet have acquired the necessary background to fully appreciate why the reader cannot be relied upon to guess the definitions and assumptions from context.

The guidelines offered below might be useful, in the short run, for students who are to be tested on technique of mathematical induction before they have had time to properly assimilate it. We must confess, however, that these guidelines might be worse than useless in the long run. They may convey the false impression that the technique of mathematical induction is something like a ritual.

If you do make use of the guidelines below, then you ought to do so under the understanding that their whole purpose is just to make sure that, when you apply the technique, you will supply your reader with all the important information that she will need in order to understand what you are saying. The quality of a proof, of course, is not determined by its degree of conformity with some or other recommended recipe. The quality of a proof is is determined only by its correctness, clarity, brevity and aesthetic appeal.

An argument by mathematical induction always has two main deductive steps. In the **base step**, you must confirm that the required conclusion holds in some initial cases. In the **reduction step** you must deal with all the other cases by reducing them to some smaller cases.

The following scheme outlines just one way of presenting an induction argument.

• **Preliminary notation and other definitions:** If necessary, introduce appropriate notation or other definitions.
• **Base step:** Confirm that the required conclusion holds for some initial cases. The cases to be treated here are precisely those not treated by the reduction step below.
• **Begin the reduction step by stating the inductive assumption:** In preparation for treatment of an arbitrary case, specify which smaller cases are to be assumed.
• **Complete the reduction step by deducing a given case:** Using the inductive assumption, deduce the required conclusion in the arbitrary case under consideration.

Of course, the four parts of the argument have to be coordinated with each other. Usually, the core of the argument is in the fourth part. One has to anticipate how the fourth part wil run before one can finish the other three parts.

Let us illustrate the scheme with two examples.

Deconstruction of the response we gave to the first part of Exercise 1.4.B. Proof, by induction, that $F_0 + ... + F_n = F_{n+2} - 1$ *for all natural numbers* n.

Preliminary notation and other definitions: Let $S_n = F_0 + ... + F_n$.

Base Step: Clearly, $S_0 = F_0 = 0 = F_2 - 1$.

Begin the reduction step by stating the inductive assumption: The inductive assumption is that $n \geq 1$ and that $S_{n-1} = F_{n+1} - 1$.

Complete the reduction step by deducing a given case: By direct calculation, $S_n - S_{n-1} = F_n = F_{n+2} - F_{n+1} = F_{n+2} - 1 - (F_{n+1} - 1) = F_{n+2} - 1 - S_{n-1}$. Cancelling, we deduce that $S_n = F_{n+2} - 1$. □

Thus, in the reduction step, we showed that $S_n = F_{n+2} - 1$ for all positive n. Only the case $n = 0$ had to be treated in the base step.

Deconstruction of the response we gave to the first part of Exercise 1.5.A. Proof, by induction, that $5F_n = L_{n+1} + L_{n-1}$ *for all positive integers* n.

Preliminary notation and other definitions: (No preliminary definitions need to be introduced.)

Base Step: Recall that $F_1 = F_2 = 1$ and that $L_0 = 2$ and $L_1 = 1$ and $L_2 = 3$ and $L_3 = 4$. Hence $5F_1 = 5 = L_2 + L_0$ and $5F_2 = 5 = L_3 + L_1$.

Begin the reduction step by stating the Inductive assumption: The inductive assumption is that $n \geq 3$ and that $5F_{n-1} = L_n + L_{n-2}$ and $5F_{n-2} = L_{n-1} + L_{n-3}$.

Complete the reduction step by deducing a given case: Using the inductive assumption, $5F_n = 5F_{n-1} + 5F_{n-2} = L_n + L_{n-1} + L_{n-2} + L_{n-3} = L_{n+1} + L_{n-1}$. □

In this second illustration, the reduction step dealt with all cases except for the cases $n = 1$ and $n = 2$. Those two cases had to be treated separately in the base step.

♣ ♡ ♢ ♠

Very often, in a proof by mathematical induction, the components of the argument are not presented in the order listed above. Sometimes, the inductive assumption is specified only towards the end of the reduction step. Sometimes, the reduction step is treated before the base step.

In the more advanced literature, applications of the strong form of mathematical induction do not always deal with the base step separately. With careful phrasing, the base step can sometimes be incorporated into the rest of the argument by means of some literally vacuous logic. The statement "All flying pigs can talk" says that the class of flying pigs is contained in the class of talking animals. Let us suppose that no flying pigs exist. Then the class of flying pigs, being empty, must be contained in the class of talking animals. In other words, the statement "All flying pigs can talk" holds. The statement "If all flying pigs can talk, then the sea is boiling hot" is equivalent to the statement "The sea is boiling hot". Similarly, since there are no natural numbers less than zero, the statement P_0 is equivalent to the statement "If P_m holds for all natural numbers m less than 0, then P_0 holds". In view of these vacuous observations, the Principle of Mathematical Induction can be cunningly restated as follows.

Strong Form of the Principle of Mathematical Induction, reformulated: *For each natural number n, let P_n be a statement. Suppose that, for each natural number n, if P_m holds for all natural numbers m less than n, then P_n holds. Then P_n holds for all natural numbers n.*

The only advantage of this cunning manoeuvre is that it sometimes allows the length of an induction argument to be reduced by about one line. Let us reiterate the point that the condition on P_0 has not been removed. It has been incorporated into the general condition on P_n. Those of us who feel confused by the vacuous trick might do best to ignore it until they have acquired more experience. Anyway, we shall not be making use of it anywhere in this book. Of course, in all arguments by mathematical induction, the base step does have to be considered, one way or another.

Our response to the next exercise may help to reinforce the point that the guidelines above ought not to be memorized.

