Computational Complexity:

A Conceptual Perspective

Oded Goldreich

Department of Computer Science and Applied Mathematics
Weizmann Institute of Science, Rehovot, Israel.

September 16, 2006

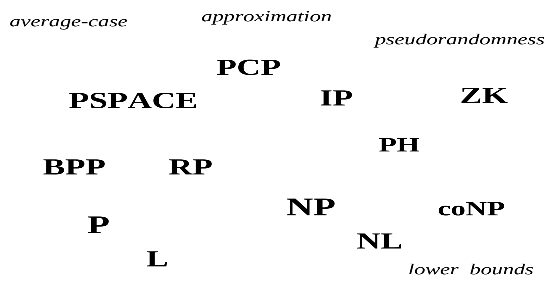
Chapter 1

Introduction and Preliminaries

You can start by putting the DO NOT DISTURB *sign.*

Cay, in *Desert Hearts* (1985).

The current chapter consists of two parts. The first part provides a high-level introduction to complexity theory. This introduction is much more detailed than the laconic statements made in the preface, but is quite sparse when compared to the richness of the field. In addition, the introduction contains several important comments regarding the contents, approach and style of the current book.



The second part of this chapter provides the necessary preliminaries to the rest of the book. It includes a discussion of computational tasks and computational models, as well as natural complexity measures associated with the latter. More specifically, this part recalls the basic notions and results of computability theory (including the definition of Turing machines, some undecidability results, the notion of universal machines, and the definition of oracle machines). In addition, this part presents the basic notions underlying non-uniform models of computation (like Boolean circuits).

1.1 Introduction

This section consists of two parts: the first part refers to the area itself, whereas the second part refers to the current book. The first part provides a brief overview of Complexity Theory (Section 1.1.1) as well as some reflections about its characteristics (Section 1.1.2). The second part describes the contents of this book (Section 1.1.3), the considerations underlying the choice of topics as well as the way they are presented (Section 1.1.4), and various notations and conventions (Section 1.1.5).

1.1.1 A brief overview of Complexity Theory

Out of the tough came forth sweetness[1]

Judges, 14:14

Complexity Theory is concerned with the study of the *intrinsic complexity* of computational tasks. Its "final" goals include the determination of the complexity of any well-defined task. Additional "final" goals include obtaining an understanding of the relations between various computational phenomena (e.g., relating one fact regarding computational complexity to another). Indeed, we may say that the former type of goals is concerned with *absolute* answers regarding specific computational phenomena, whereas the latter type is concerned with questions regarding the *relation* between computational phenomena.

Interestingly, the current success of Complexity Theory in coping with the latter type of goals has been more significant. In fact, the failure to resolve questions of the "absolute" type, led to the flourishing of methods for coping with questions of the "relative" type. Putting aside for a moment the frustration caused by the failure, we must admit that there is something fascinating in the success: in some sense, establishing relations between phenomena is more revealing than making statements about each phenomenon. Indeed, the first example that comes to mind is the theory of NP-completeness. Let us consider this theory, for a moment, from the perspective of these two types of goals.

Complexity theory has failed to determine the intrinsic complexity of tasks such as finding a satisfying assignment to a given (satisfiable) propositional formula or finding a 3-coloring of a given (3-colorable) graph. But it has established that these two seemingly different computational tasks are in some sense the same (or, more precisely, are computationally equivalent). The author finds this success amazing and exciting, and hopes that the reader shares his feeling. The same feeling of wonder and excitement is generated by many of the other discoveries of Complexity theory. Indeed, the reader is invited to join a fast tour of some of the other questions and answers that make up the field of Complexity theory.

We will indeed start with the "P versus NP Question". Our daily experience is that it is harder to solve a problem than it is to check the correctness of a solution

[1]The quote is commonly used to mean that benefit arose out of misfortune.

(e.g., think of either a puzzle or a research problem). Is this experience merely a coincidence or does it represent a fundamental fact of life (or a property of the world)? Could you imagine a world in which solving any problem is not significantly harder than checking a solution to it? Would the term "solving a problem" not lose its meaning in such a hypothetical (and impossible in our opinion) world? The denial of the plausibility of such a hypothetical world (in which "solving" is not harder than "checking") is what "P different from NP" actually means, where P represents tasks that are efficiently solvable and NP represents tasks for which solutions can be efficiently checked.

The mathematically (or theoretically) inclined reader may also consider the task of proving theorems versus the task of verifying the validity of proofs. Indeed, finding proofs is a special type of the aforementioned task of "solving a problem" (and verifying the validity of proofs is a corresponding case of checking correctness). Again, "P different from NP" means that there are theorems that are harder to prove than to be convinced of their correctness when presented with a proof. This means that the notion of a proof is meaningful (i.e., that proofs do help when trying to be convinced of the correctness of assertions). Here NP represents sets of assertions that can be efficiently verified with the help of adequate proofs, and P represents sets of assertions that can be efficiently verified from scratch (i.e., without proofs).

In light of the foregoing discussion it is clear that the P-versus-NP Question is a fundamental scientific question of far-reaching consequences. The fact that this question seems beyond our current reach led to the development of the theory of NP-completeness. Loosely speaking, this theory identifies a set of computational problems that are as hard as NP. That is, the fate of the P-versus-NP Question lies with each of these problems: if any of these problems is easy to solve then so are all problems in NP. Thus, showing that a problem is NP-complete provides evidence to its intractability (assuming, of course, "P different than NP"). Indeed, demonstrating NP-completeness of computational tasks is a central tool in indicating hardness of natural computational problems, and it has been used extensively both in computer science and in other disciplines. NP-completeness indicates not only the conjectured intractability of a problem but rather also its "richness" in the sense that the problem is rich enough to "encode" any other problem in NP. The use of the term "encoding" is justified by the exact meaning of NP-completeness, which in turn is based on establishing relations between different computational problems (without referring to their "absolute" complexity).

The foregoing discussion of the P-versus-NP Question also hints to *the importance of representation*, a phenomenon that is central to complexity theory. In general, complexity theory is concerned with problems the solutions of which are implicit in the problem's statement (or rather in the instance). That is, the problem (or rather its instance) contains all necessary information, and one merely needs to process this information in order to supply the answer.[2] Thus, complexity theory is

[2]In contrast, in other disciplines, solving a problem may require gathering information that is not available in the problem's statement. This information may either be available from auxiliary (past) records or be obtained by conducting new experiments.

concerned with manipulation of information, and its transformation from one representation (in which the information is given) to another representation (which is the one desired). Indeed, a solution to a computational problem is merely a different representation of the information given; that is, a representation in which the answer is explicit rather than implicit. For example, the answer to the question of whether or not a given Boolean formula is satisfiable is implicit in the formula itself (but the task is to make the answer explicit). Thus, complexity theory clarifies a central issue regarding representation; that is, the distinction between what is explicit and what is implicit in a representation. Furthermore, it even suggests a quantification of the level of non-explicitness.

In general, complexity theory provides new viewpoints on various phenomena that were considered also by past thinkers. Examples include the aforementioned concepts of proofs and representation as well as concepts like randomness, knowledge, interaction, secrecy and learning. We next discuss some of these concepts and the perspective offered by complexity theory.

The concept of *randomness* has puzzled thinkers for ages. Their perspective can be described as ontological: they asked "what is randomness" and wondered whether it exist at all (or is the world deterministic). The perspective of complexity theory is behavioristic: it is based on defining objects as equivalent if they cannot be told apart by any efficient procedure. That is, a coin toss is (defined to be) "random" (even if one believes that the universe is deterministic) if it is infeasible to predict the coin's outcome. Likewise, a string (or a distribution of strings) is "random" if it is infeasible to distinguish it from the uniform distribution (regardless of whether or not one can generate the latter). Interestingly, randomness (or rather pseudorandomness) defined this way is efficiently expandable; that is, under a reasonable complexity assumption (to be discussed next), short pseudorandom strings can be deterministically expanded into long pseudorandom strings. Indeed, it turns out that randomness is intimately related to intractability. Firstly, note that the very definition of pseudorandomness refers to intractability (i.e., the infeasibility of distinguishing a pseudorandomness object from a uniformly distributed object). Secondly, as hinted above, a complexity assumption that refers to the existence of functions that are easy to evaluate but hard to invert (called *one-way functions*) imply the existence of deterministic programs (called *pseudorandom generators*) that stretch short random seeds into long pseudorandom sequences. In fact, it turns out that the existence of pseudorandom generators is equivalent to the existence of one-way functions.

Complexity theory offers its own perspective on the concept of *knowledge* (and distinguishes it from information). It views knowledge as the result of a hard computation. Thus, whatever can be efficiently done by anyone is not considered knowledge. In particular, the result of an easy computation applied to publicly available information is not considered knowledge. In contrast, the value of a hard to compute function applied to publicly available information is knowledge, and if somebody provides you with such a value then it has provided you with knowledge. This discussion is related to the notion of *zero-knowledge* interactions, which are interactions in which no knowledge is gained. Such interactions may still be useful, because they may assert the *correctness* of specific data that was provided beforehand.

The foregoing paragraph has explicitly referred to *interaction*. It has pointed one possible motivation for interaction: gaining knowledge. It turns out that interaction may help in a variety of other contexts. For example, it may be easier to verify an assertion when allowed to interact with a prover rather than when reading a proof. Put differently, interaction with some teacher may be more beneficial than reading any book. We comment that the added power of such *interactive proofs* is rooted in their being randomized (i.e., the verification procedure is randomized), because if the verifier's questions can be determined beforehand then the prover may just provide the transcript of the interaction as a traditional written proof.

Another concept related to knowledge is that of *secrecy*: knowledge is something that one party has while another party does not have (and cannot feasibly obtain by itself) – thus, in some sense knowledge is a secret. In general, complexity theory is related to *Cryptography*, where the latter is broadly defined as the study of systems that are easy to use but hard to abuse. Typically, such systems involve secrets, randomness and interaction as well as a complexity gap between the ease of proper usage and the infeasibility of causing the system to deviate from its prescribed behavior. Thus, much of Cryptography is based on complexity theoretic assumptions and its results are typically transformations of relatively simple computational primitives (e.g., one-way functions) into more complex cryptographic applications (e.g., a secure encryption scheme).

We have already mentioned the context of *learning* when referring to learning from a teacher versus learning from a book. Recall that complexity theory provides evidence to the advantage of the former. This is in the context of gaining knowledge about publicly available information. In contrast, computational learning theory is concerned with learning objects that are only partially available to the learner (i.e., learning a function based on its value at a few random locations or even at locations chosen by the learner). Complexity theory sheds light on the intrinsic limitations of learning (in this sense).

Complexity theory deals with a variety of computational tasks. We have already mentioned two fundamental types of tasks: *searching for solutions* (or "finding solutions") and *making decisions* (e.g., regarding the validity of assertion). We have also hinted that in some cases these two types of tasks can be related. Now we consider two additional types of tasks: *counting the number of solutions* and *generating random solutions*. Clearly, both the latter tasks are at least as hard as finding arbitrary solutions to the corresponding problem, but it turns out that for some natural problems they are not significantly harder. Specifically, under some natural conditions on the problem, approximately counting the number of solutions and generating an approximately random solution is not significantly harder than finding an arbitrary solution.

Having mentioned the notion of *approximation*, we note that the study of the complexity of finding approximate solutions has also received a lot of attention. One type of approximation problems refers to an objective function defined on the set of potential solutions. Rather than finding a solution that attains the optimal value, the approximation task consists of finding a solution that attains an "almost optimal" value, where the notion of "almost optimal" may be understood in different ways giving rise to different levels of approximation. Interestingly, in many cases even a very relaxed level of approximation is as difficult to achieve as the original (exact) search problem (i.e., finding an approximate solution is as hard as finding an optimal solution). Surprisingly, these hardness of approximation results are related to the study of *probabilistically checkable proofs*, which are proofs that allow for ultra-fast probabilistic verification. Amazingly, every proof can be efficiently transformed into one that allows for probabilistic verification based on probing a *constant* number of bits (in the alleged proof). Turning back to approximation problems, we note that in other cases a reasonable level of approximation is easier to achieve than solving the original (exact) search problem.

Approximation is a natural relaxation of various computational problems. Another natural relaxation is the study of *average-case complexity*, where the "average" is taken over some "simple" distributions (representing a model of the problem's instances that may occur in practice). We stress that, although it was not stated explicitly, the entire discussion so far has referred to "worst-case" analysis of algorithms. We mention that worst-case complexity is a more robust notion than average-case complexity. For starters, one avoids the controversial question of what are the instances that are "important in practice" and correspondingly the selection of the class of distributions for which average-case analysis is to be conducted. Nevertheless, a relatively robust theory of average-case complexity has been suggested, albeit it is far less developed than the theory of worst-case complexity.

In view of the central role of randomness in complexity theory (as evident, say, in the study of pseudorandomness, probabilistic proof systems, and cryptography), one may wonder as to whether the randomness needed for the various applications can be obtained in real-life. One specific question, which received a lot of attention, is the possibility of "purifying" randomness (or "extracting good randomness from bad sources"). That is, can we use "defected" sources of randomness in order to implement almost perfect sources of randomness. The answer depends, of course, on the model of such defected sources. This study turned out to be related to complexity theory, where the most tight connection is between some type of *randomness extractors* and some type of pseudorandom generators.

So far we have focused on the time complexity of computational tasks, while relying on the natural association of efficiency with time. However, time is not the only resource one should care about. Another important resource is *space*: the amount of (temporary) memory consumed by the computation. The study of space complexity has uncovered several fascinating phenomena, which seem to indicate a fundamental difference between space complexity and time complexity. For example, in the context of space complexity, verifying proofs of validity of assertions (of any specific type) has the same complexity as verifying proofs of invalidity for the same type of assertions.

In case the reader feels dizzy, it is no wonder. We took an ultra-fast air-tour of some mountain tops, and dizziness is to be expected. Needless to say, the rest of the book is in a totally different style. We will climb some of these mountains by foot, step by step, and will stop to look around and reflect.

Absolute Results (a.k.a. Lower-Bounds). As stated up-front, absolute results are not known for many of the "big questions" of complexity theory (most notably the P-versus-NP Question). However, several highly non-trivial absolute results have been proved. For example, it was shown that using negation can speed-up the computation of monotone functions (which do not require negation for their mere computation). In addition, many promising techniques were introduced and employed with the aim of providing a low-level analysis of the progress of computation. However, as stated in the preface, the focus of this book is elsewhere.

1.1.2 Characteristics of Complexity Theory

We are successful because we use the right level of abstraction

Avi Wigderson (1996)

Using the "right level of abstraction" seems to be a main characteristic of the Theory of Computation at large. The right level of abstraction means abstracting away second-order details, which tend to be context-dependent, while using definitions that reflect the main issues (rather than abstracting them away too). Indeed, using the right level of abstraction calls for an extensive exercising of good judgment, and one indication for having chosen the right questions is the result of their study.

One major choice of the theory of computation, which is currently taken for granted, is the *choice of a model of computation and corresponding complexity measures and classes.* Two extreme choices that were avoided are a too realistic model and a too abstract model. On the one hand, the main model of computation used in complexity theory does not try to reflect (or mirror) the specific operation of real-life computers used at a specific point in time. Such a choice would have made it very hard to develop complexity theory as we know it and to uncover the fundamental relations discussed in this book: The mass of details would have obscured the view. On the other hand, avoiding any reference to a concrete model (like in the case of recursive function theory) does not encourage the introduction and study of natural measures of complexity. Indeed, as we shall see in Section 1.2.3, the choice was (and is) to use a simple model of computation (which does not mirror real-life computers), while avoiding any effects that are specific to that model (by keeping a eye on a host of variants and alternative models). The freedom from the specifics of the basic model is obtained by considering complexity classes that are invariant under a change of model (as long as the alternative model is "reasonable").

Another major choice is to use *asymptotic analysis.* Specifically, we consider the complexity of an algorithm as a function of its input length, and study the asymptotic behavior of this function. It turns out that structure that is hidden by concrete quantities appears at the limit. Furthermore, depending on the case, we classify functions according to different criteria. For example, in case of time complexity we consider classes of functions that are closed under multiplication, whereas in case of space complexity we consider closure under addition. In each case, the choice is governed by the nature of the complexity measure being considered. Indeed, one could have developed a theory without using these conventions, but this would have resulted in a far more cumbersome theory. For example, rather than saying that finding a satisfying assignment for a given formula is polynomial-time reducible to deciding satisfiability other of formulae, one could have said stated the exact functional dependence of the complexity of the search problem on the complexity of the decision problem.

Both aforementioned choices are common to other branches of the theory of computation. What makes complexity theory unique is its commitment to the most basic question of the theory of computation: *what can be efficiently computed*? This commitment is reflected by the area's *primary focus* on the class of efficient procedures, regardless of their goal. This focus is responsible for the area's failures and successes: It has failed to obtain a definite answer regarding the power and limitations of efficient procedures, but it has obtained many connections between computational phenomena that reflect what efficient procedures can or cannot do.

1.1.3 Contents of this book

This book consists of ten chapters and seven appendices. The chapters constitute the core of this book and are written in a style adequate for a textbook, whereas the appendices provide additional perspective and are written in the style of a survey article.

Section 1.2 and Chapter 2 are a prerequisite to the rest of the book. Technically, notions and results that appear in these parts are extensively used in the rest of the book. More importantly, the former parts are the conceptual framework that shapes the field and provides a good perspective on the questions and answers provided. Section 1.2 and Chapter 2 provide the very basic material that must be understood by anybody having an interest in complexity theory. The rest of the book covers more advanced material. Although some advanced chapters refer to material in other advanced chapters, the relation between these chapters is not a fundamental one. Thus, one may choose to read and/or teach an arbitrary subset of the advanced chapters and do it in an arbitrary order, provided one is willing to follow the relevant references to parts of other chapters. Needless to say, we recommend following the order presented in this book.

The rest of this section provides a brief summary of the contents of the various parts. This summary is intended for the teacher and/or the expert, and the student is referred to the more reader-friendly summaries that appear in the book's prefix.

Section 1.2: Preliminaries. This section provides the relevant background on computability theory, which is the basis for the rest of this book (as well as for complexity theory at large). Most importantly, it contains a discussion of central notions such as search and decision problems, algorithms that solve such problems, and their complexity. In addition, this section presents non-uniform models of computation (e.g., Boolean circuits).

Chapter 2: P, NP and NP-completeness. This chapter presents the P-vs-NP Question both in terms of search problems and in terms of decision problems. The second main topic of this chapter is the theory of NP-completeness. The chapter also provides a treatment of the general notion of a (polynomial-time) reduction, with special emphasis on self-reducibility. Additional topics include the existence of problems in NP that are neither NP-complete nor in P, optimal search algorithms, the class coNP, and promise problems.

Chapter 3: Variations on P and NP. This chapter provides a treatment of non-uniform polynomial-time (P/poly) and of the Polynomial-time Hierarchy (PH). Each of the two classes is defined in two equivalent ways (e.g., P/poly is defined both in terms of circuits and in terms of "machines that take advice"). In addition, it is shown that if NP is contained in P/poly then PH collapses to its second level.

Chapter 4: More Resources, More Power? This chapter provides separation results asserting, for example, that there are functions computable in cubic-time but not in quadratic-time. These results depend on using bounding functions that can be computed without exceeding the amount of resources that they specify. In contrast, if this condition is not satisfied then increasing the resources may have no effect.

Chapter 5: Space Complexity. Among the results presented in this chapter are a log-space algorithm for testing connectivity of (undirected) graphs, a proof that $\mathcal{NL} = \text{co}\mathcal{NL}$, and complete problems for $\mathcal{NL}$ and $\mathcal{PSPACE}$ (under log-space and poly-time reductions, respectively).

Chapter 6: Randomness and Counting. This chapter focuses on various randomized complexity classes (i.e., $\mathcal{BPP}$, $\mathcal{RP}$, and $\mathcal{ZPP}$) and the counting class $\#\mathcal{P}$. The results presented in this chapter include $\mathcal{BPP} \subseteq \mathcal{P}/\text{poly} \cap \Sigma_2$, the $\#\mathcal{P}$-completeness of the permanent, the connection between approximate counting and uniform generation of solutions, and the randomized reductions of approximate counting to $\mathcal{NP}$ and of $\mathcal{NP}$ to solving problems with unique solutions.

Chapter 7: The Bright Side of Hardness. We consider two conjectures that are related to $\mathcal{P} \neq \mathcal{NP}$. The first conjecture is that there are problems in $\mathcal{E}$ that are not solvable by (non-uniform) families of small (say polynomial-size) circuits, whereas the second conjecture is equivalent to the notion of *one-way functions*. Most of this chapter is devoted to converting these conjectures into tools that can be used for non-trivial derandomizations of $\mathcal{BPP}$ and for a host of cryptographic applications.

Chapter 8: Pseudorandom Generators. The pivot of this chapter is the notion of *computational indistinguishability* and corresponding notions of pseudorandomness. The definition of general-purpose pseudorandom generators (running in polynomial-time and withstanding any polynomial-time distinguisher) is presented as a special case of a general paradigm. The chapter also contains a presentation of other instantiations of the latter paradigm, including generators aimed at derandomizing complexity classes such as $\mathcal{BPP}$, generators withstanding space-bounded distinguishers, and some special-purpose generators.

Chapter 9: Probabilistic Proof Systems. This chapter provides an introduction to three types of probabilistic proof systems: *interactive proofs*, *zero-knowledge proofs*, and *probabilistic checkable proofs*. These proof systems share a common (untraditional) feature – they carry a probability of error; yet, this probability is explicitly bounded and can be reduced by successive application of the proof system. The gain in allowing this untraditional relaxation is substantial, as they enable the construction of proof systems with properties that seem impossible to achieve via traditional proof systems.

Chapter 10: Relaxing the Requirement. This chapter provides a treatment of two types of approximation problems and a theory of average-case (or rather typical-case) complexity. The traditional type of approximation problems refers to search problems and consists of a relaxation of standard optimization problems. The second type is known as "property testing" and consists of a relaxation of standard decision problems. The theory of average-case complexity involves several non-trivial definitional choices (e.g., an adequate choice of the class of distributions).

Appendix A: Glossary of Complexity Classes. The glossary provides self-contained definitions of most complexity classes mentioned in the book.

Appendix B: On the Quest for Lower Bounds. The first part, devoted to Circuit Complexity, reviews lower bounds for the *size* of (restricted) circuits that solve natural computational problems. The second part, devoted to Proof Complexity, reviews lower bounds on the length of (restricted) propositional proofs of natural tautologies.

Appendix C: On the Foundations of Modern Cryptography. The first part of this appendix augments the partial treatment of one-way functions, pseudorandom generators and zero-knowledge proofs, which is included in Chapters 7–9. Using these basic tools, the second part provides a treatment of basic cryptographic applications such as Encryption, Signatures, and General Cryptographic Protocols.

Appendix D: Probabilistic Preliminaries and Advanced Topics in Randomization. The probabilistic preliminaries include conventions regarding random variables and overviews of three useful inequalities (i.e., Markov Inequality, Chebyshev's Inequality, and Chernoff Bound). The advanced topics include constructions families of *hashing* functions and variants of the Leftover Hashing

Lemma, and overviews of *samplers* and *extractors* (i.e., the problem of randomness extraction).

Appendix E: Explicit Constructions. This appendix focuses on various computational aspects of error correcting codes and expander graphs. On the topic of codes, the appendix contains a review of the Hadamard code, Reed-Solomon codes, Reed-Muller codes, and a construction of a binary code of constant rate and constant relative distance. Also included are a brief review of the notions of locally testable and locally decodable codes, and a list-decoding bound. On the topic of expander graphs, the appendix contains a review of standard definitions and properties as well as a presentation of the Margulis-Gabber-Galil and the Zig-Zag constructions.

Appendix F: Some Omitted Proofs. This appendix contains some proofs that are beneficial as alternatives to the original and/or standard presentations. Included are proofs that $\mathcal{PH}$ is reducible to $\#\mathcal{P}$ via randomized Karp-reductions, and that $\mathcal{IP}(f) \subseteq \mathcal{AM}(O(f)) \subseteq \mathcal{AM}(f)$, for any function f such that $f(n) \in \{2, ..., \text{poly}(n)\}$.

Appendix G: Some Computational Problems. This appendix contains a brief introduction to graph algorithms, Boolean formulae, and finite fields.

Bibliography. As stated in Section 1.1.4, we tried to keep the bibliographic list as short as possible (and still reached a couple of hundreds of entries). As a result many relevant references were omitted. We tried, however, not to omit references to key papers in an area. In general, our choice of references was biased in favour of textbooks and survey articles.

Absent from this book. As stated in the preface, the current book does not provide a uniform cover of the various areas of complexity theory. Notable omissions include the areas of *circuit complexity* (cf. [43, 222]) and *proof complexity* (cf. [24]), which are briefly reviewed in Appendix B. Additional topics that are commonly covered in complexity theory courses but omitted here include the study of *branching programs* and *decision trees* (cf. [223]), *parallel computation* [133], and *communication complexity* [141]. Finally, we mention a two areas that we consider related to complexity theory, although this view is not very common. These areas are *distributed computing* [15] and *computational learning theory* [135].

1.1.4 Approach and style of this book

Explanations and motivations are merely methods for introducing superficial redundancy. Since the role of redundancy is to ensure error-correction, one better leave the implementation of these mechanisms to experts, trusting them to use the best error-correcting codes known at the time.

Leonid A Levin (1984)

Although people say that Levin is an extremist, the foregoing quote perfectly represent a common opinion regarding the presentation of scientific work. According to this opinion, the most important aspect of a scientific work is the technical result that it achieves and the rest is redundancy introduced for the sake of "error correction" and/or comfort. It is further believed that, like in a work of art, the interpretation of the work should be left with the reader (or viewer or listener).

The author disagrees with the aforementioned opinions, and argues that there is a fundamental difference between art and science, and that this difference refers exactly to the meaning of a piece of work. Science is concerned with meaning (and not with form), and in its quest for truth and/or understanding it follows philosophy (and not art). The author holds the opinion that the most important aspects of a scientific work are the intuitive question that it addresses, the reason that it addresses this question, the way it phrases the question, the approach that underlies its answer, and the ideas that are embedded in the answer. Following this view, it is important to communicate these aspects of the work, and the current book is written accordingly.

These issues are even more acute when it comes to complexity theory, because this field is extremely rich in conceptual content. Unfortunately, this content is rarely communicated (explicitly) in books and/or extensive surveys of the area.[3] The annoying (and quite amazing) consequences are students that have only a vague understanding of the *meaning* and general relevance of the fundamental notions and results that they were taught. The author's view is that these consequences are easy to avoid by taking the time to explicitly discuss the *meaning* of definitions and results. A related issue is using the "right" definitions (i.e., those that reflect better the fundamental nature of the notion being defined) and teaching things in the (conceptually) "right" order.

1.1.4.1 Focus on conceptual issues

In accordance with the foregoing, this book starts from the intuitive questions addressed by complexity theory, explains the fundamental importance of these questions, the specific ways that they are phrased, the approaches that underly the answers, and the ideas that are embedded in these answers. Thus, a significant portion of the text is devoted to motivating discussions that refer to the concepts and ideas that underly the definitions and results.

The material is organized around conceptual themes, which reflect fundamental notions and/or general questions. Specific computational problems are rarely referred to, with exceptions that are used either for sake of clarity or because the specific problem happens to capture a general conceptual phenomenon. For exam-

[3]It is tempting to speculate on the reasons for this phenomenon. One speculation is that communicating the conceptual content of complexity theory involves making bold philosophical assertions that are technically straightforward, whereas this combination does not fit the character of most researchers in complexity theory.

ple, in this book, complete problems are always secondary to the class for which they are complete.[4]

We tried to avoid the presentation of material that, in our opinion, is neither the "last word" on the subject nor represents the "right" way of approaching the subject. Thus, we do not always present the "best" known result.

1.1.4.2 On a few specific choices

Our technical presentation often differs from the standard one. In many cases this is due to conceptual considerations. At times, this leads to some technical simplifications. In this section we only discuss general themes and/or choices that have a global impact on much of the presentation.

Avoiding non-deterministic machines. We try to avoid non-deterministic machines as much as possible. As argued in several places (e.g., Section 2.1.4), we believe that these fictitious "machines" have a negative effect both from a conceptual and technical point of view. The conceptual damage caused by using non-deterministic machines is that it is unclear why one should care about what such machines can do. Needless to say, the reason to care is clear when noting that these fictitious "machines" offer a (convenient or rather slothful) way of phrasing fundamental issues. The technical damage caused by using non-deterministic machines is that they tend to confuse the students. Furthermore, they do not offer the best way to handle more advanced issues (e.g., counting classes).

In contrast, we use search problems as the basis for much of the presentation. Specifically, the class $\mathcal{PC}$ (see Definition 2.3), which consists of search problems having efficiently checkable solutions, plays a central role in our presentation. Indeed, defining this class is slightly more complicated than the standard definition of $\mathcal{NP}$ (based on non-deterministic machines), but the technical benefits start accumulating as we proceed. Needless to say, the class $\mathcal{PC}$ is a fundamental class of computational problems and this fact serves as the main motivation to its presentation. (Indeed, the most conceptually appealing phrasing of the P-vs-NP Question consists of asking whether every search problem in $\mathcal{PC}$ can be solved efficiently.)

Avoiding model-dependent effects. Our focus is on the notion of efficient computation. A rigorous definition of this notion seems to require reference to some concrete model of computation; however, all questions and answers considered in this book are invariant under the choice of such a concrete model, provided of course that the model is "reasonable" (which, needless to say, is a matter of intuition). Indeed, the foregoing text reflects the tension between the need to make rigorous definitions and the desire to be independent of technical choices,

[4]We admit that a very natural computational problem can give rise to a class of problems that are computationally equivalent to it, and that in such a case the class may be less interesting than the original problem. Furthermore, in some cases, the historical evolution actually went from a specific computational problem to a class of problems that are computationally equivalent to it. However, in all cases presented in this book, a retrospective evaluation suggests that the class is actually more important than the original problem (see, e.g., $\mathcal{NP}$ and $\#\mathcal{P}$).

which are unavoidable when making such definitions. Furthermore, in contrast to common beliefs, the foregoing comments refer not only to time complexity but also to space complexity. However, in both cases, the claim of invariance may not hold for marginally small resources (e.g., linear-time or sub-logarithmic space).

In contrast to the foregoing paragraph, in some cases we choose to be specific. The most notorious case is the association of efficiency with polynomial-time (see §1.2.3.4). Indeed, all the questions and answers regarding efficient computation can be phrased without referring to polynomial-time (i.e., by stating explicit functional relations between the complexities of the problems involved), but such a generalized treatment will be painful to follow.

1.1.4.3 On the presentation of technical material

In general, the more complex the issues are, the more levels of expositions we employ, starting from the most high-level exposition, and when necessary providing more than one level of details. In particular, whenever a proof is not very simple, we try to present the key ideas first and implementation details later. We also try to clearly indicate the passage from a high-level presentation to implementation details (e.g., by using phrases such as "details follow"). In some cases, especially in the case of advanced results, only proof sketches are provided and the implication is that the reader should be able to fill-up the missing details.

Few results are stated without proof. In some of these cases the proof idea or a proof overview is provided, but the reader is *not* supposed to be able to fill-up the highly non-trivial details. (We clearly indicate that this is the case in the text.) One notable example is the proof of the PCP Theorem (Theorem 9.16).

1.1.4.4 Organizational principles

Each of the main chapters starts with a high-level summary and ends with chapter notes and exercises. The latter are not aimed at testing or inspiring creativity, but are rather designed to help and verify the basic understanding of the main text.

As stated in the preface, this book focuses on the high-level approach to complexity theory and the low-level approach (i.e., lower bounds) is briefly reviewed in Appendix B. Other appendices contain material that is closely related to complexity theory but is not an integral part of it (e.g., the Foundations of Cryptography).[5]

In an attempt to keep the bibliographic list from becoming longer than an average chapter, we omitted many relevant references. One trick used towards this end is referring to lists of references in other texts, especially when these texts are cited anyhow. Indeed, our choices of references were biased in favour of textbooks and survey articles, because we believe that they provide the best way to further learn about a research direction and/or approach. We tried, however, not to omit references to key papers in an area. In some cases, when we needed a reference for

[5]As further articulated in Section 7.1, we recommend not including a basic treatment of cryptography within a course on complexity theory. Indeed, cryptography may be claimed to be the most appealing application of complexity theory, but a superficial treatment of cryptography (from this perspective) is likely to be misleading and cause more harm than good.

a result of interest and could not resort to the aforementioned trick, we cited also less central work.

As a matter of policy, we tried to avoid credits in the main text. The few exceptions are either pointers to texts that provide details that we chose to omit or usage of terms (bearing researchers' names) that are too popular to avoid.

Teaching note: The text also includes some teaching notes, which are typeset as this one. Some of these notes express quite opinionated recommendations and/or justify various expositional choices made in the text.

1.1.4.5 Additional notes

The author's guess is that the text will be criticized for lengthy discussions of technically trivial issues. Indeed, most researchers dismiss various conceptual clarifications as being trivial and devote all their attention to the technically challenging parts of the material. The consequence is students that master the technical material but are confused about its meaning. In contrast, the author recommends not being embarrassed of devoting time to conceptual clarifications, even if some students may view them as obvious.

The motivational discussions presented in the text do not necessarily represent the original motivation of the researchers that pioneered a specific study and/or contributed greatly to it. Instead, these discussions provide what the author considers to be a good motivation and/or perspective on the corresponding concepts.

1.1.5 Standard notations and other conventions

Following are some notations and conventions that are freely used in this book.

Standard asymptotic notation: When referring to integral functions, we use the standard asymptotic notation; that is, for $f, g : \mathbb{N} \to \mathbb{N}$, we write $f = O(g)$ (resp., $f = \Omega(g)$) if there exists a constant $c > 0$ such that $f(n) \leq c \cdot g(n)$ (resp., $f(n) \geq c \cdot g(n)$) holds for all $n \in \mathbb{N}$. We usually denote by poly an unspecified polynomial, and write $f(n) = \text{poly}(n)$ instead of "there exists a polynomial p such that $f(n) \leq p(n)$ for all $n \in \mathbb{N}$." We also use the notation $f = \widetilde{O}(g)$ that mean $f(n) = \text{poly}(\log n) \cdot g(n)$, and $f = o(g)$ (resp., $f = \omega(g)$) that mean $f(n) < c \cdot g(n)$ (resp., $f(n) > c \cdot g(n)$) for every constant $c > 0$ and all sufficiently large n.

Integrality issues: Typically, we ignore integrality issues. This means that we may assume that $\log_2 n$ is an integer rather than using a more cumbersome form as $\lfloor \log_2 n \rfloor$. Likewise, we may assume that various equalities are satisfied by integers (e.g., $2^n = m^m$), even when this cannot possibly be the case (e.g., $2^n = 3^m$). In all these cases, one should consider integers that approximately satisfy the relevant equations (and deal with the problems that emerge by such approximations, which will be ignored in the current text).

Standard combinatorial and graph theory terms and notation: For any set S, we denote by 2^S the set of all subsets of S (i.e., $2^S = \{S' : S' \subseteq S\}$). For a natural number $n \in \mathbb{N}$, we denote $[n] \stackrel{\rm def}{=} \{1, ..., n\}$. Many of the computational problems refer to finite (undirected) graphs. Such a graph, denoted $G = (V, E)$, consists of a set of vertices, denoted V, and a set of edges, denoted E, which are unordered pairs of vertices. By default, graphs are undirected, whereas directed graphs consists of vertices and directed edges, where a directed edge is an order pair of vertices. We also refer to other graph theoretic terms such as connectivity, being acyclic (i.e., having no simple cycles), being a tree (i.e., being connected and acyclic), k-colorability, etc. For further background on graphs and computational problems regarding graphs, the reader is referred to Appendix G.1.

Typographic conventions: We denote formally defined complexity classes by caligraphic letters (e.g., $\mathcal{NP}$), but we do so only after defining these classes. Furthermore, when we wish to maintain some ambiguity regarding the specific formulation of a class of problems we use Roman font (e.g., NP may denote either a class of search problems or a class of decision problems). Likewise, we denote formally defined computational problems by typewriter font (e.g., SAT). In contrast, generic problems and algorithms will be denoted by italic font.

1.2 Computational Tasks and Models

We start by introducing the general framework for our discussion of computational tasks (or problems) This framework refers to the representation of instances and to two types of tasks (i.e., searching for solutions and making decisions). Once the stage is set, we consider two types of models of computation: uniform models that correspond to the intuitive notion of an algorithm, and non-uniform models (e.g., Boolean circuits) that facilitates a closer look at the way computation progresses.

Contents of Section 1.2. The contents of Sections 1.2.1–1.2.3 corresponds to a traditional *Computability course*, except that it includes a keen interest in universal machines (see §1.2.3.3), a discussion of the association of efficient computation with polynomial-time algorithm (§1.2.3.4), and a definition of oracle machines (§1.2.3.5). This material (with the exception of Kolmogorov Complexity) is taken for granted in the rest of the current book. (We also call the reader's attention to the discussion of generic complexity classes in Section 1.2.5.) In contrast, Section 1.2.4 presents basic preliminaries regarding non-uniform models of computation (i.e., various types of Boolean circuits), and these are only used lightly in the rest of the book. Thus, whereas Sections 1.2.1–1.2.3 (and 1.2.5) are absolute prerequisites for the rest of this book, Section 1.2.4 is not.

Teaching note: The author believes that there is no real need for a semester-long course in Computability (i.e., a course that focuses on what can be computed rather than on what can be computed efficiently). Instead, undergraduates should take a course in Computational Complexity, which should contain the computability aspects that serve as a basis for the rest of the course. Specifically, the former aspects should occupy at most 25% of the course, and the focus should be on basic complexity issues (captured by P, NP, and NP-completeness) augmented by a selection of some more advanced material. Indeed, such a course can be based on Chapters 1 and 2 of the current book (augmented by a selection of some topics from other chapters).

1.2.1 Representation

In Mathematics and related sciences, it is customary to discuss objects without specifying their representation. This is not possible in the theory of computation, where the representation of objects plays a central role. In a sense, a computation merely transforms one representation of an object to another representation of the same object. In particular, a computation designed to solve some problem merely transforms the problem instance to its solution, where the latter can be though of as a (possibly partial) representation of the instance. Indeed, the answer to any fully specified question is implicit in the question itself.