Exercise 1.5.C: *Let $(x_0, x_1, ...)$ and $(y_0, y_1, ...)$ be infinite sequences such that*

$$x_{n+3} = x_{n+2} + x_{n+1} + x_n \ , \qquad y_{n+3} = y_{n+2} + y_{n+1} + y_n$$

for all natural numbers n, also $x_0 = x_1 = 0$ and $x_2 = y_0 = y_1 = y_2 = 1$. Show that $y_n = x_{n+1} + x_{n-1}$ for all positive integers n.

Response: For each positive integer n, let $z_n = x_{n+1} + x_{n-1}$. We are to show that each $z_n = y_n$. First observe that

$$z_{n+3} = x_{n+4} + x_{n+2} = x_{n+3} + x_{n+2} + 2x_{n+1} + x_n + x_{n-1} = z_{n+2} + z_{n+1} + z_n \ .$$

Assuming, inductively, that $z_{n+2} = y_{n+2}$ and $z_{n+1} = y_{n+1}$ and $z_n = y_n$, we deduce that

$$z_{n+3} = y_{n+2} + y_{n+1} + y_n = y_{n+3} \ .$$

To complete the inductive argument, we note that, directly from the definition of z_n, we have $z_1 = 1 = y_1$ and $z_2 = 1 = y_2$ and $z_3 = 3 = y_3$. □

Nor would there be anything to be gained by memorizing various different statements of the Principle of Mathematical Induction. We have explicitly stated three or four versions of the principle: one of them in the previous section, the others in the present section. But that list is far from exhaustive. The version of the principle used in our response to the latest exercise was slightly different from all of the versions we have thus far stated. Sure enough, we could add the latest version to our list:

The Principle of Mathematical Induction, as used in Exercise 1.5.C: *For each natural number n, let P_n be a statement. Suppose that P_0 and P_1 and P_3 hold and that, for each n, if P_n and P_{n+1} and P_{n+2} hold, then P_{n+3} holds. Then P_n holds for all natural numbers n.*

Exercise 1.5.D: *Show that $L_{m+n} = L_m L_n - (-1)^n L_{m-n}$ for all natural numbers m and n with $m \geq n$.*

Response: Fixing m, defining $\Lambda_n = L_m L_n - (-1)^n L_{m-n}$, we shall show that $L_{m+n} = \Lambda_n$ for all integers n in the range $0 \leq n \leq m$. If $n = 0$ then $L_{m+n} = L_m = L_m L_0 - L_m = \Lambda_n$. If $n = 1$ then $m \geq 1$ and $L_{m+n} = L_m + L_{m-1} = L_m L_1 + L_{m-1} = \Lambda_{m,n}$. We have confirmed the required assertion in the case where n is 0 or 1. $n \leq 1$. Now suppose, inductively, that $n \geq 2$ and $L_{m+n-1} = \Lambda_{n-1}$ and $L_{m+n-2} = \Lambda_{n-2}$. Then

$$\begin{aligned} L_{m+n} &= L_{m+n-1} + L_{m+n-2} \\ &= L_m L_{n-1} - (-1)^{n-1} L_{m-(n-1)} + L_m L_{n-2} - (-1)^{n-2} L_{m-(n-2)} \\ &= L_m (L_{n-1} + L_{n-2}) - (-1)^n (L_{m-n+2} - L_{m-n+1}) = \Lambda_n \ . \end{aligned}$$ □

The latest exercise implicitly uses yet another version of the Principle of Mathematical Induction. Let us extend our list again.

The Principle of Mathematical Induction, as used in Exercise 1.5.D: *Let m be a natural number with $m \geq 2$. For each integer n in the range $0 \leq n \leq m$, let P_n be a statement. Suppose that P_0 and P_1 hold and that, for all n in the range $2 \leq n \leq m$, if P_{n-2} and P_{n-1} hold, then P_n holds. Then P_n holds for all integers n in the range $0 \leq n \leq m$.*

Evidently, further cataloging of diverse statements of the Principle of Mathematical Induction would be a tiresome and unrewarding task. Generally, one cannot grasp an

assertion by trying to memorize various different ways of expressing it. To catch onto an assertion, one must fully perceive its meaning. That way, when called upon to state it, one merely has to find a way of expressing the perceived meaning. Statements are not ideas. They are just flickering shadows of ideas. Trying to grasp at the shadows is futile.

If you are having genuine trouble catching onto the mathematical induction technique — getting low marks in assessed exercises is merely a social phenomenon, and does not count as genuine trouble — then maybe you have been focusing on the wrong kind of thing. In that case, perhaps the best course of action would be to throw away caution and to respond to the exercises in whatever way most appeals to you. After that, if you have still not caught hold of the idea, then you might try looking at the proof of Theorem 2.4.1, in the next chapter. That theorem may have enough content to afford some satisfying conceptual traction. Alternatively, you might try jumping ahead to Chapter 3, where the arguments by mathematical induction have a rather different flavour.

♣ ♡ ♢ ♠

Pedantic footnote: The name *mathematical induction* is a misnomer. Three of the major forms of reasoning recognized in the logic of science are *deductive reasoning* and *inductive reasoning* and *abductive reasoning*. All three kinds of reasoning play vital roles in the logic of mathematics. Inductive reasoning is crucial for making good decisions about which speculations to work on. Abductive reasoning is crucial for finding good speculations in the first place.

But all mathematical proofs, including those involving mathematical induction, lie entirely within the domain of deductive reasoning, and do not make any appeal to inductive or abductive reasoning. *Mathematical induction is a technique of deductive reasoning. It is not a technique of inductive reasoning.*

When a phrase such as "We argue by induction" appears in the text of a proof, one can understand the word *induction* to be a reference to *mathematical induction*. But, in other contexts, it may be best to refer to the technique by its full name, *mathematical induction*, to prevent it from being misunderstood as a reference to inductive reasoning.

See Appendix A for an discussion of the different roles played by deductive, inductive and abductive reasoning. In that appendix, we shall also make some historical comments to indicate how the misnomer arose.