Computation refers to objects that are represented in some canonical way, where such canonical representation provides an "explicit" and "full" (but not "overly redundant") description of the corresponding object. We will consider only *finite* objects like sets, graphs, numbers, and functions (and keep distinguishing these types of objects although, actually, they are all equivalent). (For example, see Appendix G.1 for a discussion of the representation of graphs.)

Strings. We consider finite objects, each represented by a finite binary sequence, called a string. For a natural number n, we denote by $\{0,1\}^n$ the set of all strings of length n, hereafter referred to as n-bit strings. The set of all strings is denoted $\{0,1\}^*$; that is, $\{0,1\}^* = \cup_{n\in\mathbb{N}}\{0,1\}^n$. For $x\in\{0,1\}^*$, we denote by $|x|$ the length of x (i.e., $x\in\{0,1\}^{|x|}$), and often denote by x_i the i^{th} bit of x (i.e., $x = x_1x_2\cdots x_{|x|}$). For $x,y\in\{0,1\}^*$, we denote by xy the string resulting from concatenation of the strings x and y.

At times, we associate $\{0,1\}^*\times\{0,1\}^*$ with $\{0,1\}^*$; the reader should merely consider an adequate encoding (e.g., the pair $(x_1\cdots x_m, y_1\cdots y_n)\in\{0,1\}^*\times\{0,1\}^*$ may be encoded by the string $x_1x_1\cdots x_mx_m01y_1\cdots y_n \in \{0,1\}^*$). Likewise, we may represent sequences of strings (of fixed or varying length) as single strings. When we wish to emphasize that such a sequence (or some other object) is to be considered as a single object we use the notation $\langle\cdot\rangle$ (e.g., "the pair (x,y) is encoded as the string $\langle x,y\rangle$").

Numbers. Unless stated differently, natural numbers will be encoded by their binary expansion; that is, the string $b_{n-1}\cdots b_1b_0 \in \{0,1\}^n$ encodes the number $\sum_{i=0}^{n-1} b_i\cdot 2^i$, where typically we assume that this representation has no leading zeros (i.e., $b_{n-1} = 1$). Rational numbers will be represented as pairs of natural numbers. In the rare cases in which one considers real numbers as part of the input to a computational problem, one actually mean rational approximations of these real numbers.

Special symbols. We denote the empty string by λ (i.e., $\lambda \in \{0,1\}^*$ and $|\lambda| = 0$), and the empty set by $\emptyset$. It will be convenient to use some special symbols that are not in $\{0,1\}^*$. One such symbol is $\perp$, which typically denotes an indication by some algorithm that something is wrong.

1.2.2 Computational Tasks

Two fundamental types of computational tasks are so-called search problems and decision problems. In both cases, the key notions are the problem's *instances* and the problem's specification.

1.2.2.1 Search problems

A search problem consists of a specification of a set of valid solutions (possibly an empty one) for each possible instance. That is, given an instance, one is required to find a corresponding solution (or to determine that no such solution exists). For example, consider the problem in which one is given a system of equations and is asked to find a valid solution. Needless to say, much of computer science is concerned with solving various search problems (e.g., finding shortest paths in a graph, sorting a list of numbers, finding an occurrence of a given pattern in a given string, etc). Furthermore, search problems correspond to the daily notion of "solving a problem" (e.g., finding one's way between two locations), and thus a discussion of the possibility and complexity of solving search problems corresponds to the natural concerns of most people.

In the following definition of solving search problems, the potential solver is a function (which may be thought of as a solving strategy), and the sets of possible solutions associated with each of the various instances are "packed" into a single binary relation.

Definition 1.1 (solving a search problem): *Let $R \subseteq \{0,1\}^* \times \{0,1\}^*$ and $R(x) \stackrel{\text{def}}{=} \{y : (x,y) \in R\}$ denote the set of solutions for the instance x. A function $f : \{0,1\}^* \to \{0,1\}^* \cup \{\perp\}$* solves the search problem of *R if for every x the following holds: if $R(x) \neq \emptyset$ then $f(x) \in R(x)$ and otherwise $f(x) = \perp$.*

Indeed, $R = \{(x,y) : y \in R(x)\}$, and the solver f is required to find a solution (i.e., given x output $y \in R(x)$) whenever one exists (i.e., the set $R(x)$ is not empty). It is also required that the solver f never outputs a wrong solution (i.e., if $R(x) \neq \emptyset$ then $f(x) \in R(x)$ and if $R(x) = \emptyset$ then $f(x) = \perp$), which in turn means that f indicates whether x has any solution.

A special case of interest is the case of search problems having a unique solution (for each possible instance); that is, the case that $|R(x)| = 1$ for every x. In this case, R is essentially a (total) function, and solving the search problem of R means computing (or evaluating) the function R (or rather the function R' defined by $R'(x) \stackrel{\text{def}}{=} y$ where $R(x) = \{y\}$). Popular examples include sorting a sequence of numbers, multiplying integers, finding the prime factorization of a composite number, etc.

1.2.2.2 Decision problems

A decision problem consists of a specification of a subset of the possible instances. Given an instance, one is required to determine whether the instance is in the specified set (e.g., the set of prime numbers, the set of connected graphs, or the set of sorted sequences). For example, consider the problem where one is given a natural number, and is asked to determine whether or not the number is a prime. One important case, which corresponds to the aforementioned search problems, is the case of the set of instances having a solution; that is, for any binary relation $R \subseteq \{0,1\}^* \times \{0,1\}^*$ we consider the set $\{x : R(x) \neq \emptyset\}$. Indeed, being able to determine whether or not a solution exists is a prerequisite to being able to solve the corresponding search problem (as per Definition 1.1). In general, decision problems refer to the natural task of making binary decision, a task that is not uncommon in daily life (e.g., determining whether a traffic light is red). In any case, in the following definition of solving decision problems, the potential solver is again a function (i.e., in this case it is a Boolean function that is supposed to indicate membership in the said set).

Definition 1.2 (solving a decision problem): *Let $S \subseteq \{0,1\}^*$. A function $f : \{0,1\}^* \to \{0,1\}$* solves the decision problem of S (or decides membership in S) *if for every x it holds that $f(x) = 1$ if and only if $x \in S$.*

We often identify the decision problem of S with S itself, and identify S with its characteristic function (i.e., with $\chi_S : \{0,1\}^* \to \{0,1\}$ defined such that $\chi_S(x) = 1$ if and only if $x \in S$). Note that if f solves the search problem of R then the Boolean function $f' : \{0,1\}^* \to \{0,1\}$ defined by $f'(x) \stackrel{\text{def}}{=} 1$ if and only if $f(x) \neq \perp$ solves the decision problem of $\{x : R(x) \neq \emptyset\}$.

Most people would consider search problems to be more natural than decision problems: typically, people seeks solutions more than they stop to wonder whether or not solutions exist. Definitely, search problems are not less important than decision problems, it is merely that their study tends to require more cumbersome formulations. This is the main reason that most expositions choose to focus on decision problems. The current book attempts to devote at least a significant amount of attention also to search problems.

1.2.2.3 Promise problems (an advanced comment)

Many natural search and decision problems are captured more naturally by the terminology of promise problems, where the domain of possible instances is a subset of $\{0,1\}^*$ rather than $\{0,1\}^*$ itself. In particular, note that the natural formulation of many search and decision problems refers to instances of a certain types (e.g., a system of equations, a pair of numbers, a graph), whereas the natural representation of these objects uses only a strict subset of $\{0,1\}^*$. For the time being, we ignore this issue, but we shall re-visit it in Section 2.4.1. Here we just note that, in typical cases, the issue can be ignored by postulating that every string represents some legitimate object (i.e., each string that is not used in the natural representation of these objects is postulated as a representation of some fixed object).

1.2.3 Uniform Models (Algorithms)

We are all familiar with computers and with the ability of computer programs to manipulate data. This familiarity seems to be rooted in the positive side of computing; that is, we have some experience regarding some things that computers can do. In contrast, complexity theory is focused at what computers cannot do, or rather with drawing the line between what can be done and what cannot be done. Drawing such a line requires a precise formulation of *all* possible computational processes; that is, we should have a clear model of *all* possible computational processes (rather than some familiarity with some computational processes).

Before being formal, let we offer a general and abstract description, which is aimed at capturing any artificial as well as natural process. Indeed, artificial processes will be associated with computers, whereas by natural processes we mean (attempts to model) the "mechanical" aspects the natural reality (be it physical, biological, or even social).

A computation is a process that modifies an environment via repeated applications of a predetermined rule. The key restriction is that this rule is *simple*: in each application it depends and affects only a (small) portion of the environment, called the active zone. We contrast the *a-priori bounded* size of the active zone (and of the modification rule) with the *a-priori unbounded* size of the entire environment. We note that, although each application of the rule has a very limited effect, the effect of many applications of the rule may be very complex. Put in other words, a computation may modify the relevant environment in a very complex way, although it is merely a process of repeatedly applying a simple rule.

As hinted, the notion of computation can be used to model the "mechanical" aspects of the natural reality; that is, the rules that determine the evolution of the reality (rather than the specifics of reality itself). In this case, the evolution process that takes place in the natural reality is the starting point of the study, and the goal of the study is finding the (computation) rule that underlies this natural process. In a sense, the goal of Science at large can be phrased as finding (simple) rules that govern various aspects of reality (or rather one's abstraction of these aspects of reality).

Our focus, however, is on artificial computation rules designed by humans in order to achieve specific desired effects on a corresponding artificial environment. Thus, our starting point is a desired functionality, and our aim is to design computation rules that effect it. Such a computation rule is referred to as an algorithm. Loosely speaking, an algorithm corresponds to a computer program written in a high-level (abstract) programming language. Let us elaborate.

We are interested in the transformation of the environment affected by the computational process (or the algorithm). Throughout (most of) this book, we will assume that, *when invoked on any finite initial environment, the computation halts after a finite number of steps*. Typically, the initial environment to which the computation is applied encodes an input string, and the end environment (i.e., at termination of the computation) encodes an output string. We consider the mapping from inputs to outputs induced by the computation; that is, for each possible input x, we consider the output y obtained at the end of a computation initiated with input x, and say that the computation maps input x to output y. Thus, a computation rule (or an algorithm) determines a function (computed by it): this function is exactly the aforementioned mapping of inputs to outputs.

In the rest of this book (i.e., outside the current chapter), we will also consider the number of steps (i.e., applications of the rule) taken by the computation on each possible input. The latter function is called the time complexity of the computational process (or algorithm). While time complexity is defined per input, we will often considers it per input length, taking the maximum over all inputs of the same length.

In order to define computation (and computation time) rigorously, one needs to specify some model of computation; that is, provide a concrete definition of environments and a class of rules that may be applied to them. Such a model corresponds to an abstraction of a real computer (be it a PC, mainframe or network of computers). One simple abstract model that is commonly used is that of *Turing machines* (see, §1.2.3.1). Thus, specific algorithms are typically formalized by corresponding Turing machines (and their time complexity is represented by the time complexity of the corresponding Turing machines). We stress, however, that most results in the Theory of Computation hold regardless of the specific computational model used, as long as it is "reasonable" (i.e., satisfies the aforementioned simplicity condition and can perform some obviously simple computations).

What is being computed? The forgoing discussion has implicitly referred to algorithms (i.e., computational processes) as means of computing functions. Specifically, an algorithm A computes the function $f_A : \{0,1\}^* \to \{0,1\}^*$ defined by $f_A(x)=y$ if, when invoked on input x, algorithm A halts with output y. However, algorithms can also serve as means of "solving search problems" or "making decisions" (as in Definitions 1.1 and 1.2). Specifically, we will say that algorithm A solves the search problem of R (resp., decides membership in S) if f_A solves the search problem of R (resp., decides membership in S). In the rest of this exposition we associate the algorithm A with the function f_A computed by it; that is, we write $A(x)$ instead of $f_A(x)$. For sake of future reference, we summarize the foregoing discussion.

Definition 1.3 (algorithms as problem-solvers): *We denote by $A(x)$ the output of algorithm A on input x. Algorithm A* solves the search problem *R* (resp., the decision problem S) *if A, viewed as a function, solves R* (resp., S).

Organization of the rest of Section 1.2.3. In §1.2.3.1 we provide a sketchy description of the model of Turing machines. This is done merely for sake of providing a concrete model that supports the study of computation and its complexity, whereas most of the material in this book will not depend on the specifics of this model. In §1.2.3.2 and §1.2.3.2 we discuss two fundamental properties of any reasonable model of computation: the existence of uncomputable functions and the existence of universal computations. The time (and space) complexity of computation is defined in §1.2.3.4. We also discuss oracle machines and restricted models of computation (in §1.2.3.5 and §1.2.3.6, respectively).

1.2.3.1 Turing machines

The model of Turing machines offer a relatively simple formulation of the notion of an algorithm. The fact that the model is very simple complicates the design of machines that solve problems of interest, but makes the analysis of such machines simpler. Since the focus of complexity theory is on the analysis of machines and not on their design, the trade-off offers by this model is suitable for our purposes. We stress again that the model is merely used as a concrete formulation of the intuitive notion of an algorithm, whereas we actually care about the intuitive notion and not about its formulation. In particular, all results mentioned in this book hold for any other "reasonable" formulation of the notion of an algorithm.

The model of Turing machines is not meant to provide an accurate (or "tight") model of real-life computers, but rather to capture their inherent limitations and abilities (i.e., a computational task can be solved by a real-life computer if and only if it can be solved by a Turing machine). In comparison to real-life computers, the model of Turing machines is extremely over-simplified and abstract away many issues that are of great concern to computer practice. However, these issues are irrelevant to the higher-level questions addressed by complexity theory. Indeed, as usual, good practice requires more refined understanding than the one provided by a good theory, but one should first provide the latter.

Historically, the model of Turing machines was invented before modern computers were even built, and was meant to provide a concrete model of computation and a definition of computable functions.[6] Indeed, this concrete model clarified fundamental properties of computable functions and plays a key role in defining the complexity of computable functions.

The model of Turing machines was envisioned as an abstraction of the process of an algebraic computation carried out by a human using a sheet of paper. In such a process, at each time, the human looks at some location on the paper, and depending on what he/she sees and what he/she has in mind (which is little...), he/she modifies the contents of this location and shifts his/her look to an adjacent location.

The actual model. Following is a high-level description of the model of Turing machines; the interested reader is referred to standard textbooks (e.g., [197]) for

[6] In contrast, the abstract definition of "recursive functions" yields a class of "computable" functions defined recursively in terms of the composition of such functions.

further details. Recall that we need to specify the set of possible environments, the set of machines (or computation rules), and the effect of applying such a rule on an environment.

- The main component in the environment of a Turing machine is an infinite sequence of cells, each capable of holding a single symbol (i.e., member of a finite set $\Sigma \supset \{0,1\}$). In addition, the environment contains the current location of the machine on this sequence, and the internal state of the machine (which is a member of a finite set Q). The aforementioned sequence of cells is called the tape, and its contents combined with the machine's location and its internal state is called the instantaneous configuration of the machine.

- The Turing machine itself consists of a finite rule (i.e., a finite function), called the transition function, which is defined over the set of all possible symbol-state pairs. Specifically, the transition function is a mapping from $\Sigma \times Q$ to $\Sigma \times Q \times \{-1,0,+1\}$, where $\{-1,+1,0\}$ correspond to a movement instruction (which is either "left" or "right" or "stay", respectively). In addition, the machine's description specifies an initial state and a halting state, and the computation of the machine halts when the machine enters its halting state.[7]

  We stress that, in contrast to the finite description of the machine, the tape has an a priori unbounded length (and is considered, for simplicity, as being infinite).

- A single computation step of such a Turing machine depends on its current location on the tape, on the contents of the corresponding cell and on the internal state of the machine. Based on the latter two elements, the transition function determines a new symbol-state pair as well as a movement instruction (i.e., "left" or "right" or "stay"). The machine modifies the contents of the said cell and its internal state accordingly, and moves as directed. That is, suppose that the machine is in state q and resides in a cell containing the symbol σ, and suppose that the transition function maps (σ, q) to (σ', q', D). Then, the machine modifies the contents of the said cell to σ', modifies its internal state to q', and moves one cell in direction D. Figure 1.1 shows a single step of a Turing machine that, when in state 'b' and seeing a binary symbol σ, replaces σ with the symbol $\sigma + 2$, maintains its internal state, and moves one position to the right.[8]

  Formally, we define the successive configuration function that maps each instantaneous configuration to the one resulting by letting the machine take a single step. This function modifies its argument in a very minor manner, as described in the foregoing; that is, the contents of at most one cell (i.e., at

[7]Envisioning the tape as extending from left to right, we also use the convention by which if the machine tries to move left of the end of the tape then it is considered to have halted.

[8]Figure 1.1 corresponds to a machine that, when in the initial state (i.e., 'a'), replaces the symbol σ by $\sigma + 4$, modifies its internal state to 'b', and moves one position to the right. Indeed, "marking" the leftmost cell (in order to allow for recognizing it in the future), is a common practice in the design of Turing machines.

which the machine currently resides) is changed, and in addition the internal state of the machine and its location may change too.

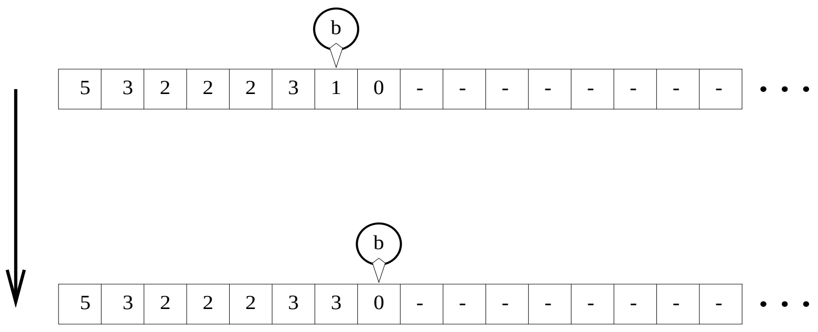


Figure 1.1: A single step by a Turing machine.

The initial environment (or configuration) of a Turing machine consists of the machine residing in the first (i.e., left-most) cell and being in its initial state. Typically, one also mandates that, in the initial configuration, a prefix of the tape's cells hold bit values, which concatenated together are considered the input, and the rest of the tape's cells hold a special symbol (which in Figure 1.1 is denoted by '-'). Once the machine halts, the output is defined as the contents of the cells that are to the left of its location (at termination time).[9] Thus, each machine defines a function mapping inputs to outputs, called the function computed by the machine.

Multi-tape Turing machines. We comment that in most expositions, one refers to the location of the "head of the machine" on the tape (rather than to the "location of the machine on the tape"). The standard terminology is more intuitive when extending the basic model, which refers to a single tape, to a model that supports a constant number of tapes. In the model of multi-tape machines, each step of the machine depends and effects the cells that are at the head location of the machine on each tape. As we shall see in Chapter 5 (and in §1.2.3.4), the extension of the model to multi-tape Turing machines is crucial to the definition of space complexity. A less fundamental advantage of the model of multi-tape Turing machines is that it facilitates the design of machines that compute functions of interest.

Teaching note: We strongly recommend avoiding the standard practice of teaching the student to program with Turing machines. These exercises seem very painful and pointless. Instead, one should prove that a function can be computed by a Turing machine if and only if it is computable by a model that is closer to a real-life computer (see the following "sanity check"). For starters, one should prove that a function can be computed by a single-tape Turing machine if and only if it is computable by a multi-tape (e.g., two-tape) Turing machine.