♣ ♡ ♢ ♠

Exercise 1.5.E: Let x_0, x_1, ... be an infinite sequence of real numbers such that

$$x_n = x_0 + x_1 + ... + x_{n-2} + x_{n-1}$$

for all integers n such that $n \geq 2$. Find a formula, valid for $n \geq 2$, expressing x_n in terms of x_0 and x_1. (Hint: if at a loss, try calculating a few terms and guessing the formula. Of course, having guessed the formula, you must then try to prove it.)

Exercise 1.5.F: Show that the Fibonacci and Lucas numbers satisfy

$$2F_{m+n} = F_m L_n + F_n L_m$$

for all natural numbers m and n.

Exercise 1.5.G: Let $(x_0, x_1, x_2, ...)$ be an infinite sequence of integers such that $x_0 = 5$ and $x_1 = 14$ and

$$x_{n+2} - 5x_{n+1} + 6x_n = 0$$

for all natural numbers n. Show that $x_n = 2^n + 4 \,.\, 3^n$ for all n.

Exercise 1.5.H: Starting with a pile of n matches, where $n \geq 2$, we divide the pile into two smaller piles, then we repeatedly choose a pile and divide it into two smaller piles. After $n-1$ divisions, we produce n piles with one match each. Let a_i and b_i be the sizes of the two new piles obtained upon the i-th division. Show that

$$a_1 b_1 + a_2 b_2 + ... + a_n b_n = n(n-1)/2 \,.$$

(For example, if $n = 4$, we might divide the pile into two piles of sizes 2 and 2, then divide one of those two piles into piles of sizes 1 and 1, then divide the other pile of size 2 into piles of sizes 1 and 1. In this case, $a_1 b_1 + a_2 b_2 + a_3 b_3 = 2.2 + 1.1 + 1.1 = 6$. Alternatively, we might divide the original pile into piles with sizes 3 and 1, then divide the pile of size 3 into piles of sizes 2 and 1, then divide the pile of size 2 into piles of sizes 1 and 1. In this case, we again have $a_1 b_1 + a_2 b_2 + a_3 b_3 = 3.1 + 2.1 + 1.1 = 6$.)

1.6: A winning strategy for the game of nim

The material in this section is of a somewhat recreational nature, and it is also rather complicated. It can be skipped by those of us who are impatient to get on with material at the core of discrete mathematics. Some of the notions introduced in this section will reappear in later chapters but, when that happens, we shall introduce the notions again, in a self-contained way. Let us not force entertainment on the reluctant.

Two versions of the game of nim — the normal version, and the misère version — were described in the previous section. Recall, in the normal version, the taker of the last match wins. In the misère version, the taker of the last match loses. Both versions have much the same winning strategy. The winning moves for normal nim are the same as the winning moves for misère nim, except in the case where less than two piles have more than one match. To describe the winning strategy, we shall need the notion of a binary string.

We define a **binary string** to be a finite sequence $(a_{n-1}, a_{n-2}, ..., a_2, a_1, a_0)$ where n is a positive integer and, for each integer j in the range $0 \leq j \leq n-1$, the term a_j is the integer 0 or the integer 1. We usually write $(a_{n-1}, ..., a_1, a_0)$ more briefly as $a_{n-1}...a_1a_0$.

Given a binary string $a = (a_{n-1}, ...a_0)$, we call n the **length** of a. We call a_j the digit of a with **index** j. A little more briefly, we also call a_j the **j-indexed** digit of a.

An example of a binary string is 000101101110. It has length nine. Its 0-indexed digit is 0. Its 1-indexed digit, its 2-indexed digit and its 3-indexed digit are all 1. Its 4-indexed digit is 0.

When all the digits of a binary string a are 0, we call a the **zero string** with length n. For example, 0000 is the zero string with length 4.

We have said that, for a binary string, the possible values of the digits are the integers 0 and 1. Let us mention that, in abstract, it does not matter what the symbols 0 and 1 represent, just as long as they are two distinct things, $0 \neq 1$. In some applications, 0 and 1 represent the states of a switch, *off* and *on*. In other applications, they represent the logical values *false* and *true*. They could, if we wished, repesent the moral values *bad* and *good*. However, in the application below, it is convenient to understand that each digit of a binary string is either the integer 0 or the integer 1.

Let us adopt the convention that, whenever some symbol denotes a binary string, then the same symbol decorated with subscript j denotes the j-indexed digit of the binary string. For instance, given a binary string b with length n, then b_j is the j-indexed digit of b, in other words, $b = b_{n-1}b_{n-2}...b_1b_0$.

Many of us, probably, are familiar with the following conventional way of using binary strings to represent natural numbers. Given a positive integer n and a natural number x such that $x < 2^n$, we write $\text{bin}_n(x)$ to denote the binary string a of length n such that

$$x = 2^{n-1}a_{n-1} + 2^{n-2}a_{n-2} + ... + 2^2a_2 + 2a_1 + a_0 \ .$$

We call $\text{bin}_n(x)$ the **n-digit binary representation** of x.

Note that x and $\text{bin}_n(x)$ are two different things: x is a positive integer, whereas $\text{bin}_n(x)$ is a binary string. For example, suppose that x is the number of days in a leap year. Then the nine-digit binary representation of x is 101101110. The ten-digit binary representation of x is 0101101110. For comparison, we point out that the three-digit decimal representation of x is 366. The five-digit decimal representation of x is 00366.

The following operation on binary strings, called **addition**, has diverse applications. As we shall see in Chapter 6, it is of fundamental importance in coding theory. For a positive integer n, and two binary strings a and b with length n, we define the **sum** of a and b to be the binary string $s = a + b$ such that

$$s_j = \begin{cases} 1 & \text{if } a_j \neq b_j, \\ 0 & \text{if } a_j = b_j. \end{cases}$$

For example, $1100 + 0110 = 1010$.

Plainly, $a + b = b + a$. It is also easy to see that, given three binary strings a, b, c, all of the same length, then $(a + b) + c = a + (b + c)$. For example

$$(0010 + 1100) + 0101 = 1110 + 0101 = 1011 = 0010 + 1001 = 0010 + (1100 + 0101) \ .$$

So we can write $a+b+c$ unambiguously. The j-indexed digit of $a+b+c$ is 1 if and only if one or three of the terms a, b, c have j-index digit 1. More generally, there is a very straightforward way of evaluating the sum of several binary strings. Given binary strings a, b, c, d, e, f, g, all of the same length, then the j-indexed digit of $a+b+c+d+e+f+g$ is 1 if and only if an odd number of the terms a, b, c, d, e, f, g has j-indexed digit 1. As another little example, $0010 + 1100 + 0101 + 1101 = 0110$.

It may be worth reiterating our observation that natural numbers and binary strings are two different kinds of thing. There is, of course, a familiar operation on numbers called addition. We have just defined an operation on binary strings and we have called it addition. Those two addition operations are quite different from each other. However, provided we take care not to confuse natural numbers with binary strings, no ambiguity can arise. As a sum of natural numbers, $7+5 = 12$; the 4-digit binary strings representing 7 and 5 and 12 are 0111 and 0101 and 1100, but $0111+0101 \neq 1100$. As a sum of binary strings, $0111 + 0101 = 0010$, which is the 4-digit binary string representing 2, but $7 + 5 \neq 2$.

♣ ♡ ♢ ♠

To express a state of play in a game of nim, we number the piles of matches from 1 to m and we let x_i be the number of matches in the i-th pile. We say that the game is in **position** $(x_1, ..., x_m)$. It will be convenient to allow the possibility that $x_i = 0$ for some or all of the indices i. Of course, if $x_i = 0$ for all i, then there are no matches on the table and the game has already finished.

The position $(x_1, ..., x_m)$ is called a **winning position for normal nim** provided, in the normal version of the game, the next player to move can win against all possible play. The position is called a **losing position for normal nim** provided the other player can win against all possible play. It makes sense to deem the final position $(0, 0, ..., 0, 0)$

to be a losing position. Indeed, in that position, the player due to move next cannot make a move but, anyway, that player has already lost, the other player having already removed the last match. The notions of a **winning position for misère nim** and a **losing position for misère nim** are defined similarly, the final position understood to be a winning position for misère nim.

The next remark is obvious. For both versions of the game, it recursively characterizes the distinction between the winning positions and the losing positions.

Remark 1.6.1: *For normal nim or misère nim, every position is either a winning position or else a losing position. A position is a losing position if and only if all the positions that can result from it, after one move, are winning positions.*

Let us choose a positive integer n such that $x_i < 2^n$ for all i. In other words, all the piles have less than 2^n matches. The binary string

$$\text{nim}(x_1, x_2, ..., x_m) = \text{bin}_n(x_1) + \text{bin}(x_2) + ... + \text{bin}_n(x_m)$$

is called the **nim string** of the position $(x_1, ..., x_m)$. For example, putting $n = 4$, then

$$\text{nim}(2, 12, 5, 13) = 0010 + 1100 + 0101 + 1101 = 0110 \ .$$

The choice of n does not have any essential significance. If we were to replace n with a larger integer, then the effect would just be to append some leading zeros to all the binary strings that will come into consideration. Putting $n = 9$, then

$$\text{nim}(2, 12, 5, 13) = 000000010 + 000001100 + 000000101 + 000001101 = 000000110 \ .$$

Theorem 1.6.2: (Bouton's Theorem) *In normal nim, a position* $(x_1, ..., x_n)$ *is a losing position if and only if* $\text{nim}(x_1, ..., x_m)$ *is the zero string.*

In Exercise 1.6.B, we shall prove the theorem and we shall also prove that, when $\text{nim}(x_1, ..., x_n)$ is not the zero string, the following process yields a winning move in normal nim.

- Determine the binary strings $\text{bin}_n(x_i)$ and the nim string $z = \text{nim}(x_1, ..., x_m)$.
- Let j be the largest natural number less than n such that $z_j = 1$.
- Choose i such that the j-indexed digit of $\text{bin}_n(x_i)$ is 1.
- Calculate the natural number w represented by the binary string $z + \text{bin}_n(x_i)$.
- Remove $x_i - w$ matches from pile i, thus leaving w matches remaining there.

Let us give an example to illustrate the winning strategy for normal nim. In this example, the nim string at the start of the game is the zero string. The first player cannot apply the procedure, and it will be the second player who wins. We suppose that there are 4 piles of matches with sizes $x_1 = 4$ and $x_2 = 12$ and $x_3 = 5$ and $x_4 = 13$. The binary representations of those natural numbers are 0100, 1100, 0101, 1101, ordered respectively. The game proceeds as follows.

Move 1A: Alfred, the first player, calculates the nim string

$$\text{nim}(4, 12, 5, 13) = 0100 + 1100 + 0101 + 1101 = 0000 \ .$$

Since this is the zero string, he is unable to apply the above procedure. Lacking a strategy, he removes 2 matches from pile 1. The sizes of the piles become 2, 12, 5, 13.
Move 1B: Boudica, the second player, follows the procedure specified above. Let us present it in detail.
• After observing that $\mathrm{bin}_4(2) = 0010$ and $\mathrm{bin}_4(12) = 1100$ and $\mathrm{bin}_4(5) = 0101$ and $\mathrm{bin}_4(13) = 1101$, she calculates the nim string