[9]By an alternative convention, the machine halts while residing in the left-most cell, and the output is defined as the maximal prefix of the tape contents that contains only bit values.

The Church-Turing Thesis: The entire point of the model of Turing machines is its simplicity. That is, in comparison to more "realistic" models of computation, it is simpler to formulate the model of Turing machines and to analyze machines in this model. The Church-Turing Thesis asserts that nothing is lost by considering the Turing machine model: *A function can be computed by some Turing machine if and only if it can be computed by some machine of any other* "reasonable and general" *model of computation.*

This is a thesis, rather than a theorem, because it refers to an intuitive notion that is left undefined on purpose (i.e., the notion of a *reasonable and general model of computation*). The model should be reasonable in the sense that it should refer to computation rules that are "simple" in some intuitive sense. On the other hand, the model should allow to compute functions that intuitively seem computable. At the very least the model should allow to emulate Turing machines (i.e., compute the function that given a description of a Turing machine and an instantaneous configuration returns the successive configuration).

A philosophical comment. The fact that a thesis is used to link an intuitive concept to a formal definition is common practice in any science (or, more broadly, in any attempt to reason rigorously about intuitive concepts). The moment an intuition is rigorously defined, it stops being an intuition and becomes a definition, and the question of the correspondence between the original intuition and the derived definition arises. This question can never be rigorously treated, because it relates to two objects, where one of them is undefined. Thus, the question of correspondence between the intuition and the definition always transcends a rigorous treatment (i.e., it always belongs to the domain of the intuition).

A sanity check: Turing machines can emulate an abstract RAM. To gain confidence in the Church-Turing Thesis, one may attempt to define an abstract Random-Access Machine (RAM), and verify that it can be emulated by a Turing machine. An abstract RAM consists of an infinite number of memory cells, each capable of holding an integer, a finite number of similar registers, one designated as program counter, and a program consisting of instructions selected from a finite set. The set of possible instructions includes the following instructions:

- `reset`(r), where r is an index of a register, results in setting the value of register r to zero.
- `inc`(r), where r is an index of a register, results in incrementing the content of register r. Similarly `dec`(r) causes a decrement.
- `load`(r_1, r_2), where r_1 and r_2 are indices of registers, results in loading to register r_1 the contents of the memory location m, where m is the current contents of register r_2.
- `store`(r_1, r_2), stores the contents of register r_1 in the memory, analogously to `load`.
- `cond-goto`(r, ℓ), where r is an index of a register and ℓ does not exceed the program length, results in setting the program counter to $\ell - 1$ if the content of register r is non-negative.

The program counter is incremented after the execution of each instruction, and the next instruction to be executed by the machine is the one to which the program counter points (and the machine halts if the program counter exceeds the program's length). The input to the machine may be defined as the contents of the first n memory cells, where n is placed in a special input register. We note that the RAM model satisfies the Church-Turing Thesis, but in order to make it closer to real-life computers we may augment the model with additional instructions that are available on such computers (e.g., the instruction $\texttt{add}(r_1,r_2)$ (resp., $\texttt{mult}(r_1,r_2)$) that results in adding (resp., multiplying) the contents of registers r_1 and r_2 and placing the result in register r_1). We suggest proving that this abstract RAM can be emulated by a Turing machine.[10] (Hint: note that during the emulation, we only need to hold the input, the contents of all registers, and the contents of the memory cells that were accessed during the computation.)[11]

Observe that the abstract RAM model is significantly more cumbersome than the Turing machine model. Furthermore, seeking a sound choice of the instruction set (i.e., the instructions to be allowed in the model) creates a vicious cycle (because the sound guideline would have been to allow only instructions that correspond to "simple" operations, whereas the latter correspond to easily computable functions...). This vicious cycle was avoided by trusting the reader to consider only instructions that are available in some real-life computer. (We comment that this empirical consideration is justifiable in the current context, because our current goal is merely linking the Turing machine model with the reader's experience of real-life computers.)

1.2.3.2 Uncomputable functions

Strictly speaking, the current subsection is not necessary for the rest of this book, but we feel that it provides a useful perspective.

In contrast to what every layman would think, we know that not all functions are computable. Indeed, an important message to be communicated to the world is that *not every well-defined task can be solved* by applying a "reasonable" procedure (i.e., a procedure that has a simple description that can be applied to any instance of the problem at hand). Furthermore, not only is it the case that there exist uncomputable functions, but it is rather the case that most functions are uncomputable. In fact, only relatively few functions are computable.

Theorem 1.4 (on the scarcity of computable functions): *The set of computable functions is countable, whereas the set of all functions* (from strings to string) *has*

[10]We emphasize this direction of the equivalence of the two models, because the RAM model is introduced in order to convince the reader that Turing machines are not too weak (as a model of general computation). The fact that they are not too strong seems self-evident. Thus, it seems pointless to prove that the RAM model can emulate Turing machines. Still, note that this is indeed the case, by using the RAM's memory cells to store the contents of the cells of the Turing machine's tape.

[11]Thus, at each time, the Turning machine's tape contains a list of the RAM's memory cells that were accessed so far as well as their current contents. When we emulate a RAM instruction, we first check whether the relevant RAM cell appears on this list, and augment the list by a corresponding entry or modify this entry as needed.

cardinality $\aleph$.

We stress that the theorem holds for any reasonable model of computation. In fact, it only relies on the postulate that each machine in the model has a finite description (i.e., can be described by a string).

Proof: Since each computable function is computable by a machine that has a finite description, there is a 1-1 correspondence between the set of computable functions and the set of strings (which in turn is in 1-1 correspondence to the natural numbers). On the other hand, there is a 1-1 correspondence between the set of Boolean functions (i.e., functions from strings to a bit) and the set of real number in $[0,1)$. This correspondence associates each real $r \in [0,1)$ to the function $f : \mathbb{N} \to \{0,1\}$ such that $f(i)$ is the i^{th} bit in the binary expansion of r. ■

The Halting Problem: In contrast to the preliminary discussion, at this point we consider also machines that may not halt on some inputs. (The functions computed by such machines are partial functions that are defined only on inputs on which the machine halts.) Again, we rely on the postulate that each machine in the model has a finite description, and denote the description of machine M by $\langle M\rangle \in \{0,1\}^*$. The halting function, $\mathtt{h} : \{0,1\}^* \times \{0,1\}^* \to \{0,1\}$, is defined such that $\mathtt{h}(\langle M\rangle, x) \stackrel{\text{def}}{=} 1$ if and only if M halts on input x. The following result goes beyond Theorem 1.4 by pointing to an explicit function (of natural interest) that is not computable.

Theorem 1.5 (undecidability of the halting problem): *The halting function is not computable.*

The term undecidability means that the corresponding decision problem cannot be solved by an algorithm. That is, Theorem 1.5 asserts that the decision problem associated with the set $\mathtt{h}^{-1}(1) = \{(\langle M\rangle, x) : \mathtt{h}(\langle M\rangle, x) = 1\}$ is not solvable by an algorithm (i.e., there exists no algorithm that, given a pair $(\langle M\rangle, x)$, decides whether or not M halts on input x). Actually, the following proof shows that there exists no algorithm that, given $\langle M\rangle$, decides whether or not M halts on input $\langle M\rangle$.

Proof: We will show that even the restriction of $\mathtt{h}$ to its "diagonal" (i.e., the function $\mathtt{d}(\langle M\rangle) \stackrel{\text{def}}{=} \mathtt{h}(\langle M\rangle, \langle M\rangle)$) is not computable. Note that the value of $\mathtt{d}(\langle M\rangle)$ refers to the question of what happens when we feed M with its own description, which is indeed a "nasty" (but legitimate) thing to do. We will actually do worse: towards the contradiction, we will consider the value of $\mathtt{d}$ when evaluated at a (machine that is related to a) machine that supposedly computes $\mathtt{d}$.

We start by considering a related function, $\mathtt{d}'$, and showing that this function is uncomputable. This function is defined on purpose so to foil any attempt to compute it; that is, for every machine M, the value $\mathtt{d}'(\langle M\rangle)$ is defined to differ from $M(\langle M\rangle)$. Specifically, the function $\mathtt{d}' : \{0,1\}^* \to \{0,1\}$ is defined such that $\mathtt{d}'(\langle M\rangle) \stackrel{\text{def}}{=} 1$ *if and only if* M *halts on input* $\langle M\rangle$ *with output* 0. (That is, $\mathtt{d}'(\langle M\rangle) = 0$ if either M does not halt on input $\langle M\rangle$ or its output does not equal the value 0.) Now, suppose, towards the contradiction, that $\mathtt{d}'$ is computable by some machine, denoted $M_{\mathtt{d}'}$. Note that machine $M_{\mathtt{d}'}$ is supposed to halt on every input, and so $M_{\mathtt{d}'}$ halts on input $\langle M_{\mathtt{d}'}\rangle$. But, by definition of $\mathtt{d}'$, it holds that $\mathtt{d}'(\langle M_{\mathtt{d}'}\rangle) = 1$ if and only if $M_{\mathtt{d}'}$ halts on input $\langle M_{\mathtt{d}'}\rangle$ with output 0 (i.e., if and only if $M_{\mathtt{d}'}(\langle M_{\mathtt{d}'}\rangle) = 0$). Thus, $M_{\mathtt{d}'}(\langle M_{\mathtt{d}'}\rangle) \neq \mathtt{d}'(\langle M_{\mathtt{d}'}\rangle)$ in contradiction to the hypothesis that $M_{\mathtt{d}'}$ computes $\mathtt{d}'$.

We next prove that $\mathtt{d}$ is uncomputable, and thus $\mathtt{h}$ is uncomputable (because $\mathtt{d}(z) = \mathtt{h}(z,z)$ for every z). To prove that $\mathtt{d}$ is uncomputable, we show that if $\mathtt{d}$ is computable then so is $\mathtt{d}'$ (which we already know not to be the case). Indeed, let A be an algorithm for computing $\mathtt{d}$ (i.e., $A(\langle M\rangle) = d(\langle M\rangle)$ for every machine M). Then we construct an algorithm for computing $\mathtt{d}'$, which given $\langle M'\rangle$, invokes A on $\langle M''\rangle$, where M'' is defined to operate as follows:

1. On input x, machine M'' emulates M' on input x.
2. If M' halts on input x with output 0 then M'' halts.
3. If M' halts on input x with an output different from 0 then M'' enters an infinite loop (and thus does not halt).
4. Otherwise (i.e., M' does not halt on input x), then machine M'' does not halt (because it just stays stuck in Step 1 forever).

Note that the mapping from $\langle M'\rangle$ to $\langle M''\rangle$ is easily computable (by augmenting M' with instructions to test its output and enter an infinite loop if necessary), and that $\mathtt{d}(\langle M''\rangle) = \mathtt{d}'(\langle M'\rangle)$, because M'' halts on x if and only if M'' halts on x with output 0. We thus derived an algorithm for computing $\mathtt{d}'$ (i.e., transform the input $\langle M'\rangle$ into $\langle M''\rangle$ and output $A(\langle M''\rangle)$), which contradicts the already established fact by which $\mathtt{d}'$ is uncomputable. ■

Turing-reductions. The core of the second part of the proof of Theorem 1.5 is an algorithm that solves one problem (i.e., computes $\mathtt{d}'$) by using as a subroutine an algorithm that solves another problem (i.e., computes $\mathtt{d}$ (or $\mathtt{h}$)). In fact, the first algorithm is actually an algorithmic scheme that refers to a "functionally specified" subroutine rather than to an actual (implementation of such a) subroutine, which may not exist. Such an algorithmic scheme is called a Turing-reduction (see formulation in §1.2.3.5). Hence, we have Turing-reduced the computation of $\mathtt{d}'$ to the computation of $\mathtt{d}$, which in turn Turing-reduces to $\mathtt{h}$. The "natural" ("positive") meaning of a Turing-reduction of f' to f is that when given an algorithm for computing f we obtain an algorithm for computing f'. In contrast, the proof of Theorem 1.5 uses the "unnatural" ("negative") counter-positive: if (as we know) there exists no algorithm for computing $f' = \mathtt{d}'$ then there exists no algorithm for computing $f = \mathtt{d}$ (which is what we wanted to prove). Jumping ahead, we mention that resource-bounded Turing-reductions (e.g., polynomial-time reductions) play a central role in complexity theory itself, and again they are used mostly in a "negative" way. We will define such reductions and extensively use them in subsequent chapters.

Rice's Theorem. The undecidability of the halting problem (or rather the fact that the function `d` is uncomputable) is a special case of a more general phenomenon: Every non-trivial decision problem *regarding the function computed by a given Turing machine* has no algorithmic solution. We state this fact next, clarifying what is the aforementioned class of problems. (Again, we refer to Turing machines that may not halt on all inputs.)

Theorem 1.6 (Rice's Theorem): *Let $\mathcal{F}$ be a non-trivial subset*[12] *of the set of all computable partial functions, and let $S_{\mathcal{F}}$ be the set of strings that describe machines that compute functions in $\mathcal{F}$. Then deciding membership in $S_{\mathcal{F}}$ cannot be solved by an algorithm.*

Theorem 1.6 can be proved by a Turing-reduction from `d`. We do not provide a proof because this is too remote from the main subject matter of the book. We stress that Theorems 1.5 and 1.6 hold for any reasonable model of computation (referring both to the potential solvers and to the machines the description of which is given as input to these solvers). Thus, Theorem 1.6 means that *no algorithm can determine any non-trivial property of the function computed by a given computer program* (written in any programming language). For example, *no algorithm can determine whether or not a given computer program halts on each possible input.* The relevance of this assertion to the project of program verification is obvious.

The Post Correspondence Problem. We mention that undecidability arises also outside of the domain of questions regarding computing devices (given as input). Specifically, we consider the Post Correspondence Problem in which the input consists of two sequences of strings, $(\alpha_1, ..., \alpha_k)$ and $(\beta_1, ..., \beta_k)$, and the question is whether or not there exists a sequence of indices $i_1, ..., i_\ell \in \{1, ..., k\}$ such that $\alpha_{i_1} \cdots \alpha_{i_\ell} = \beta_{i_1} \cdots \beta_{i_\ell}$. (We stress that the length of this sequence is not bounded.)[13]

Theorem 1.7 *The Post Correspondence Problem is undecidable.*

Again, the omitted proof is by a Turing-reduction from `d` (or `h`).[14]

1.2.3.3 Universal algorithms

So far we have used the postulate that, in any reasonable model of computation, each machine (or computation rule) has a finite description. Furthermore, we also used the fact that such model should allow for the easy modification of such descriptions such that the resulting machine computes an easily related function

[12]The set S is called a non-trivial subset of U if both S and $U \setminus S$ are non-empty. Clearly, if $\mathcal{F}$ is a trivial set of computable functions then the corresponding decision problem can be solved by a "trivial" algorithm that outputs the corresponding constant bit.

[13]In contrast, the existence of an adequate sequence of a specified length can be determined in time that is exponential in this length.

[14]We mention that the reduction maps an instance $(\langle M \rangle, x)$ of `h` to a pair of sequences such that only the first string in each sequence depends on x, whereas the other strings as well as their number depend only on M.

(see the proof of Theorem 1.5). Here we go one step further and postulate that the description of machines (in this model) is "effective" in the following natural sense: there exists an algorithm that, given a description of a machine (resp., computation rule) and a corresponding environment, determines the environment that results from performing a single step of this machine on this environment (resp. the effect of a single application of the computation rule). This algorithm can, in turn, be implemented in the said model of computation (assuming this model is general; see the Church-Turing Thesis). Successive applications of this algorithm leads to the notion of a universal machine, which (for concreteness) is formulated next in terms of Turing machines.

Definition 1.8 (universal machines): *A* universal Turing machine *is a Turing machine that on input a description of a machine M and an input x returns the value of $M(x)$ if M halts on x and otherwise does not halt.*

That is, a universal Turing machine computes the partial function $\mathtt{u}$ that is defined over pairs $(\langle M\rangle, x)$ such that M halts on input x, in which case it holds that $\mathtt{u}(\langle M\rangle, x) = M(x)$. We note that if M halts on all possible inputs then $\mathtt{u}(\langle M\rangle, x)$ is defined for every x. We stress that the mere fact that we have defined something does not mean that it exists. Yet, as hinted in the foregoing discussion and obvious to anyone who has written a computer program (and thought about what he/she was doing), universal Turing machines do exist.

Theorem 1.9 *There exists a universal Turing machine.*

Theorem 1.9 asserts that the partial function $\mathtt{u}$ is computable. In contrast, it can be shown that any extension of $\mathtt{u}$ to a total function is uncomputable. That is, for any total function $\hat{\mathtt{u}}$ that agrees with the partial function $\mathtt{u}$ on all the inputs on which the latter is defined, it holds that $\hat{\mathtt{u}}$ is uncomputable.[15]

Proof: Given a pair $(\langle M\rangle, x)$, we just emulate the computation of machine M on input x. This emulation is straightforward, because (by the effectiveness of the description of M) we can iteratively determine the next instantaneous configuration of the computation of M on input x. If the said computation halts then we will obtain its output and can output it (and so, on input $(\langle M\rangle, x)$, our algorithm returns $M(x)$). Otherwise, we turn out emulating an infinite computation, which means that our algorithm does not halt on input $(\langle M\rangle, x)$. Thus, the foregoing emulation procedure constitutes a universal machine (i.e., yields an algorithm for computing $\mathtt{u}$). ∎

As hinted already, the existence of universal machines is the fundamental fact underlying the paradigm of general-purpose computers. Indeed, a specific Turing

[15] The claim is easy to prove for the total function $\hat{\mathtt{u}}$ that extends $\mathtt{u}$ and assigns the special symbol $\perp$ to inputs on which $\mathtt{u}$ is undefined (i.e., $\hat{\mathtt{u}}(\langle M\rangle, x) \stackrel{\text{def}}{=} \perp$ if $\mathtt{u}$ is not defined on $(\langle M\rangle, x)$ and $\hat{\mathtt{u}}(\langle M\rangle, x) \stackrel{\text{def}}{=} \mathtt{u}(\langle M\rangle, x)$ otherwise). In this case $\mathtt{h}(\langle M\rangle, x) = 1$ if and only if $\hat{\mathtt{u}}(\langle M\rangle, x) \neq \perp$, and so the halting function $\mathtt{h}$ is Turing-reducible to $\hat{\mathtt{u}}$. In the general case, we may adapt the proof of Theorem 1.5 by using the fact that, for a machine M that halts on every input, it holds that $\hat{\mathtt{u}}(\langle M\rangle, x) = \mathtt{u}(\langle M\rangle, x)$ for every x (and in particular for $x = \langle M\rangle$).

machine (or algorithm) is a device that solves a specific problem. A priori, solving each problem would have required building a new physical device that allows for this problem to be solved in the physical world (rather than as a thought experiment). The existence of a universal machine asserts that it is enough to build one physical device; that is, a general purpose computer. Any specific problem can then be solved by writing a corresponding program to be executed (or emulated) by the general purpose computer. Thus, universal machines correspond to general purpose computers, and provide the basis for separating hardware from software. In other words, the existence of universal machines says that software can be viewed as (part of the) input.