$$z = \mathrm{nim}(2, 12, 5, 13) = 0010 + 1100 + 0101 + 1101 = 0110 \ .$$

• She notes that the largest integer j satisfying $z_j = 1$ is $j = 2$
• She can choose i to be 2 or 3 or 4 because the binary strings 1100 and 0101 and 1101 all have 2-indexed digit 1. Let us say she chooses i to be 3.
• She evaluates the binary string $\mathrm{bin}_4(x_3) + z = 0101 + 0110 = 0011$ and observes that it represents the integer $w = 3$.
• She removes $5 - 3 = 2$ matches from pile 3. The sizes of the piles are now 2, 12, 3, 13.
Move 2A: Unfortunately for Alfred, the procedure is again inapplicable, because the nim string is again the zero string

$$\mathrm{nim}(2, 13, 3, 13) = 0010 + 1100 + 0101 + 1101 = 0000 \ .$$

Forlornly, he removes 6 matches from pile 2, leaving piles with sizes 2, 6, 3, 13.
Move 2B: For Boudica, the nim string is now

$$\mathrm{nim}(2, 6, 3, 13) = 0010 + 0110 + 0011 + 0111 = 1010 \ .$$

Putting $i = 4$, noting that the binary string $w = 1101 + 1010 = 0111$ represents the integer 7, she removes $13 - 7 = 6$ matches from pile 4, leaving piles with sizes 2, 6, 3, 7.
Move 3A: The nim string is again the zero string. Stoically, Alfred removes 3 matches from pile 4, leaving piles with sizes 2, 6, 3, 4.
Move 3B: Boudica removes all 3 matches from pile 3 because

$$\mathrm{nim}(2, 6, 3, 4) = 0010 + 0110 + 0011 + 0100 = 0011 \ .$$

The piles now have sizes 2, 6, 0, 4.
Move 4A: Alfred, again faced with a nim string equal to the zero string, removes 5 from pile 2, leaving sizes 2, 1, 0, 4.
Move 4B: Boudica removes 1 match from pile 4, leaving sizes 2, 1, 0, 3.
Move 5A: Alfred removes 1 match from pile 1, leaving sizes 1, 1, 0, 3.
Move 5B: Boudica removes all 3 matches from pile 4, leaving sizes 1, 1, 0, 0.
Move 6A: Alfred removes a penultimate match.
Move 6B: Boudica, in victory, removes the last one.

We neglected to record Boudica's calculations from Move 4B onwards. We shall rectify that omission in Exercise 1.6.A.

The winning strategy for the misère version of the game is the same, except that, when there is exactly one pile with more than one match, the unique winning move is to remove matches from that pile in such a way as to ensure that an odd number of matches

remain, no two of them in the same pile. The moves thereafter are forced. Boudica's moves, above, would still be correct for the misère version, except that, on Move 5B, she would remove only 2 matches from pile 4, leaving sizes 1, 1, 0, 1. Alfred would then inevitably remove the last match on Move 7A.

♣ ♡ ♢ ♠

Pedantic footnote: FISH Roland Sprague 1935, Patrick Grundy 1939, independent. How much should we say about the Sprague–Grundy Theorem. How much about the rise of combinatorial game theory in the 1960s? Shall we present that theorem in Chapter 3? ENDFISH.

♣ ♡ ♢ ♠

Exercise 1.6.A: Check that, in the nim game recorded above, Moves 4B and 5B conform with the specified procedure.

Exercise 1.6.B ♠: Prove that, when a position $(x_1, ..., x_n)$ has non-zero nim string, the process described above can be carried out and results in a position whose nim string is the zero string. Prove that Bouton's Theorem holds and that, for normal nim, if the initial position has non-zero nim string, then the process yields a win for the next player to move. (Hint: If you have not yet grasped how the process works, apply it to some examples until understanding dawns. If you do see how the process works, but cannot figure out what needs to be proved, note that, for instance, it is not obvious that w is less than x_i. For Bouton's Theorem, argue by induction as in Exercise 1.5.B.)

Exercise 1.6.C: In misère nim, which positions $(x_1, ..., x_m)$ are losing positions?

Exercise 1.6.D ♠: In normal 3-subtraction nim, a further rule is imposed, namely, that at most 3 matches can be removed on each move. As before, at least one match must be removed, and all the removed matches must be from the same pile. Which positions $(x_1, ..., x_m)$ are the losing positions? What is the winning strategy? (Hint: Very often, to solve general problems, it is helpful to consider special or particular cases. You might first consider the particular case where there is only 1 match. Having analysed that, you might move on to some other small particular cases, say, where there are 2 or 3 or 4 or 5 matches, all of them in the same pile. You might then be able to answer the questions in the special case where there is only one non-empty pile. After that, you might advance to the case where there are exactly two piles. Of course, in your response to the exercise, you need not record all the exploratory work you did to find the answers. You need only supply answers and proofs.)

Responses to Exercises in Chapter 1

Nearly always, good responses to exercises include proofs. Sometimes, the conclusion to be proved is already stated in the exercise. Sometimes, the exercise is in the form of a question. Nearly always, a question raised in exercise has only one correct answer, though the correct answer can often be concisely expressed in more than one way. When the correct answer is not obvious, it ought to be justified with a proof. Usually, correct assertions have more than one good proof. Always, any proof can be expressed in various different ways.

So the only exercises that have just one good response are those which are in the form of a question whose answer is obvious and expressible in only one way. Such exercises are vanishingly rare. Almost always, an exercise has many different good responses.

After doing some exercises, and after comparing your responses to mine, you might decide that you wish to emulate some features of my style. That would be okay. Or, if you dislike every feature of the responses recorded below, then maybe you will take better to some other styles found somewhere else. As one studies mathematics, one does pick up stylistic features from others. Of course, you must apply your own judgement.

Never compose a response to an excercise except in the belief that you fully understand what you are writing. You will get nowhere, or worse than nowhere, if you imitate without comprehension. In an exam script, nothing is more pathetic than an incoherent salad of remembered bits and pieces of text, apparently written down in the desperate hope that they may be relevant somehow. Fakery in mathematics, as in all sciences, is anathema.

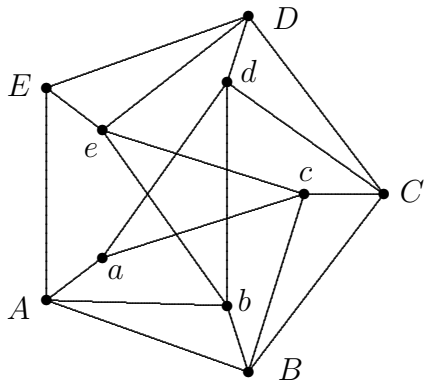
There are no strange tacit rules about how to respond to exercises. You can treat an exercise simply as a request that has has been made by another person, say, in a letter. You can respond just as you would in a letter of reply, bearing in mind that you are writing to someone who is exclusively interested with the resolution of the problem raised in the exercise. If she asked for a proof of a stated assertion, then you are to supply a proof. If she asked a question, then you are to supply the correct answer and, if the correct answer is not obvious, you must also supply a proof.

Let us mention that, in the pedagogical literature, a good response to an exercise is sometimes called an *answer* or a *solution*. That terminology is potentially misdirecting, both of those words having several different meanings in ordinary English. An exercise having the form "Does the problem have a *solution*?" might have correct *answer* "No" or "Yes". A good *response* to such an exercise might be to state the correct answer and then to explain, very clearly and deductively, why the stated answer is correct. Even if the answer is "yes", it may be possible to prove it without actually finding a solution.

Response to 1.1.C: Recall that the 64 squares of a chessboard can be coloured in such a way that each square is black or white, each square coloured differently from its 2 or 3 or 4 nearest neighbours. Half of the squares are black, the other half white. The two removed squares are of the same colour. So the mutilated board has more squares of one colour than of the other colour. But each dominoes piece would cover one black square and one white square. So the answer to the question is: No. □

Comment: The answer "No", without any justification, would be of no help at all to your reader, since it would not supply her with any good reason for accepting that your answer is correct.

Response to 1.1.D: Part (1). Adding 3 edges to the given graph yields a new graph with at least 4 edges of odd degree. By easy half of the Euler Path Theorem, the new graph has no path as specified. Part (1) is established. Part (3) can be demonstrated similarly, noting that the addition of 4 edges yields a graph with at least 2 edges with odd degree. For part (2), we add edges as indicated in the diagram. Labelling the vertices a, ..., e, A, ..., E as indicated, there is a path from a to E visiting, in order, the vertices: a, d, b, e, c, a, A, b, B, c, C, d, D, e, E D, C, B, A, E.



Comment: For part (2), we cannot apply Proposition 1.1.1, the easy half of the Euler Path Theorem. That proposition gives a necessary condition for an Euler path, we mean to say, a condition whic must hold if an Euler path exists. But the proposition does not give a sufficient condition, we mean, a condition which guarantees the existence of an Euler path.

The hard half of the Euler Path Theorem, stated in the next section and proved in Chapter 3, does give a sufficient condition. But we cannot make use of that here, because we have not proved that part of the theorem.

Response to 1.1.E: Cutting each edge at the middle, so that each half-edge is now attached to a single vertex, then the number of half-edges attached to each vertex x is equal to the degree of x. But the total number of half-edges is even. So there must be an even number of vertices x such that x has an odd number of half-edges attached to it. □

Comment: In Chapter 3, when we have just a bit more graph theory behind us, we shall be able to give a smoother response to this exercise. See Exercise 3.2.X.

Response to 1.3.C: For a contradiction, suppose that each troll receives a different number of rocks. The troll with the least number of rocks has at least 0 rocks. The troll with the second least has at least 1. The troll with the third least has at least 2, and so on. The troll with the most rocks has at least 14. So the total number of rocks must be at least $0 + 1 + 2 + ... + 13 + 14 = 7.15 = 105$. This contradicts the condition that there are only 100 rocks to distribute. □

Response to 1.3.D: We make use of the fact that, given an integer a such that a^2 is divisible by 3, then a is divisible by 3. Suppose, for a contradiction, that $\sqrt{15}$ is rational. Let a and b be positive integers such that $a/b = \sqrt{15}$ and a is as small as possible. Then $a^2 = 15b^2$ which is divisible by 3. So $a = 3\alpha$ for some positive integer α. It follows that $3\alpha^2 = 5b^2$, which is divisble by 3. So $b = 3\beta$ for some positive integer β. We deduce that $\alpha^2 = 15\beta^2$, in other words, $\alpha/\beta = \sqrt{15}$. But α is smaller than β. This contradicts the assumption that a is as small as possible. We have shown that $\sqrt{15}$ is irrational.

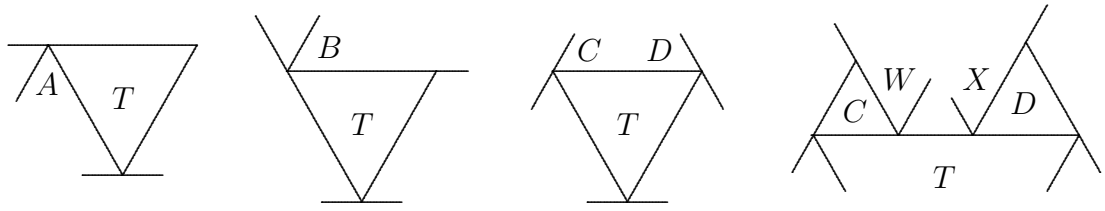
Now suppose, for a contradiction again, that $\sqrt{3} + \sqrt{5} = c/d$ for positive integers c and d. Then $8 + 2\sqrt{15} = (\sqrt{3} + \sqrt{5})2 = c^2/d^2$. Hence $\sqrt{15} = (c^2 - 8d^2)/2d^2$. But that contradicts the irrationality of $\sqrt{15}$. We conclude that $\sqrt{3} + \sqrt{15}$ is irrational. □

Response to 1.3.E: No, such a tour is impossible. For a contradiction, suppose such a tour exists. Let U, D, L, R be the number of moves, from one cell to the next one, in the upwards, downwards, leftwards, rightwards directions, respectively. Then $U - D = 4$. So U and D and either both odd or else both even. It follows that $U + D$ is even. We also have $L = R$, so $L + R$ is even. But the total number of moves made in the tour is $U + D + L + R = 49$, which is odd. This is a contradiction, as required. □

Comment: Notwithstanding a superficial resemblence to Exercise 1.1.B, we did not make use of any theory discussed in the text. Perhaps it is wicked of an author to discuss some theory and then to set an exercise which deceptively looks like an application. Still, at least in research, likely looking pieces of theory do sometimes turn out to represent quite the wrong approach.