In addition to their practical importance, the existence of universal machines (and their variants) has important consequences in the theories of computability and computational complexity. Here we merely note that Theorem 1.9 implies that many questions about the behavior of a universal machine on certain input types are undecidable. For example, it follows that, for some fixed machines (i.e., universal ones), there is no algorithm that determines whether or not the (fixed) machine halts on a given input. Revisiting the proof of Theorem 1.7 (see Footnote 14), it follows that the Post Correspondence Problem remains undecidable even if the input sequences are restricted to have a specific length (i.e., k is fixed). A more important application of universal machines to the theory of computability follows.

A detour: Kolmogorov Complexity. The existence of universal machines, which may be viewed as universal languages for writing effective and succinct descriptions of objects, plays a central role in Kolmogorov Complexity. Loosely speaking, the latter theory is concerned with the length of (effective) descriptions of objects, and views the minimum such length as the inherent "complexity" of the object; that is, "simple" objects (or phenomena) are those having short description (resp., short explanation), whereas "complex" objects have no short description. Needless to say, these (effective) descriptions have to refer to some fixed "language" (i.e., to a fixed machine that, given a succinct description of an object, produces its explicit description). Fixing any machine M, a string x is called a description of s with respect to M if $M(x) = s$. The complexity of s with respect to M, denoted $K_M(s)$, is the length of the shortest description of s with respect to M. Certainly, we want to fix M such that every string has a description with respect to M, and furthermore such that this description is not "significantly" longer than the description with respect to a different machine M'. The following theorem make it natural to use a universal machine as the "point of reference" (i.e., as the aforementioned M).

Theorem 1.10 (complexity w.r.t a universal machine): *Let U be a universal machine. Then, for every machine M', there exists a constant c such that $K_U(s) \leq K_{M'}(s) + c$ for every string s.*

The theorem follows by (setting $c = O(|\langle M' \rangle|)$ and) observing that if x is a description of s with respect to M' then $(\langle M' \rangle, x)$ is a description of s with respect to U. Here it is important to use an adequate encoding of pairs of strings (e.g., the pair $(\sigma_1 \cdots \sigma_k, \tau_1 \cdots \tau_\ell)$ is encoded by the string $\sigma_1\sigma_1 \cdots \sigma_k\sigma_k 01\tau_1 \cdots \tau_\ell$). Fixing any universal machine U, we define the Kolmogorov Complexity of a string s as $K(s) \stackrel{\rm def}{=} K_U(s)$. The reader may easily verify the following facts:

1. $K(s) \leq |s| + O(1)$, for every s.

   (Hint: apply Theorem 1.10 to a machine that computes the identity mapping.)

2. There exist infinitely many strings s such that $K(s) \ll |s|$.

   (Hint: consider $s = 1^n$. Alternatively, consider any machine M such that $|M(x)| \gg |x|$ for every x.)

3. Some strings of length n have complexity at least n. Furthermore, for every n and i,
   $$|\{s \in \{0,1\}^n : K(s) \leq n - i\}| < 2^{n-i+1}$$
   (Hint: different strings must have different descriptions with respect to U.)

It can be shown that *the function K is uncomputable*. The proof is related to the paradox captured by the following "description" of a natural number: `the largest natural number that can be described by an English sentence of up-to a thousand letters`. (The paradox amounts to observing that if the above number is well-defined then so is `the integer-successor of the largest natural number that can be described by an English sentence of up-to a thousand letters`.) Needless to say, the foregoing sentences presuppose that any English sentence is a legitimate description in some adequate sense (e.g., in the sense captured by Kolmogorov Complexity). Specifically, the foregoing sentences presuppose that we can determine the Kolmogorov Complexity of each natural number, and furthermore that we can effectively produce the largest number that has Kolmogorov Complexity not exceeding some threshold. Indeed, the paradox provides a proof to the fact that the latter task cannot be performed; that is, there exists no algorithm that given t produces the lexicographically last string s such that $K(s) \leq t$, because if such an algorithm A would have existed then $K(s) \leq O(|\langle A \rangle|) + \log t$ and $K(s0) < K(s) + O(1) < t$ in contradiction to the definition of s.

1.2.3.4 Time and space complexity

Fixing a model of computation (e.g., Turing machines) and focusing on algorithms that halt on each input, we consider the number of steps (i.e., applications of the computation rule) taken by the algorithm on each possible input. The latter function is called the time complexity of the algorithm (or machine); that is, $t_A : \{0,1\}^* \to \mathbb{N}$ is called the time complexity of algorithm A if, for every x, on input x algorithm A halts after exactly $t_A(x)$ steps.

We will be mostly interested in the dependence of the time complexity on the input length, when taking the maximum over all inputs of the relevant length. That is, for t_A as above, we will consider $T_A : \mathbb{N} \to \mathbb{N}$ defined by $T_A(n) \stackrel{\rm def}{=}$

$\max_{x\in\{0,1\}^n}\{t_A(x)\}$. Abusing terminology, we sometimes refer to T_A as the time complexity of A.

The time complexity of a problem. As stated in the preface and in the introduction, typically is complexity theory not concerned with the (time) complexity of a specific algorithm. It is rather concerned with the (time) complexity of a problem, assuming that this problem is solvable at all (by some algorithm). Intuitively, the time complexity of such a problem is defined as the time complexity of the fastest algorithm that solves this problem (assuming that the latter term is well-defined).[16] More generally, we will be interested in upper and lower bounds on the (time) complexity of algorithms that solve the problem. However, the complexity of a problem may depend on the specific model of computation in which algorithms that solve it are implemented. The following Cobham-Edmonds Thesis asserts that the variation (in the time complexity) is not too big, and in particular is irrelevant to much of the current focus of complexity theory (e.g., for the P-vs-NP Question).

The Cobham-Edmonds Thesis. As just stated, the time complexity of a problem may depend on the model of computation. For example, deciding membership in the set $\{xx : x \in \{0,1\}^*\}$ can be done in linear-time on a two-tape Turing machine, but requires quadratic-time on a single-tape Turing machine.[17] On the other hand, any problem that has time complexity t in the model of multi-tape Turing machines, has complexity $O(t^2)$ in the model of single-tape Turing machines. The Cobham-Edmonds Thesis asserts that the time complexities in any two "reasonable and general" models of computation are polynomially related. That is, *a problem has time complexity t in some* "reasonable and general" *model of computation if and only if it has time complexity* $\mathrm{poly}(t)$ *in the model of* (single-tape) *Turing machines.*

Indeed, the Cobham-Edmonds Thesis strengthens the Church-Turing Thesis. It asserts not only that the class of solvable problems is invariant as far as "reasonable and general" models of computation are concerned, but also that the time complexity (of the solvable problems) in such models is polynomially related.

Efficient algorithms. As hinted in the foregoing discussions, much of complexity theory is concerned with efficient algorithms. The latter are defined as polynomial-time algorithms (i.e., algorithms that have a time complexity that is bounded by a polynomial in the length of the input). By the Cobham-Edmonds Thesis, the

[16] **Advanced comment:** As we shall see in Section 4.2.2 (cf. Theorem 4.8), the naive assumption that a "fastest algorithm" for solving a problem exists is not always justified. On the other hand, the assumption is justified in some important cases (see, e.g., Theorem 2.31).

[17] Proving the latter fact is quite non-trivial. One proof is by a "reduction" from a communication complexity problem [141, Sec. 12.2]. Intuitively, a single-tape Turing machine that decides membership in the aforementioned set can be viewed as a channel of communication between the two parts of the input. Focusing our attention on inputs of the form $y0^nz0^n$, for $y,z \in \{0,1\}^n$, each time the machine passes from the first part to the second part it carries $O(1)$ bits of information (in its internal state) while making at least n steps. The proof is completed by invoking the linear lower bound on the communication complexity of the (two-argument) identity function (i.e, $\mathtt{id}(y,z) = 1$ if $y = z$ and $\mathtt{id}(y,z) = 0$ otherwise, cf. [141, Chap. 1]).

choice of a "reasonable and general" model of computation is irrelevant to the definition of this class. The association of efficient algorithms with polynomial-time computation is grounded in the following two considerations:

- *Philosophical consideration*: Intuitively, efficient algorithms are those that can be implemented within a number of steps that is a moderately growing function of the input length. To allow for reading the entire input, at least linear time complexity should be allowed, whereas exponential time (as in "exhaustive search") must be avoided. Furthermore, a good definition of the class of efficient algorithms should be closed under natural composition of algorithms (as well as be robust with respect to reasonable models of computation and with respect to simple changes in the encoding of problems' instances).

  Selecting polynomials as the set of time-bounds for efficient algorithms satisfy all the foregoing requirements: polynomials constitute a "closed" set of moderately growing functions, where "closure" means closure under addition, multiplication and functional composition. These closure properties guarantee the closure of the class of efficient algorithm under natural composition of algorithms (as well as its robustness with respect to any reasonable and general model of computation). Furthermore, polynomial-time algorithms can conduct computations that are intuitively simple (although not necessarily trivial), and on the other hand they do not include algorithms that are intuitively inefficient (like exhaustive search).

- *Empirical consideration*: It is clear that algorithms that are considered efficient in practice have running-time that is bounded by a small polynomial (at least on the inputs that occur in practice). The question is whether any polynomial-time algorithm can be considered efficient in an intuitive sense. The belief, which is supported by past experience, is that every natural problem that can be solved in polynomial-time also has "reasonably efficient" algorithms.

We stress that the association of efficient algorithms with polynomial-time computation is not essential to most of the notions, results and questions of complexity theory. Any other class of algorithms that supports the aforementioned closure properties and allows to conduct some simple computations but not overly complex ones gives rise to a similar theory, albeit the formulation of such a theory may be much more complicated. Specifically, all results and questions treated in this book are concerned with the relation among the complexities of different computational tasks (rather than with providing absolute assertions about the complexity of some computational tasks). These relations can be stated explicitly, by stating how any upper-bound on the time complexity of one task gets translated to an upper-bound on the time complexity of another task.[18] Such cumbersome statements will maintain the contents of the standard statements; they will merely be

[18]For example, the NP-completeness of SAT (cf. Theorem 2.21) implies that any algorithm solving SAT in time T yields an algorithm that factors composite numbers in time T' such that $T'(n) = \text{poly}(n) \cdot (1 + T(\text{poly}(n)))$. (More generally, if the correctness of solutions for n-bit

much more complicated. Thus, we follow the tradition of focusing on polynomial-time computations, while stressing that this focus is both natural and provides the simplest way of addressing the fundamental issues underlying the nature of efficient computation.

Universal machines, revisited. The notion of time complexity gives rise to a time-bounded version of the universal function $\mathtt{u}$ (presented in §1.2.3.3). Specifically, we define $\mathtt{u}'(\langle M\rangle, x, t) \stackrel{\rm def}{=} y$ if on input x machine M halts within t steps and outputs the string y, and $\mathtt{u}'(\langle M\rangle, x, t) \stackrel{\rm def}{=} \perp$ if on input x machine M makes more than t steps. Unlike $\mathtt{u}$, the function $\mathtt{u}'$ is a total function. Furthermore, unlike any extension of $\mathtt{u}$ to a total function the function $\mathtt{u}'$ is computable. Moreover, $\mathtt{u}'$ is computable by a machine U' that on input $X = (\langle M\rangle, x, t)$ halts after poly(t) steps. Indeed, machine U' is a variant of a universal machine (i.e., on input X, machine U' merely emulates M for t steps rather than emulating M till it halts (and potentially indefinitely)). Note that the number of steps taken by U' depends on the specific model of computation (and that some overhead is unavoidable because emulating each step of M requires reading the relevant portion of the description of M).

Space complexity. Another natural measure of the "complexity" of an algorithm (or a task) is the amount of memory consumed by the computation. We refer to the memory used for storing some intermediate results of the computation. Since much of our focus will be on using memory that is sub-linear in the input length, it is important to use a model in which one can differentiate memory used for computation from memory used for storing the initial input or the final output. In the context of Turing machines, this is done by considering multi-tape Turing machines such that the input is presented on a special read-only tape (called the input tape), the output is written on a special write-only tape (called the output tape), and intermediate results are stored on a work-tape. Thus, the input and output tapes cannot be used for storing intermediate results. The space complexity of such a machine M is defined as a function s_M such that $s_M(x)$ is the number of cells of the work-tape that are scanned by M on input x. As in the case of time complexity, we will usually refer to $S_A(n) \stackrel{\rm def}{=} \max_{x\in\{0,1\}^n}\{s_A(x)\}$.

1.2.3.5 Oracle machines

The notion of Turing-reductions, which was discussed in §1.2.3.2, is captured by the following definition of so-called *oracle machines*. Loosely speaking, an oracle machine is a machine that is augmented such that it may pose questions to the outside. (A rigorous formulation of this notion is provided below.) We consider the case in which these questions, called queries, are answered consistently by some function $f : \{0,1\}^* \rightarrow \{0,1\}^*$, called the oracle. That is, if the machine makes a query q then the answer it obtains is $f(q)$. In such a case, we say that the oracle

instances of some search problem can be verified in time $t(n)$ then such solutions can be found in time T' such that $T'(n) = t(n) \cdot (1 + T(O(t(n))^2)).)$

machine is given access to the oracle f. For an oracle machine M, a string x and a function f, we denote by $M^f(x)$ the output of M on input x when given access to the oracle f. (Re-examining the second part of the proof of Theorem 1.5, observe that we have actually described an oracle machine that computes d' when given access to the oracle d.)

The notion of an oracle machine extends the notion of a standard computing device (machine), and thus a rigorous formulation of the former extends a formal model of the latter. Specifically, extending the model of Turing machines, we derive the following model of oracle Turing machines.

Definition 1.11 (using an oracle): *An* oracle machine *is a Turing machine with an additional tape, called the* oracle tape*, and two special states, called* oracle invocation *and* oracle spoke*. The* computation of the oracle machine M on input x and access to the oracle $f : \{0,1\}^* \to \{0,1\}^*$ *is defined based on the successive configuration function. For configurations with state different from* oracle invocation *the next configuration is defined as usual. Let γ be a configuration in which the machine's state is* oracle invocation *and suppose that the actual contents of the oracle tape is q (i.e., q is the contents of the maximal prefix of the tape that holds bit values).*[19] *Then, the configuration following γ is identical to γ, except that the state is* oracle spoke*, and the actual contents of the oracle tape is $f(q)$. The string q is called M's* query *and $f(q)$ is called the* oracle's reply*.*

We stress that the running time of an oracle machine is the number of steps made during its computation, and that the oracle's reply on each query is obtained in a single step.

1.2.3.6 Restricted models

We mention that restricted models of computation are often mentioned in the context of a course on computability, but they will play no role in the current book. One such model is the model of finite automata, which in some variant coincides with Turing machines that have space complexity zero.

In our opinion, the most important motivation for the study of these restricted models of computation is that they provide simple models for some natural (or artificial) phenomena. This motivation, however, seems only remotely related to the study of the complexity of various computational tasks. Thus, in our opinion, the study of these restricted models (e.g., any of the lower levels of Chomsky's Hierarchy [118, Chap. 9]) should be decoupled from the study of computability theory (let alone the study of complexity theory).

[19]This fits the definition of the *actual contents of a tape of a Turing machine* (cf. §1.2.3.1). A common convention is that the oracle can be invoked only when the machine's head resides at the left-most cell of the oracle tape. We comment that, in the context of space complexity, one uses two oracle tapes: a write-only tape for the query and a read-only tape for the answer.

1.2.4 Non-uniform Models (Circuits and Advice)

By a non-uniform model of computation we mean a model in which for each possible input length one considers a different computing device. That is, there is no "uniformity" requirement relating devices that correspond to different input lengths. Furthermore, this collection of devices is infinite by nature, and (in absence of a uniformity requirement) this collection may not even have a finite description. Nevertheless, each device in the collection has a finite description. In fact, the relationship between the size of the device (resp., the length of its description) and the length of the input that it handles will be of major concern. The hope is that the finiteness of all parameters (which refer to a single device in such a collection) will allow for the application of combinatorial techniques to analyze the limitations of certain settings of parameters.

In complexity theory, non-uniform models of computation are studied either towards the development of lower-bound techniques or as simplified upper-bounds on the ability of efficient algorithms. In both cases, the uniformity condition is eliminated in the interest of simplicity and with the hope (and belief) that nothing substantial is lost as far as the issues at hand are concerned.

We will focus on two related models of non-uniform computing devices: Boolean circuits (§1.2.4.1) and "machines that take advice" (§1.2.4.2). The former model is more adequate for the study of the evolution of computation (i.e., development of lower-bound techniques), whereas the latter is more adequate for modeling purposes (e.g., upper-bounding the ability of efficient algorithms). (These models will be further studied in Sections 3.1 and 4.1.)

1.2.4.1 Boolean Circuits

The most popular model of non-uniform computation is the one of Boolean circuits. Historically, this model was introduced for the purpose of describing the "logic operation" of real-life electronic circuits. Ironically, nowadays this model provides the stage for some of the most practically removed studies in complexity theory (which aim at developing methods that may eventually lead to an understanding of the inherent limitations of efficient algorithms).

A Boolean circuit is a directed acyclic graph *with labels on the vertices*, to be discussed shortly. For sake of simplicity, we disallow isolated vertices (i.e., vertices with no in-going or out-going edges), and thus the graph's vertices are of three types: *sources*, *sinks*, and *internal vertices.*

1. Internal vertices are vertices having in-coming and out-going edges (i.e., they have in-degree and out-degree at least 1). In the context of Boolean circuits, internal vertices are called gates. Each gate is labeled by a Boolean operation, where the operations that are typically considered are $\wedge$, $\vee$ and $\neg$ (corresponding to and, or and neg). In addition, we require that gates labeled $\neg$ have in-degree 1. (The in-coming degree of $\wedge$-gates and $\vee$-gates may be any number greater than zero, and the same holds for the out-degree of any gate.)

2. The graph sources (i.e., vertices with no in-going edges) are called input terminals. Each input terminal is labeled by a natural number (which is to be thought of the index of an input variable). (For sake of defining formulae (see §1.2.4.3), we allow different input terminals to be labeled by the same number.)[20]

3. The graph sinks (i.e., vertices with no out-going edges) are called output terminals, and we require that they have in-degree 1. Each output terminal is labeled by a natural number such that if the circuit has m output terminals then they are labeled $1, 2, ..., m$. That is, we disallow different output terminals to be labeled by the same number, and insist that the labels of the output terminals are consecutive numbers. (Indeed, the labels of the output terminals will correspond to the indices of locations in the circuit's output.)

For sake of simplicity, we also mandate that the labels of the input terminals are consecutive numbers.[21]

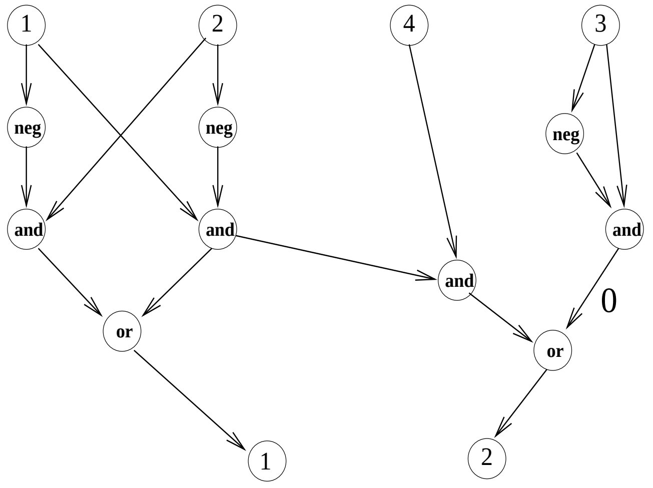


Figure 1.2: A circuit computing $f(x_1, x_2, x_3, x_4) = (x_1 \oplus x_2, x_1 \wedge \neg x_2 \wedge x_4)$.

A Boolean circuit with n different input labels and m output terminals induces (and indeed computes) a function from $\{0,1\}^n$ to $\{0,1\}^m$ defined as follows. For any fixed string $x \in \{0,1\}^n$, we iteratively define the value of vertices in the circuit

[20]This is not needed in case of general circuits, because we can just feed out-going edges of the same input terminal to many gates. Note, however, that this is not allowed in case of formulae, where all non-sinks are required to have out-degree exactly 1.

[21]This convention slightly complicates the construction of circuits that ignore some of the input values. Specifically, we use artificial gadgets that have in-coming edges from the corresponding input terminals, and compute an adequate constant. To avoid having this constant as an output terminal, we feed it into an auxiliary gate such that the value of the latter is determined by the other in-going edge (e.g., a constant 0 fed into an $\vee$-gate). See example of dealing with x_3 in Figure 1.2.

such that the input terminals are assigned the corresponding bits in $x = x_1 \cdots x_n$ and the values of other vertices are determined in the natural manner. That is:

- An input terminal with label $i \in \{1, ..., n\}$ is assigned the i^{th} bit of x (i.e., the value x_i).
- If the children of a gate (of in-degree d) that is labeled $\wedge$ have values $v_1, v_2, ..., v_d$, then the gate is assigned the value $\wedge_{i=1}^{d} v_i$. The value of a gate labeled $\vee$ (or $\neg$) is determined analogously.

  Indeed, the hypothesis that the circuit is acyclic implies that the process of determining values for the circuit's vertices is well-defined: As long as the value of some vertex is undetermined, there exists a vertex such that its value is undetermined but the values of all its children are determined. Thus, the process can make progress, and terminates when the values of all vertices (including the output terminals) are determined.

The value of the circuit on input x (i.e., the output computed by the circuit on input x) is $y = y_1 \cdots y_m$, where y_i is the value assigned by the foregoing process to the output terminal labeled i. We note that *there exists a polynomial-time algorithm that, given a circuit C and a corresponding input x, outputs the value of C on input x.* This algorithm determines the values of the circuit's vertices, going from the circuit's input terminals to its output terminals.

We say that a family of circuits $(C_n)_{n \in \mathbb{N}}$ computes a function $f : \{0,1\}^* \to \{0,1\}^*$ if for every n the circuit C_n computes the restriction of f to strings of length n. In other words, for every $x \in \{0,1\}^*$, it must hold that $C_{|x|}(x) = f(x)$.

Bounded and unbounded fan-in. We will be most interested in circuits in which each gate has at most two in-coming edges. In this case, the types of (two-argument) Boolean operations that we allow is immaterial (as long as we consider a "full basis" of such operations; i.e., a set of operations that can implement any other two-argument Boolean operation). Such circuits are called circuits of bounded fan-in. In contrast, other studies are concerned with circuits of unbounded fan-in, where each gate may have an arbitrary number of in-going edges. Needless to say, in the case of circuits of unbounded fan-in, the choice of allowed Boolean operations is important and one focuses on operations that are "uniform" (across the number of operants; e.g., $\wedge$ and $\vee$).

Circuit size as a complexity measure. The size of a circuit is the number of its edges. When considering a family of circuits $(C_n)_{n \in \mathbb{N}}$ that computes a function $f : \{0,1\}^* \to \{0,1\}^*$, we are interested in the size of C_n as a function of n. Specifically, we say that this family has size complexity $s : \mathbb{N} \to \mathbb{N}$ if for every n the size of C_n is $s(n)$. The circuit complexity of a function f, denoted s_f, is the infimum of the size complexity of all families of circuits that compute f. Alternatively, for each n we may consider the size of the smallest circuit that computes the restriction of f to n-bit strings (denoted f_n), and set $s_f(n)$ accordingly. We stress that non-uniformity is implicit in this definition, because no conditions are made regarding the relation between the various circuits used to compute the function on different input lengths.

The circuit complexity of functions. We highlight some simple facts about the circuit complexity of functions. (These facts are in clear correspondence to facts regarding Kolmogorov Complexity mentioned in §1.2.3.3.)

1. Most importantly, any Boolean function can be computed by some family of circuits, and thus the circuit complexity of any function is well-defined. Furthermore, each function has at most exponential circuit complexity.

   (Hint: $f_n : \{0,1\}^n \to \{0,1\}$ can be computed by a circuit of size $O(n2^n)$ that implements a look-up table.)

2. Some functions have polynomial circuit complexity. In particular, any function that has time complexity t (i.e., is computed by an algorithm of time complexity t) has circuit complexity poly(t). Furthermore, the corresponding circuit family is uniform (in a natural sense to be discussed in the next paragraph).

   (Hint: consider a Turing machine that computes the function, and consider its computation on a generic n-bit long input. The corresponding computation can be emulated by a circuit that consists of $t(n)$ layers such that each layer represents an instantaneous configuration of the machine, and the relation between consecutive configurations is captured by ("uniform") local gadgets in the circuit. For further details see the proof of Theorem 2.20, which presents a similar emulation.)

3. Almost all Boolean functions have exponential circuit complexity. Specifically, the number of functions mapping $\{0,1\}^n$ to $\{0,1\}$ that can be computed by some circuit of size s is at most s^{2s}.

   (Hint: the number of circuits having v vertices and s edges is at most $\binom{v}{2}^s$.)

Note that the first fact implies that families of circuits can compute functions that are uncomputable by algorithms. Furthermore, this phenomenon occurs also when restricting attention to families of polynomial-size circuits. See further discussion in §1.2.4.2.

Uniform families. A family of polynomial-size circuits $(C_n)_{n\in\mathbb{N}}$ is called uniform if given n one can construct the circuit C_n in poly(n)-time. Note that *if a function is computable by a uniform family of polynomial-size circuits then it is computable by a polynomial-time algorithm.* This algorithm first constructs the adequate circuit (which can be done in polynomial-time by the uniformity hypothesis), and then evaluate this circuit on the given input (which can be done in time that is polynomial in the size of the circuit).

Note that limitations on the computing power of arbitrary families of polynomial-size circuits certainly hold for uniform families (of polynomial-size), which in turn yield limitations on the computing power of polynomial-time algorithms. Thus, lower bounds on the circuit complexity of functions yield analogous lower bounds on their time complexity. Furthermore, as is often the case in mathematics and Science, disposing of an auxiliary condition that is not well-understood (i.e., uniformity) may turn out fruitful. Indeed, this has occured in the study of classes of restricted circuits, which is reviewed in §1.2.4.3 (and Appendix B).

1.2.4.2 Machines that take advice

General (non-uniform) circuit families and uniform circuit families are two extremes with respect to the "amounts of non-uniformity" in the computing device. Intuitively, in the former, non-uniformity is only bounded by the size of the device, whereas in the latter the amounts of non-uniformity is zero. Here we consider a model that allows to decouple the size of the computing device from the amount of non-uniformity, which may range from zero to the device's size. Specifically, we consider algorithms that "take a non-uniform advice" that depends only on the input length. The amount of non-uniformity will be defined to equal the length of the corresponding advice (as a function of the input length).

Definition 1.12 (taking advice): *We say that* algorithm A computes the function f using advice of length $\ell : \mathbb{N} \to \mathbb{N}$ *if there exists an infinite sequence* $(a_n)_{n\in\mathbb{N}}$ *such that*

1. *For every* $x \in \{0,1\}^*$*, it holds that* $A(a_{|x|}, x) = f(x)$*.*
2. *For every* $n \in \mathbb{N}$*, it holds that* $|a_n| = \ell(n)$*.*

The sequence $(a_n)_{n\in\mathbb{N}}$ *is called the* advice sequence*.*

Note that any function having circuit complexity s can be computed using advice of length $O(s \log s)$, where the log factor is due to the fact that a graph with v vertices and e edges can be described by a string of length $2e \log_2 v$. Note that the model of machines that use advice allows for some sharper bounds than the ones stated in §1.2.4.1: every function can be computed using advice of length ℓ such that $\ell(n) = 2^n$, and some uncomputable functions can be computed using advice of length 1.

Theorem 1.13 (the power of advice): *There exist functions that can be computed using one-bit advice but cannot be computed without advice.*

Proof: Starting with any uncomputable Boolean function $f : \mathbb{N} \to \{0,1\}$, consider the function f' defined as $f'(x) = f(|x|)$. Note that f is Turing-reducible to f' (e.g., on input n make any n-bit query to f', and return the answer).[22] Thus, f' cannot be computed without advice. On the other hand, f' can be easily computed by using the advice sequence $(a_n)_{n\in\mathbb{N}}$ such that $a_n = f(n)$; that is, the algorithm merely outputs the advice bit (and indeed $a_{|x|} = f(|x|) = f'(x)$, for every $x \in \{0,1\}^*$). ■

[22]Indeed, this Turing-reduction is not efficient (i.e., it runs in exponential time in $|n| = \log_2 n$), but this is immaterial in the current context.

1.2.4.3 Restricted models

As noted in §1.2.4.1, the model of Boolean circuits allows for the introduction of many natural subclasses of computing devices. Following is a laconic review of a few of these subclasses. For more detail, see Appendix B.2. Since we shall refer to various types of Boolean formulae in the rest of this book, we suggest not to skip the following two paragraphs.

Boolean formulae. In general Boolean circuits the non-sink vertices are allowed arbitrary out-degree. This means that the same intermediate value can be re-used (without being re-computed (and while increasing the size complexity by only one unit)). Such "free" re-usage of intermediate values is disallowed in Boolean formulae, which corresponds to a Boolean expression over Boolean variables. Formally, a Boolean formula is a circuit in which all non-sink vertices have out-degree 1, which means that the underlying graph is a tree (see §G.2) and the formula as an expression can be read by traversing the tree (and registering the vertices' labels in the order traversed). Indeed, we have allowed different input terminals to be assigned the same label in order to allow formulae in which the same variable occurs multiple times. As in case of general circuits, one is interested in the size of these restricted circuits (i.e., the size of families of formulae computing various functions). We mention that quadratic lower bounds are known for the formula size of simple functions (e.g., `parity`), whereas these functions have linear circuit complexity. This discrepancy is depicted in Figure 1.3.

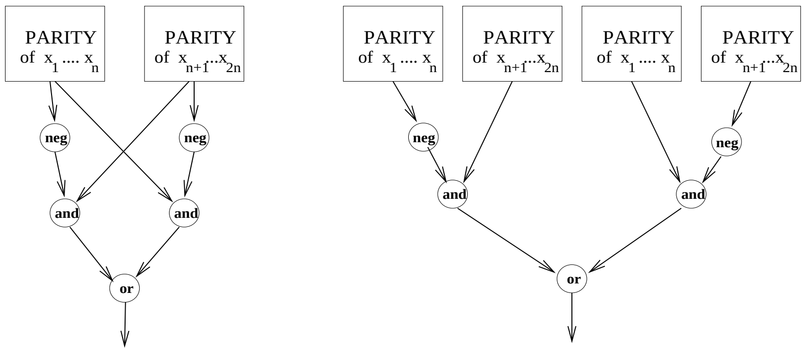


Figure 1.3: Recursive construction of parity circuits and formulae.

Formulae in CNF and DNF. A restricted type of Boolean formulae consists of formulae that are in conjunctive normal form (CNF). Such a formula consists of a conjunction of clauses, where each clause is a disjunction of literals each being either a variable or its negation. That is, such formulae are represented by layered circuits of unbounded fan-in in which the first layer consists of `neg`-gates that compute the negation of input variables, the second layer consist of `or`-gates that compute the logical-or of subsets of inputs and negated inputs, and the third layer consists of a single and-gate that computes the logical-and of the values computed in the second layer. Note that each Boolean function can be computed by a family of CNF formulae of exponential size, and that the size of CNF formulae may be exponentially larger than the size of ordinary formulae computing the same function (e.g., parity). For a constant k, a formula is said to be in k-CNF if its CNF has disjunctions of size at most k. An analogous restricted type of Boolean formulae refers to formulae that are in disjunctive normal form (DNF). Such a formula consists of a disjunction of a conjunctions of literals, and when each conjunction has at most k literals we say that the formula is in k-DNF.

Constant-depth circuits. Circuits have a "natural structure" (i.e., their structure as graphs). One natural parameter regarding this structure is the depth of a circuit, which is defined as the longest directed path from any source to any sink. Of special interest are constant-depth circuits of unbounded fan-in. We mention that sub-exponential lower bounds are known for the size of such circuits that compute a simple function (e.g., parity).

Monotone circuits. The circuit model also allows for the consideration of monotone computing devices: a monotone circuit is one having only monotone gates (e.g., gates computing $\wedge$ and $\vee$, but no negation gates (i.e., $\neg$-gates)). Needless to say, monotone circuits can only compute monotone functions, where a function $f : \{0,1\}^n \to \{0,1\}$ is called monotone if for any $x \preceq y$ it holds that $f(x) \leq f(y)$ (where $x_1 \cdots x_n \preceq y_1 \cdots y_n$ if and only if for every bit position i it holds that $x_i \leq y_i$). One natural question is whether, as far as monotone functions are concerned, there is a substantial loss in using only monotone circuits. The answer is *yes*: there exist monotone functions that have polynomial circuit complexity but require sub-exponential size monotone circuits.

1.2.5 Complexity Classes

Complexity classes are sets of computational problems. Typically, such classes are defined by fixing three parameters:

1. A type of computational problems (see Section 1.2.2). Indeed, most classes refer to decision problems, but classes of search problems, promise problems, and other types of problems will also be considered.

2. A model of computation, which may be either uniform (see Section 1.2.3) or non-uniform (see Section 1.2.4).

3. A complexity measure and a function (or a set of functions), which put together limit the class of computations of the previous item; that is, we refer to the class of computations that have complexity not exceeding the specified function (or set of functions). For example, in §1.2.3.4, we mentioned time complexity and space complexity, which apply to any uniform model of computation. We also mentioned polynomial-time computations, which are

computations in which the time complexity (as a function) does not exceed some polynomial (i.e., a member of the set of polynomial functions).

The most common complexity classes refer to decision problems, and are sometimes defined as classes of sets rather than classes of the corresponding decision problems. That is, one often says that a set $S \subseteq \{0,1\}^*$ is in the class $\mathcal{C}$ rather than saying that *the problem of deciding membership in* S is in the class $\mathcal{C}$. Likewise, one talks of classes of relations rather than classes of the corresponding search problems (i.e., saying that $R \subseteq \{0,1\}^* \times \{0,1\}^*$ is in the class $\mathcal{C}$ means that *the search problem of* R is in the class $\mathcal{C}$).

Chapter Notes

It is quite remarkable that the theories of uniform and non-uniform computational devices have emerged in two single papers. We refer to Turing's paper [213], which introduced the model of Turing machines, and to Shannon's paper [191], which introduced Boolean circuits.

In addition to introducing the Turing machine model and arguing that it corresponds to the intuitive notion of computability, Turing's paper [213] introduces universal machines and contains proofs of undecidability (e.g., of the Halting Problem).

The Church-Turing Thesis is attributed to the works of Church [53] and Turing [213]. In both works, this thesis is invoked for claiming that the fact that Turing machines cannot solve some problem implies that this problem cannot be solved in any "reasonable" model of computation. The RAM model is attributed to von Neumann's report [220].

The association of efficient computation with polynomial-time algorithms is attributed to the papers of Cobham [54] and Edmonds [66]. It is interesting to note that Cobham's starting point was his desire to present a philosophically sound concept of efficient algorithms, whereas Edmonds's starting point was his desire to articulate why certain algorithms are "good" in practice.

Rice's Theorem is proven in [182], and the undecidability of the Post Correspondence Problem is proven in [171]. The formulation of machines that take advice (as well as the equivalence to the circuit model) originates in [131].

Chapter 2

P, NP and NP-Completeness

Forasmuch as many have taken in hand to set forth in order a declaration of those things which are most surely believed among us; Even as they delivered them unto us, who from the beginning were eyewitnesses, and ministers of the word; It seems good to me also, having had perfect understanding of all things from the very first, to write unto thee in order, most excellent Theophilus; That thou mightest know the certainty of those things, wherein thou hast been instructed.

Luke, 1:1–4

The main focus of this chapter is the P-vs-NP Question and the theory of NP-completeness. Additional topics covered in this chapter include the general notion of a polynomial-time reduction (with a special emphasis on self-reducibility), the existence of problems in NP that are neither NP-complete nor in P, the class coNP, optimal search algorithms, and promise problems.

Summary: Loosely speaking, the P-vs-NP Question refers to search problems for which the correctness of solutions can be efficiently checked (i.e., there is an efficient algorithm that given a solution to a given instance determines whether or not the solution is correct). Such search problems correspond to the class NP, and the question is whether or not all these search problems can be solved efficiently (i.e., is there an efficient algorithm that given an instance finds a correct solution). Thus, the P-vs-NP Question can be phrased as asking *whether or not finding solutions is harder than checking the correctness of solutions.*

An alternative formulation, in terms of decision problems, refers to assertions that have efficiently verifiable proofs (of relatively short length). Such sets of assertions correspond to the class NP, and the question is

whether or not proofs for such assertions can be found efficiently (i.e., is there an efficient algorithm that given an assertion determines its validity and/or finds a proof for its validity). Thus, the P-vs-NP Question can be phrased as asking *whether or not discovering proofs is harder than verifying their correctness*; that is, is proving harder than verifying (or are proofs valuable at all).

Indeed, it is widely believed that the answer to the two equivalent formulations is that finding (resp., discovering) is harder than checking (resp., verifying); that is, that *P is different than NP*. The fact that this natural conjecture is unsettled seems to be one of the big sources of frustration of complexity theory. The author's opinion, however, is that this feeling of frustration is out of place. In any case, at present, when faced with a hard problem in NP, we cannot expect to prove that the problem is not in P (unconditionally). The best we can expect is a conditional proof that the said problem is not in P, based on the assumption that NP is different from P. The contrapositive is proving that if the said problem is in P, then so is any problem in NP (i.e., NP equals P). This is where the theory of NP-completeness comes into the picture.

The theory of NP-completeness is based on the notion of a reduction, which is a relation between computational problems. Loosely speaking, one computational problem is reducible to another problem if it is possible to efficiently solve the former when provided with an (efficient) algorithm for solving the latter. Thus, the first problem is not harder to solve than the second one. A problem (in NP) is NP-complete if any problem in NP is reducible to it. Thus, the fate of the entire class NP (with respect to inclusion in P) rests with each individual NP-complete problem. In particular, showing that a problem is NP-complete implies that this problem is not in P unless NP equals P. Amazingly enough, NP-complete problems exist, and furthermore hundreds of natural computational problems arising in many different areas of mathematics and science are NP-complete.

We stress that NP-complete problems are not the only hard problems in NP (i.e., if P is different than NP then NP contains problems that are neither NP-complete nor in P). We also note that the P-vs-NP Question is not about inventing sophisticated algorithms or ruling out their existence, but rather boils down to the analysis of a single known algorithm; that is, we will present an optimal search algorithm for any problem in NP, while having not clue about its time complexity.