Response to 1.3.F: Suppose, for a contradiction, that an equilateral triangle can be cut up as specified. Let T be the smallest triangle with the abutting property described in the hint. We may assume that T is oriented with a horizontal edge at the top, the bottom vertex abutting against edge.

Consider only the lines formed by the edges of the triangles. The top edge of T does not extend leftwards, as in the first diagram below, because otherwise the triangle A would have the abutting property, yet A would be smaller than T, contradicting the minimality of T. Similarly, the top edge of T does not extend rightwards.

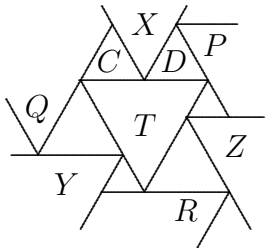


The left edge of T does not extend upwards, as in the second diagram, because then the triangle B, having the abutting property, would be larger than T, forcing the top edge of T to extend rightwards, which we have already shown is impossible. Similarly, the right edge of T does not extend upwards.

We deduce that all three vertices of T abut an edge. Triangle C, in the third diagram, is distinct from triangle D, because otherwise the triangle $C = D$ would have the same size as T, contradicting the condition that no two triangles have the same size.

The triangles C and D must meet at a vertex because otherwise the triangles W and X, depicted in the fourth diagram, must be distinct, yet both W and X have the abutting property and must be larger than T, plainly an impossibility.

Applying the same argument to the other two edges of T, we deduce that T shares and edge with exactly six other triangles. Furthermore, the edges of T are abutted by three triangles X, Y, Z, as depicted below. By the minimality of T, the three triangles X, Y, Z are all larger than T and, perforce, larger than the six triangles sharing an edge with T. So there exist triangles P, Q, R abutting X, Y, Z, respectively. The sum of the edge-lengths of P and Q and R is less than or equal to the perimeter of T. In other words, the average of those three edge-lengths is less than or equal to the edge-length of T. Since P, Q, R are of different sizes, one of them must be smaller than T. But P, Q, R have the abutting property. This contradicts the minimality of T. □



Comment: Perhaps the latest exercise does not really belong to discrete mathematics. Nor, perhaps, does it belong to any other area or branch of mathematics. Still, it does seem to nicely illustrate the vandalistic character of argument by contradiction.

Response to 1.4.C: Plainly, $x_n = 3^{n+1}$. □

Comment: Note that, following the rules of style indicated in this section, our response is a sentence, and it begins with a word. Instead of just writing "$x_n = 3^{n+1}$", we wrote "Plainly, $x_n = 3^{n+1}$."

If at a loss to think of a way to begin a sentence with a word rather than a symbol, the phrase "We have" can come in useful. We could have written "We have $x_n = 3^{n+1}$". But, since proofs are required to be succinct, it is better, when possible, to select a word or phrase that might communicate some helpful guidance

Very frequently, when we are reading a mathematical text, we get stuck trying to figure out why a stated assertion is correct. In fact, when the text is a difficult one, we may be struggling with a long list of enigmatic assertions which, we hope, will sooner or later achieve resolution as we penetrate more deeply into the material. Often, though, an assertion that has baffled us eventually turns out to have been obvious, our ineptitude having arisen only because we were looking at it in an unproductive way.

So it can sometimes be genuinely helpful to inform a reader that an stated assertion is obvious. That way, if she is having any trouble with it, then she knows that, instead of searching hard for some clever trick, she just needs to snap out of her mental block and look at the assertion afresh. The words *plainly*, *clearly*, *obviously* are so closely synonymous with each other that they can be used interchangeably, though one of them may have a more appropriate nuance than the other two in some contexts.

Response to 1.4.D: Let $S = 1 + z + z^2 + ... + z^{n-1}$. Then $zS = z + z^2 + ... + z^{n-1} + z^n$. Therefore $(z-1)S = z^n - 1$. $\square$

Comment: The conclusion could also be obtained using mathematical induction, but the above argument is more straightforward.

Response 1 to 1.4.E: Clearly, the condition holds if and only if $n = 0$ or $n \geq 4$. $\square$

Response 2 to 1.4.E: We shall show, by induction, that condition holds when $n = 0$ or $n \geq 4$. The case $n \leq 4$ is clear. Supposing that $n \geq 4$ and $n! \geq 2^n$, then $(n+1)! > 2^{n+1}$ because $(n+1)!/n! = n + 1 > 2 = 2^{n+1}/2^n$. $\square$

Comment: In principle, one cannot approve of both of the above responses. If the required conclusion is obvious, then Response 1 must be satisfactory and Response 2 must be just a non-deductive list of trivia. On the other hand, if the required conclusion is not obvious, then Response 1 must be inadequate, and Response 2 would seem to be acceptable. In practise, when we are unsure about such matters, we just make a decision one way or the other.

Our equivocal verdict on the need for proof, here, does not imply any inconsistency in the concept of proof. As we indicated in this section, mathematical proof is not some kind of compulsive ritual for agonizing over very easy problems. Mathematical proof is substantially useful only when the content is difficult.