Teaching note: Indeed, we suggest presenting the P-vs-NP Question both in terms of search problems and in terms of decision problems. Furthermore, in the latter case, we suggest introducing NP by explicitly referring to the terminology of proof systems. As for the theory of NP-completeness, we suggest emphasizing the mere existence of NP-complete problems.

Prerequisites: We assume familiarity with the notions of search and decision problems (see Section 1.2.2), algorithms (see Section 1.2.3) and their time complexity (see §1.2.3.4). We will also refer to the notion of an oracle machine (see §1.2.3.5).

Organization: In Section 2.1 we present the two formulations of the P-vs-NP Question. The general notion of a reduction is presented in Section 2.2, where we highlight its applicability outside the domain of NP-completeness. Section 2.3 is devoted to the theory of NP-completeness, whereas Section 2.4 treats three relatively advanced topics (i.e., the framework of promise problems, the existence of optimal search algorithms for NP, and the class coNP).

Teaching note: This chapter has more teaching notes than any other chapter in the book. This reflects the author's concern regarding the way in which this fundamental material is often taught. Specifically, it is the author's impression that the material covered in this chapter is often taught in wrong ways, which fail to communicate its fundamental nature.

2.1 The P versus NP Question

Our daily experience is that it is harder to solve a problem than it is to check the correctness of a solution. Is this experience merely a coincidence or does it represent a fundamental fact of life (or a property of the world)? This is the essence of the P versus NP Question, where *P represents search problems that are efficiently solvable and NP represents search problems for which solutions can be efficiently checked.*

Another natural question captured by the P versus NP Question is whether proving theorems is harder that verifying the validity of these proofs. In other words, the question is whether deciding membership in a set is harder than being convinced of this membership by an adequate proof. In this case, *P represents decision problems that are efficiently solvable, whereas NP represents sets that have efficiently checkable proofs of membership.*

These two meanings of the P versus NP Question are rigorously presented and discussed in Sections 2.1.1 and 2.1.2, respectively. The equivalence of the two formulations is shown in Section 2.1.3, and the common belief that P is different from NP is further discussed in Section 2.1.5. Let us start by recalling the notion of efficient computation.

Teaching note: Most students have heard of P and NP before, but we suspect that many have not obtained a good explanation of what the P vs NP Question actually represents. This unfortunate situation is due to using the standard technical definition of NP (which refers to the fictitious and confusing device called a non-deterministic polynomial-time machine). Instead, we advocate the use of the more cumbersome definitions sketched in the forgoing paragraphs (and elaborated in Sections 2.1.1 and 2.1.2), which clearly capture the fundamental nature of NP.

The notion of efficient computation. Recall that we associate efficient computation with polynomial-time algorithms.[1] This association is justified by the fact that polynomials are a class of moderately growing functions that is closed under operations that correspond to natural composition of algorithms. Furthermore, the class of polynomial-time algorithms is independent of the specific model of computation, as long as the latter is "reasonable" (cf. the Cobham-Edmonds Thesis). Both issues are discussed in §1.2.3.4.

A note on the representation of problem instances. As noted in Section 1.2.2, many natural (search and decision) problems are captured more naturally by the terminology of promise problems (cf. Section 2.4.1), where the domain of possible instances is a subset of $\{0,1\}^*$ rather than $\{0,1\}^*$ itself. For example, computational problems in graph theory presume some simple encoding of graphs as strings, but this encoding is typically not onto (i.e., not all strings encode graphs), and thus not all strings are legitimate instances. However, in these cases, the set of legitimate instances (e.g., encodings of graphs) is efficiently recognizable (i.e., membership in it can be decided in polynomial-time). Thus, artificially extending the set of instances to the set of all possible strings (and allowing trivial solutions for the corresponding dummy instances) does not change the complexity of the original problem. We further discuss this issue in Section 2.4.1.

2.1.1 The search version: finding versus checking

Teaching note: Complexity theorists are so accustomed to focus on decision problem that they seem to forget that search problems are at least as natural as decision problems. Furthermore, to many non-experts, search problems may seem even more natural than decision problems: Typically, people seeks solutions more than they pause to wonder whether or not solutions exist. Thus, we recommend starting with a formulation of the P-vs-NP Question in terms of search problems. Admittingly, the cost is more cumbersome formulations, but it is more than worthwhile.

Much of computer science is concerned with solving various search problems (as in Definition 1.1). Examples of such problems include finding a solution to a system of linear (or polynomial) equations, finding a prime factor of a given integer, finding a spanning tree in a graph, finding a short traveling salesman tour in a metric space, and finding a scheduling of jobs to machines such that various constraints are satisfied. Furthermore, search problems correspond to the daily notion of "solving problems" and thus are of natural general interest. In the current section, we will consider the question of *which search problems can be solved efficiently.*

One type of search problems that cannot be solved efficiently consists of search problems for which the solutions are too long in terms of the problem's instances.

[1]**Advanced comment:** In this chapter, we consider *deterministic* (polynomial-time) algorithms as the basic model of efficient computation. A more liberal view, which includes also *probabilistic* (polynomial-time) algorithms is presented in Chapter 6. We stress that the most important facts and questions that are addressed in the current chapter hold also with respect to probabilistic polynomial-time algorithms.

In such a case, merely typing the solution amounts to an activity that is deemed inefficient. Thus, we focus our attention on search problems that are not in this class. That is, we consider only search problems in which the length of the solution is bounded by a polynomial in the length of the instance. Recalling that search problems are associated with binary relations (see Definition 1.1), we focus our attention on polynomially bounded relations.

Definition 2.1 (polynomially bounded relations): *We say that $R \subseteq \{0,1\}^* \times \{0,1\}^*$ is* polynomially-bounded *if there exists a polynomial p such that for every $(x,y) \in R$ it holds that $|y| \leq p(|x|)$.*

Recall that $(x,y) \in R$ means that y is a solution to the problem instance x, where R represents the problem itself. For example, in the case of finding a prime factor of a given integer, we refer to a relation R such that $(N,P) \in R$ if P is a prime factor of N.

For a polynomially bounded relation R it makes sense to ask whether or not, given a problem instance x, one can efficiently find an adequate solution y (i.e., find y such that $(x,y) \in R$). The polynomial bound on the length of the solution (i.e., y) guarantees that a negative answer is not merely due to the length of the required solution.

2.1.1.1 The class P as a natural class of search problems

Recall that we are interested in the class of search problems that can be solved efficiently; that is, problems for which solutions (whenever they exist) can be found efficiently. Restricting our attention to polynomially bounded relations, we identify the corresponding fundamental class of search problem (or binary relation), denoted $\mathcal{PF}$ (standing for "Polynomial-time Find"). (The relationship between $\mathcal{PF}$ and the standard definition of P will be discussed in Sections 2.1.3 and 2.2.3.) The following definition refers to the formulation of solving search problems provided in Definition 1.1.

Definition 2.2 (efficiently solvable search problems):

- *The search problem of a polynomially bounded relation $R \subseteq \{0,1\}^* \times \{0,1\}^*$ is* efficiently solvable *if there exists a polynomial time algorithm A such that, for every x, it holds that $A(x) \in R(x) \stackrel{\text{def}}{=} \{y : (x,y) \in R\}$ if and only if $R(x)$ is not empty. Furthermore, if $R(x) = \emptyset$ then $A(x) = \perp$, indicating that x has no solution.*

- *We denote by $\mathcal{PF}$ the class of search problems that are efficiently solvable* (and correspond to polynomially bounded relations). *That is, $R \in \mathcal{PF}$ if R is polynomially bounded and there exists a polynomial time algorithm that given x finds y such that $(x,y) \in R$* (or asserts that no such y exists).

Note that $R(x)$ denotes the set of valid solutions for the problem instance x. Thus, the solver A is required to find a valid solution (i.e., satisfy $A(x) \in R(x)$) whenever such a solution exists (i.e., $R(x)$ is not empty). On the other hand, if the instance x has no solution (i.e., $R(x) = \emptyset$) then clearly $A(x) \notin R(x)$. The extra condition (also made in Definition 1.1) requires that in this case $A(x) = \perp$. Thus, algorithm A always outputs a correct answer, which is a valid solution in the case that such a solution exists and otherwise provides an indication that no solution exists.

We have defined a fundamental class of problems, and we do know of many natural problems in this class (e.g., solving linear equations over the rationals, finding a perfect matching in a graph, etc). However, we must admit that we do not have a good understanding regarding the actual contents of this class (i.e., we are unable to characterize many natural problems with respect to membership in this class). This situation is quite common in complexity theory, and seems to be a consequence of the fact that complexity classes are defined in terms of the "external behavior" (of potential algorithms) rather than in terms of the "internal structure" (of the problem). Turning back to $\mathcal{PF}$, we note that, while it contains many natural search problems, there are also many natural search problems that are not known to be in $\mathcal{PF}$. A natural class containing a host of such problems is presented next.

2.1.1.2 The class NP as another natural class of search problems

Natural search problems have the property that valid solutions can be efficiently recognized. That is, given an instance x of the problem R and a candidate solution y, one can efficiently determine whether or not y is a valid solution for x (with respect to the problem R; i.e., whether or not $y \in R(x)$). The class of all such search problems is a natural class *per se*, because it is not clear why one should care about a solution unless one can recognize a valid solution once given. Furthermore, this class is a natural domain of candidates for $\mathcal{PF}$, because the ability to efficiently recognize a valid solution seems to be a natural (albeit not absolute) prerequisite for a discussion regarding the complexity of finding such solutions.

We restrict our attention again to polynomially bounded relations, and consider the class of relations for which membership of pairs in the relation can be decided efficiently. We stress that we consider deciding membership of given pairs of the form (x, y) in a fixed relation R, and not deciding membership of x in the set $S_R \stackrel{\text{def}}{=} \{x : R(x) \neq \emptyset\}$. (The relationship between the following definition and the standard definition of NP will be discussed in Sections 2.1.3–2.1.4 and 2.2.3.)

Definition 2.3 (search problems with efficiently checkable solutions):

- *The search problem of a polynomially bounded relation $R \subseteq \{0,1\}^* \times \{0,1\}^*$ has* efficiently checkable solutions *if there exists a polynomial time algorithm A such that, for every x and y, it holds that $A(x, y) = 1$ if and only if $(x, y) \in R$.*

- *We denote by $\mathcal{PC}$* (standing for "Polynomial-time Check") *the class of search problems that correspond to polynomially-bounded binary relations that have efficiently checkable solutions. That is, $R \in \mathcal{PC}$ if the following two conditions hold:*

1. *For some polynomial p, if $(x, y) \in R$ then $|y| \leq p(|x|)$.*
2. *There exists a polynomial-time algorithm that given (x, y) determines whether or not $(x, y) \in R$.*

The class $\mathcal{PC}$ contains thousands of natural problems (e.g., finding a traveling salesman tour of length that does not exceed a given threshold, finding the prime factorization of a given composite, etc). In each of these natural problems, the correctness of solutions can be checked efficiently (e.g., given a traveling salesman tour it is easy to compute its length and check whether or not it exceeds the given threshold).[2]

The class $\mathcal{PC}$ is the natural domain for the study of which problems are in $\mathcal{PF}$, because the ability to efficiently recognize a valid solution is a *natural* prerequisite for a discussion regarding the complexity of finding such solutions. We warn, however, that $\mathcal{PF}$ contains (unnatural) problems that are not in $\mathcal{PC}$ (see Exercise 2.1).

2.1.1.3 The P versus NP question in terms of search problems

Is it the case that every search problem in $\mathcal{PC}$ is in $\mathcal{PF}$? That is, if one can efficiently check the correctness of solutions with respect to some (polynomially-bounded) relation R, then is it necessarily the case that the search problem of R can be solved efficiently? In other words, if it is *easy to check* whether or not a given solution for a given instance is correct, then is it also *easy to find* a solution to a given instance?

If $\mathcal{PC} \subseteq \mathcal{PF}$ then this would mean that whenever solutions to given instances can be efficiently checked (for correctness) it is also the case that such solutions can be efficiently found (when given only the instance). This would mean that all reasonable search problems (i.e., all problems in $\mathcal{PC}$) are easy to solve. Needless to say, such a situation would contradict the intuitive feeling (and the daily experience) that some reasonable search problems are hard to solve. Furthermore, in such a case, the notion of "solving a problem" will lose its meaning (because finding a solution will not be significantly more difficult than checking its validity).

On the other hand, if $\mathcal{PC} \setminus \mathcal{PF} \neq \emptyset$ then there exist reasonable search problems (i.e., some problems in $\mathcal{PC}$) that are hard to solve. This conforms with our basic intuition by which some reasonable problems are easy to solve whereas others are hard to solve. Furthermore, it reconfirms the intuitive gap between the notions of solving and checking (asserting that in some cases "solving" is significantly harder than "checking").

2.1.2 The decision version: proving versus verifying

As we shall see in the sequel, the study of search problems (e.g., the $\mathcal{PC}$-vs-$\mathcal{PF}$ Question) can be "reduced" to the study of decision problems. Since the latter

[2]In the traveling salesman problem (TSP), the instance is a matrix of distances between cities and a threshold, and the task is to find a tour that passes all cities and covers a total distance that does not exceed the threshold.

problems have a less cumbersome terminology, complexity theory tends to focus on them (and maintains its relevance to the study of search problems via the aforementioned reduction). Thus, the study of decision problems provides a convenient way for studying search problems. For example, the study of the complexity of deciding the satisfiability of Boolean formulae provides a convenient way for studying the complexity of finding satisfying assignments for such formulae.

We wish to stress, however, that decision problems are interesting and natural *per se* (i.e., beyond their role in the study of search problems). After all, some people do care about the truth, and so determining whether certain claims are true is a natural computational problem. Specifically, determining whether a given object (e.g., a Boolean formula) has some predetermined property (e.g., is satisfiable) constitutes an appealing computational problem. The P-vs-NP Question refers to the complexity of solving such problems for a wide and natural class of properties associated with the class NP. The latter class refers to properties that have "efficient proof systems" allowing for the verification of the claim that a given object has a predetermined property (i.e., is a member of a predetermined set). Jumping ahead, we mention that the P-vs-NP Question refers to the question of whether properties that have efficient proof systems can also be decided efficiently (without proofs). Let us clarify all these notions.

Properties of objects are modeled as subsets of the set of all possible objects (i.e., a property is associated with the set of objects having this property). For example, the property of being a prime is associated with the set of prime numbers and the property of being connected (resp., having a Hamiltonian path) is associated with the set of connected (resp., Hamiltonian) graphs. Thus, we focus on deciding membership in sets (as in Definition 1.2). The standard formulation of the P-vs-NP Question refers to the equality of two natural classes of decision problems, denoted P and NP (and defined in §2.1.2.1 and §2.1.2.2, respectively).

2.1.2.1 The class P as a natural class of decision problems

Needless to say, we are interested in the class of decision problems that are efficiently solvable. This class is traditionally denoted $\mathcal{P}$ (standing for Polynomial-time). The following definition refers to the formulation of solving decision problems (provided in Definition 1.2).

Definition 2.4 (efficiently solvable decision problems):

- *A decision problem $S \subseteq \{0,1\}^*$ is* efficiently solvable *if there exists a polynomial time algorithm A such that, for every x, it holds that $A(x) = 1$ if and only if $x \in S$.*
- *We denote by $\mathcal{P}$ the class of decision problems that are efficiently solvable.*

As in Definition 2.2, we have defined a fundamental class of problems, which contains many natural problems (e.g., determining whether or not a given graph is connected), but we do not have a good understanding regarding its actual contents (i.e., we are unable to characterize many natural problems with respect to membership in this class). In fact, there are many natural decision problems that are not known to reside in $\mathcal{P}$, and a natural class containing a host of such problems is presented next. This class of decision problems is denoted NP (for reasons that will become evident in Section 2.1.4).

2.1.2.2 The class NP and NP-proof systems

We view NP as the class of decision problems that have efficiently verifiable proof systems. Loosely speaking, we say that a set S has a proof system if instances in S have valid proofs of membership (i.e., proofs accepted as valid by the system), whereas instances not in S have no valid proofs. Indeed, proofs are defined as strings that (when accompanying the instance) are accepted by the (efficient) verification procedure. We say that V is a verification procedure for membership in S if it satisfies the following two conditions:

1. Completeness: True assertions have valid proofs; that is, proofs accepted as valid by V. Bearing in mind that assertions refer to membership in S, this means that for every $x \in S$ there exists a string y such that $V(x,y) = 1$ (i.e., V accepts y as a valid proof for the membership of x in S).

2. Soundness: False assertions have no valid proofs. That is, for every $x \notin S$ and every string y it holds that $V(x,y) = 0$, which means that V rejects y as a proof for the membership of x in S.

We note that the soundness condition captures the "security" of the verification procedure; that is, its ability not to be fooled by anything into proclaiming a wrong assertion. The completeness condition captures the "viability" of the verification procedure; that is, its ability to be convinced of any valid assertion, when presented with an adequate proof. (We stress that, in general, proof systems are defined in terms of their verification procedures, which must satisfy adequate completeness and soundness conditions.) Our focus here is on efficient verification procedures that utilize relatively short proofs (i.e., proofs that are of length that is polynomially bounded by the length of the corresponding assertion).[3]

Let us consider a couple of examples before turning to the actual definition. For example, the set of Hamiltonian graphs has a verification procedure that, given a pair (G, P), accepts if and only if P is a Hamiltonian path in the graph G. In this case P serves as a proof that G is Hamiltonian. Note that such proofs are relatively short (i.e., the path is actually shorter than the description of the graph) and are easy to verify. Needless to say, this proof system satisfies the

[3]**Advanced comment:** In continuation to Footnote 1, we note that in this chapter we consider *deterministic* (polynomial-time) verification procedures, and consequently the completeness and soundness conditions that we state here are error-less. In contrast, in Chapter 9, we will consider various types of probabilistic (polynomial-time) verification procedures as well as probabilistic completeness and soundness conditions. A common theme that underlies both treatments is that efficient verification is interpreted as meaning verification by a process that runs in time that is polynomial in the length of the assertion. In the current chapter, we use the equivalent formulation that considers the running time as a function of the total length of the assertion and the proof, but require that the latter has length that is polynomially bounded by the length of the assertion.

aforementioned completeness and soundness conditions. In the case of satisfiable Boolean formulae, given a formula ϕ and a truth assignment τ, the verification procedure instantiates ϕ (according to τ), and accepts if and only if simplifying the resulting Boolean expression yields the value `true`. In this case τ serves as a proof that ϕ is satisfiable, and the alleged proofs are indeed relatively short and easy to verify.

Definition 2.5 (efficiently verifiable proof systems):

- *A decision problem $S \subseteq \{0,1\}^*$ has an* efficiently verifiable proof system *if there exists a polynomial p and a polynomial-time* (verification) *algorithm V such that the following two conditions hold:*

  1. Completeness: *For every $x \in S$, there exists y of length at most $p(|x|)$ such that $V(x,y) = 1$.*

     (Such a string y is called an NP-witness for $x \in S$.)

  2. Soundness: *For every $x \notin S$ and every y, it holds that $V(x,y) = 0$.*

  Thus, $x \in S$ if and only if there exists y of length at most $p(|x|)$ such that $V(x,y) = 1$.