Response 1 to 1.4.F: FISH.

Response 1 to 1.4.G: FISH.

Response 1 to 1.4.H: The second equality is clear. It remains to show that $S_n = T_n$, where $S_n = 1^3 + 2^3 + ... + n^3$ and $T_n = n^2(n+1)^2/4$. We argue by induction. First note that $S_1 = 1 = T_1$. Supposing, now, that $n \geq 2$ and $S_{n-1} = T_{n-1}$, then

$$4(T_n - T_{n-1}) = n^2((n+1)^2 - (n-1)^2) = 4n^3 = 4(S_n - S_{n-1}) \ .$$

Dividing by 4 and adding S_{n-1}, we obtain $S_n = T_n$, as required. $\square$

Response 2 to 1.4.H: The second equality holds because

$$1 + 2 + ... + n = n(n+1)/2 \ .$$

Let $S_n = 1^3 + 2^3 + ... + n^3$ and $T_n = (1 + 2 + ... + n)^2$. We are to show that $S_n = T_n$ for all integers $n \geq 1$. The case $n = 1$ is trivial. Suppose, inductively, that $n \geq 2$ and $S_{n-1} = T_{n-1}$. We have

$$T_n - T_{n-1} = n^2 + 2n(1 + 2 + ... + (n-2) + (n-1)) \ .$$

But $1 + 2 + ... + (n-2) + (n-1) = n(n-1)/2$. Therefore

$$T_n - T_{n-1} = n^2 - n^2(n-1) = n^3 = S_n - S_{n-1} \ .$$

Adding the integer $S_{n-1} = T_{n-1}$ to both sides, we deduce that $S_n = T_n$, as required. $\square$

Response to 1.4.I: Let $X_n = F_0^2 + ... + F_n^2$ and $Y_n = F_n F_{n-1}$. We are to show, by induction, that $X_n = Y_n$ for all natural numbers n. The case $n = 0$ is clear. For $n \geq 1$, we have

$$X_n - X_{n-1} = F_n^2 = F_n(F_{n+1} - F_{n-1}) = Y_n - Y_{n-1} \ .$$

Inductively assuming that $X_{n-1} = Y_{n-1}$, we deduce that $X_n = Y_n$. The required formula for the Fibonacci numbers is now established. The required formula for the Lucas numbers can be demonstrated similarly. □

Response 1 to 1.5.E: For $n \geq 3$, we have $x_n = 2x_{n-1}$. It follows that, for $n \geq 2$, we have $x_n = 2^{n-2}x_2 = 2^{n-1}(x_0 + x_1)$. □

Response 2 to 1.5.E: We argue by induction on n. The case $n = 2$ is clear. Suppose that $n \geq 3$ and that $x_m = 2^{m-1}(x_0 + x_1)$ for all integers m in the range $2 \leq m < n$. Then $x_2 + x_3 + ... + x_{n-1} = (1 + 2 + ... + 2^{n-2})(x_0 + x_1) = (2^{n-1} - 1)(x_0 + x_1)$, hence $x_n = x_0 + x_1 + (2^{n-1} - 1)(x_0 + x_1) = 2^{n-1}(x_0 + x_1)$. □

Comment: The first proof, being shorter and much easier, is better than the second. Still, the second proof is valid and might reasonably be produced by someone who has not found the first proof.

Response to 1.5.F: FISH.

Response to 1.5.G: FISH.

Response to 1.5.H: FISH.

Response to 1.6.A: FISH.

Response to 1.6.B: We employ the notation that appeared in the description of the process. Since the binary string $z = \text{nim}(x_1, ..., x_n)$ is non-zero, there must exist a largest positive integer j such that $1 \leq j \leq n$ and $z_j \neq 0$. Since z is the sum of the binary strings $\text{bin}_n(x_i)$, there must exist some i such that the binary string $a = \text{bin}_n(x_i)$ has j-indexed digit $a_j = 1$. Let $b = z + a$. Since $b_j = 0$ and $a_j = 1$ and $b_k = a_k$ for all $k > j$, the natural number w represented by b must be less than x_i. That is to say, the positive integer $x_i - w$ is less than or equal to x_i. It is therefore possible to remove $x_i - w$ matches from pile i. We have shown that the process can be carried out.

After removing the matches, the new position is the same as the initial position except that the term x_i has been replaced by w. Let z' denote the nim string of the new position. We are to show that z' is the zero string. Recall, the nim string of the initial position is the sum $z = \text{bin}_n(x_1) + ... + \text{bin}_n(x_m)$. The nim string z' can be calculated as a sum in a similar way, the only change being that the term $a = \text{bin}_n(x_i)$ is replaced by b. Since $b = z + a$, we have $z' = z + z$, which is the zero string.

Now let us demonstrate Bouton's Theorem. We are to show that z is the zero string if and only if the position $(x_1, ..., x_n)$ is a losing position for normal nim. We argue by induction on the number of matches $x_1 + ... + x_m$. When there are no matches, the required conclusion is trivial. Now assume that there is at least one match and that the required conclusion holds for all positions with fewer matches.

Any move changes the nim string. Supposing that z is the zero string, then any

move will result in a position with non-zero nim string. Since the possible resulting positions have fewer matches, the inductive assumption implies that all of those positions are winning positions. In view of Remark 1.6.1, we deduce that the initial position $(x_1, ..., x_m)$ is a losing position.

Conversely, supposing that z is non-zero then, by applying the process discussed above, the next player can produce a new position whose nim string is the zero string. Since the new position has fewer matches, the inductive assumption implies that the new position is a losing position. In view of Remark 1.6.1 again, we deduce thaty $(x_1, ..., x_m)$ is a winning position.

The demonstration of the theorem is complete. The above argument also shows that application of the process is a winning strategy. □

Response to 1.6.C: FISH.

Response to 1.6.D: FISH.

foxcat