  In such a case, we say that S has an NP-proof system*, and refer to V as its* verification procedure (or as the proof system itself).

- *We denote by $\mathcal{NP}$ the class of decision problems that have efficiently verifiable proof systems.*

We note that the term *NP-witness* is commonly used, although in most cases V is not called a proof system (nor a verification procedure of such a system). In some cases, V (or the set of pairs accepted by V) is called a witness relation of S. We stress that the same set S may have many different NP-proof systems (see Exercise 2.2), and that in some cases the difference is not artificial (see Exercise 2.3).

Teaching note: Using Definition 2.5, it is typically easy to show that natural decision problems are in $\mathcal{NP}$. All that is needed is designing adequate NP-proofs of membership, which is typically quite straightforward and natural, because natural decision problems are typically phrased as asking about the existence of a structure (or object) that can be easily verified as valid. For example, `SAT` is defined as the set of satisfiable Boolean formulae, which means asking about the existence of satisfying assignments. Indeed, we can efficiently check whether a given assignment satisfies a given formula, which means that we have (a verification procedure for) an NP-proof system for `SAT`.

Note that for any search problem R in $\mathcal{PC}$, the set of instances that have a solution with respect to R (i.e., the set $S_R \stackrel{\rm def}{=} \{x : R(x) \neq \emptyset\}$) is in $\mathcal{NP}$. Specifically, for any $R \in \mathcal{PC}$, consider the verification procedure V such that $V(x,y) \stackrel{\rm def}{=} 1$ if and only if $(x,y) \in R$, and note that the latter condition can be decided in poly$(|x|)$-time. Thus, *any search problem in $\mathcal{PC}$ can be viewed as a problem of searching* *for* (efficiently verifiable) *proofs* (i.e., NP-witnesses for membership in the set of instances having solutions). Furthermore, any NP-proof system gives rise to a natural search problem in $\mathcal{PC}$; that is, the problem of searching for a valid proof (i.e., an NP-witness) for the given instance (i.e, the verification procedure V yields the search problem that corresponds to $R = \{(x,y) : V(x,y)=1\}$).

Teaching note: The last paragraph suggests another easy way of showing that natural decision problems are in $\mathcal{NP}$: just thinking of the corresponding natural search problem. The point is that natural decision problems (in $\mathcal{NP}$) are phrased as referring to whether a solution exists (for the corresponding natural search problem). For example, in the case of `SAT`, the question is whether there exists a satisfying assignment to given Boolean formula, and the corresponding search problem is finding such an assignment. But in all these cases, it is easy to check the correctness of solutions; that is, the corresponding search problem is in $\mathcal{PC}$, which implies that the decision problem is in $\mathcal{NP}$.

Observe that $\mathcal{P} \subseteq \mathcal{NP}$ holds: A verification procedure for claims of membership in a set $S \in \mathcal{P}$ may just ignore the alleged NP-witness and run the decision procedure that is guaranteed by the hypothesis $S \in \mathcal{P}$; that is, $V(x,y) = A(x)$, where A is the aforementioned decision procedure. Indeed, the latter verification procedure is quite an abuse of the term (because it makes no use of the proof); however, it is a legitimate one. As we shall shortly see, the P-vs-NP Question refers to the question of whether such proof-oblivious verification procedures can be used for every set that has some efficiently verifiable proof system. (Indeed, given that $\mathcal{P} \subseteq \mathcal{NP}$, the P-vs-NP Question is whether $\mathcal{NP} \subseteq \mathcal{P}$.)

2.1.2.3 The P versus NP question in terms of decision problems

Is it the case that NP-proofs are useless? That is, is it the case that for every efficiently verifiable proof system one can easily determine the validity of assertions without looking at the proof? If that were the case, then proofs would be meaningless, because they would have no fundamental advantage over directly determining the validity of the assertion. The conjecture $\mathcal{P} \neq \mathcal{NP}$ asserts that proofs are useful: there exists sets in $\mathcal{NP}$ that cannot be decided by a polynomial-time algorithm, and so for these sets obtaining a proof of membership (for some instances) is useful (because we cannot efficiently determine membership by ourselves).

In the foregoing paragraph we viewed $\mathcal{P} \neq \mathcal{NP}$ as asserting the advantage of obtaining proofs over deciding the truth by ourselves. That is, $\mathcal{P} \neq \mathcal{NP}$ asserts that (in some cases) verifying is easier than deciding. A slightly different perspective is that $\mathcal{P} \neq \mathcal{NP}$ asserts that finding proofs is harder than verifying their validity. This is the case because, for any set S that has an NP-proof system, the ability to efficiently find proofs of membership with respect to this system (i.e., finding an NP-witness of membership in S for any given $x \in S$), yields the ability to decide membership in S. Thus, for $S \in \mathcal{NP} \setminus \mathcal{P}$, it must be harder to find proofs of membership in S than to verify the validity of such proofs (which can be done in polynomial-time).

2.1.3 Equivalence of the two formulations

As hinted several times, *the two formulations of the P-vs-NP Questions are equivalent.* That is, every search problem having efficiently checkable solutions is solvable in polynomial time (i.e., $\mathcal{PC} \subseteq \mathcal{PF}$) if and only if membership in any set that has an NP-proof system can be decided in polynomial time (i.e., $\mathcal{NP} \subseteq \mathcal{P}$). Recalling that $\mathcal{P} \subseteq \mathcal{NP}$ (whereas $\mathcal{PF}$ is not contained in $\mathcal{PC}$ (Exercise 2.1)), we prove the following.

Theorem 2.6 *$\mathcal{PC} \subseteq \mathcal{PF}$ if and only if $\mathcal{P} = \mathcal{NP}$.*

Proof: Suppose, on the one hand, that the inclusion holds for the search version (i.e., $\mathcal{PC} \subseteq \mathcal{PF}$). We will show that this implies the existence of an efficient algorithm for finding NP-witnesses for any set in $\mathcal{NP}$, which in turn implies that this set is in $\mathcal{P}$. Specifically, let S be an arbitrary set in $\mathcal{NP}$, and V be the corresponding verification procedure (i.e., satisfying the conditions in Definition 2.5). Then $R \stackrel{\rm def}{=} \{(x,y) : V(x,y) = 1\}$ is a polynomially bounded relation in $\mathcal{PC}$, and by the hypothesis its search problem is solvable in polynomial time (i.e., $R \in \mathcal{PC} \subseteq \mathcal{PF}$). Denoting by A the polynomial-time algorithm solving the search problem of R, we decide membership in S in the obvious way. That is, on input x, we output 1 if and only if $A(x) \neq \perp$, where the latter event holds if and only if $A(x) \in R(x)$, which in turn occurs if and only if $R(x) \neq \emptyset$ (equiv., $x \in S$). Thus, $\mathcal{NP} \subseteq \mathcal{P}$ (and $\mathcal{NP} = \mathcal{P}$) follows.

Suppose, on the other hand, that $\mathcal{NP} = \mathcal{P}$. We will show that this implies an efficient algorithm for determining whether a given string y' is a prefix of some solution to a given instance x of a search problem in $\mathcal{PC}$, which in turn yields an efficient algorithm for finding solutions. Specifically, let R be an arbitrary search problem in $\mathcal{PC}$. Then the set $S'_R \stackrel{\rm def}{=} \{\langle x, y'\rangle : \exists y'' \text{ s.t. } (x, y'y'') \in R\}$ is in $\mathcal{NP}$ (because $R \in \mathcal{PC}$), and hence S'_R is in $\mathcal{P}$ (by the hypothesis $\mathcal{NP} = \mathcal{P}$). This yields a polynomial-time algorithm for solving the search problem of R, by extending a prefix of a potential solution bit-by-bit (while using the decision procedure to determine whether or not the current prefix is valid). That is, on input x, we first check whether or not $(x, \lambda) \in S'_R$ and output $\perp$ (indicating $R(x) = \emptyset$) in case $(x, \lambda) \notin S'_R$. Next, we proceed in iterations, maintaining the invariant that $(x, y') \in S'_R$. In each iteration, we set $y' \leftarrow y'0$ if $(x, y'0) \in S'_R$ and $y' \leftarrow y'1$ if $(x, y'1) \in S'_R$. If none of these conditions hold (which happens after at most polynomially many iterations) then the current y' satisfies $(x, y') \in R$. Thus, for an abritrary $R \in \mathcal{PC}$ we obtain that $R \in \mathcal{PF}$, and $\mathcal{PC} \subseteq \mathcal{PF}$ follows. ■

Reflection: The first part of the proof of Theorem 2.6 associates with each set S in $\mathcal{NP}$ a natural relation R (in $\mathcal{PC}$). Specifically, R consists of all pairs (x, y) such that y is an NP-witness for membership of x in S. Thus, the search problem of R consists of finding such an NP-witness, when given x as input. Indeed, R is called the witness relation of S, and solving the search problem of R allows to decide membership in S. Thus, $R \in \mathcal{PC} \subseteq \mathcal{PF}$ implies $S \in \mathcal{P}$. In the second part of the proof, we associate with each $R \in \mathcal{PC}$ a set S'_R (in $\mathcal{NP}$), but S'_R is more "expressive" than the set $S_R \overset{\text{def}}{=} \{x : \exists y \text{ s.t. } (x,y) \in R\}$ (which gives rise to R as its witness relation). Specifically, S'_R consists of strings that encode pairs (x, y') such that y' is a prefix of some string in $R(x) = \{y : (x,y) \in R\}$. The key observation is that deciding membership in S'_R allows to solve the search problem of R; that is, $S'_R \in \mathcal{P}$ implies $R \in \mathcal{PF}$.

Conclusion: Theorem 2.6 justifies the traditional focus on the decision version of the P-vs-NP Question. Indeed, given that both formulations of the question are equivalent, we may just study the less cumbersome one.

2.1.4 The traditional definition of NP

Unfortunately, Definition 2.5 is not the commonly used definition of $\mathcal{NP}$. Instead, traditionally, $\mathcal{NP}$ is defined as the class of sets that can be decided by a *fictitious* device called a non-deterministic polynomial-time machine (which explains the source of the notation NP). The reason that this class of fictitious devices is interesting is due to the fact that it captures (indirectly) the definition of NP-proofs. Since the reader may come across the traditional definition of $\mathcal{NP}$ when studying different works, the author feels obliged to provide the traditional definition as well as a proof of its equivalence to Definition 2.5.

Definition 2.7 (non-deterministic polynomial-time Turing machines):

- *A* non-deterministic Turing machine *is define as in §1.2.3.1, except that the transition function maps symbol-state pairs to subsets of triples* (rather than to a single triple) *in* $\Sigma \times Q \times \{-1, 0, +1\}$. *Accordingly, the configuration following a specific instantaneous configuration may be one of several possibilities, each determine by a different possible triple. Thus, the* computations of a non-deterministic machine *on a* (fixed) *given input may result in different outputs.*

  In the context of decision problems one typically considers the question of whether or not there exists a computation that starting with a fixed input halts with output 1. We say that the non-deterministic machine M accept x *if there exists a computation of* M*, on input* x*, that halts with output 1. The* set accepted by a non-deterministic machine *is the set of inputs that are accepted by the machine.*

- *A* non-deterministic polynomial-time Turing machine *is defined as one that makes a number of steps that is polynomial in the length of the input. Traditionally,* $\mathcal{NP}$ *is defined as the class of sets that are accepted by some non-deterministic polynomial-time Turing machine.*

We stress that Definition 2.7 refers to a fictitious model of computation. Specifically, Definition 2.7 makes no reference to the number (or fraction) of possible computations of the machine (on a specific input) that yield a specific output.[4] Definition 2.7 only refers to whether or not computations leading to a certain output exist (for a specific input). The question of what does the mere existence of such possible computations mean in terms of real-life is not addressed, because the model of a non-deterministic machine is not meant to provide a reasonable model of a real-life computer. The model is meant to capture something completely different (i.e., it is meant to provide an elegant definition of the class $\mathcal{NP}$, while relying on the fact that Definition 2.7 is equivalent to Definition 2.5).

Teaching note: Whether or not Definition 2.7 is elegant is a matter of taste. For sure, many students find Definition 2.7 quite confusing, possibly because they assume that it represents some natural model of computation and allow themselves to be fooled by their intuition regarding such models. (Needless to say, the students' intuition regarding computation is irrelevant when applied to a fictitious model.)

Note that, unlike other definitions in this chapter, Definition 2.7 makes explicit reference to a specific model of computation. Still, a similar extension can be applied to other models of computation by considering adequate non-deterministic computation rules. Also note that, without loss of generality, we may assume that the transition function maps each possible symbol-state pair to exactly two triples (cf. Exercise 2.4).

Theorem 2.8 *Definition 2.5 is equivalent to Definition 2.7. That is, a set S has an NP-proof system if and only if there exists a non-deterministic polynomial-time machine that accepts S.*

Proof Sketch: Suppose, on one hand, that the set S has an NP-proof system, and let us denote the corresponding verification procedure by V. Consider the following non-deterministic polynomial-time machine, denoted M. On input x, machine M makes an adequate $m = \text{poly}(|x|)$ number of non-deterministic steps, producing (non-deterministically) a string $y \in \{0,1\}^m$, and then emulates $V(x,y)$. We stress that these non-deterministic steps may result in producing any m-bit string y. Recall that $x \in S$ if and only if there exists y of length at most $\text{poly}(|x|)$ such that $V(x,y) = 1$. This implies that the set accepted by M equals S.

Suppose, on the other hand, that there exists a non-deterministic polynomial-time machine M that accepts the set S. Consider a deterministic machine M' that on input (x,y), where y has adequate length, emulates a computation of M on input x while using y to determine the non-deterministic steps of M. That is, the i^{th} step of M on input x is determined by the i^{th} bit of y (which indicates which of the two possible moves to make at the current step). Note that $x \in S$ if and only if there exists y of length at most $\text{poly}(|x|)$ such that $M'(x,y) = 1$. Thus, M' gives rise to an NP-proof system for S. $\square$

[4]**Advanced comment:** In contrast, the definition of a probabilistic machine refers to this number (or, equivalently, to the probability that the machine produces a specific output, when the probability is essentially taken uniformly over all possible computations). Thus, a probabilistic machine refers to a natural model of computation that can be realized provided we can equip the machine with a source of randomness. For details, see Section 6.1.

2.1.5 In support of P different from NP

Intuition and concepts constitute... the elements of all our knowledge, so that neither concepts without an intuition in some way corresponding to them, nor intuition without concepts, can yield knowledge.

Immanuel Kant (1724–1804)

Kant talks on the importance of *both* philosophical considerations (referred to as "concepts") and empirical considerations (referred to as "intuition") to science (referred to as (sound) "knowledge").

It is widely believed that P is different than NP; that is, that $\mathcal{PC}$ contains search problems that are not efficiently solvable, and that there are NP-proof systems for sets that cannot be decided efficiently. This belief is supported by both philosophical and empirical considerations.

- *Philosophical considerations*: Both formulations of the P-vs-NP Question refer to natural questions about which we have strong intuition. The notion of solving a (search) problem seems to presume that, at least in some cases (if not in general), finding a solution is significantly harder than checking whether a presented solution is correct. This translates to $\mathcal{PC} \setminus \mathcal{PF} \neq \emptyset$. Likewise, the notion of a proof seems to presume that, at least in some cases (if not in general), the proof is useful in determining the validity of the assertion; that is, that deciding the validity of an assertion may be made significantly easier when provided with a proof. This translates to $\mathcal{P} \neq \mathcal{NP}$, which also implies that it is significantly harder to find proofs than to verify their correctness, which again coincides with the daily experience of researchers and students.

- *Empirical considerations*: The class NP (or rather $\mathcal{PC}$) contains thousands of different problems for which no efficient solving procedure is known. Many of these problems have arisen in vastly different disciplines, and were the subject of extensive research of numerous different communities of scientists and engineers. These essentially independent studies have all failed to provide efficient algorithms for solving these problems, a failure that is extremely hard to attribute to sheer coincidence or a stroke of bad luck.

Throughout the rest of this book, we will adopt the common belief that P is different from NP. At some places, we will explicitly use this conjecture (or even stronger assumptions), whereas in other places we will present results that are interesting (if and) only if $\mathcal{P} \neq \mathcal{NP}$ (e.g., the entire theory of NP-completeness becomes uninteresting if $\mathcal{P} = \mathcal{NP}$).

The $\mathcal{P} \neq \mathcal{NP}$ conjecture is indeed very appealing and intuitive. The fact that this natural conjecture is unsettled seems to be one of the sources of frustration of complexity theory. The author's opinion, however, is that this feeling of frustration is not in place. In contrast, the fact that complexity theory evolves around natural questions that are so difficult to resolve makes its study very exciting.

2.1.6 Two technical comments regarding NP

Recall that when defining $\mathcal{PC}$ (resp., $\mathcal{NP}$) we have explicitly confined our attention to search problems of polynomially bounded relations (resp., NP-witnesses of polynomial length). An alternative formulation may allow a binary relation R to be in $\mathcal{PC}$ (resp., $S \in \mathcal{NP}$) if membership of (x, y) in R can be decided in time that is polynomial in the length of x (resp., the verification of a candidate NP-witness y for membership of x in S is required to be performed in poly($|x|$)-time). Indeed, this mean that the validity of y can be determined without reading all of it (which means that some substring of y can be used as the effective y in the original definitions).

We comment that problems in $\mathcal{PC}$ (resp., $\mathcal{NP}$) can be solved in exponential-time (i.e., time exp(poly($|x|$)) for input x). This can be done by an exhaustive search among all possible candidate solutions (resp., all possible candidate NP-witnesses). Thus, $\mathcal{NP} \subseteq \mathcal{EXP}$, where $\mathcal{EXP}$ denote the class of decision problems that can be solved in exponential-time (i.e., time exp(poly($|x|$)) for input x).

2.2 Polynomial-time Reductions

We present a general notion of (polynomial-time) reductions among computational problems, and view the notion of a "Karp-reduction" as an important special case that suffices (and is more convenient) in many cases. Reductions play a key role in the theory of NP-completeness, which is the topic of Section 2.3, but we stress the fundamental nature of the notion of a reduction and highlight two specific applications (i.e., reducing search and optimization problems to decision problems). Furthermore, in the latter applications, it will be important to use the general notion of a reduction (i.e., "Cook-reduction" rather than "Karp-reduction").

Teaching note: We assume that many students have heard of reductions, but we fear that most have obtained a conceptually poor view of their fundamental nature. In particular, we fear that reductions are identified with the theory of NP-completeness, while reductions have important applications that have little to do with NP-completeness (or completeness with respect to some other class). Furthermore, we believe that it is important to show that natural search and optimization problems can be reduced to decision problems.

2.2.1 The general notion of a reduction

Reductions are procedures that use "functionally specified" subroutines. That is, the functionality of the subroutine is specified, but its operation remains unspecified and its running-time is counted at unit cost. Analogously to algorithms, which are modeled by Turing machines, reductions can be modeled as *oracle* (Turing) machines. A reduction solves one computational problem (which may be either a search or a decision problem) by using oracle (or subroutine) calls to another computational problem (which again may be either a search or a decision problem).

The notion of a general algorithmic reduction was discussed in §1.2.3.2 and §1.2.3.5. These reductions, called Turing-reductions (cf. §1.2.3.2) and modeled by oracle machines (cf. §1.2.3.5), made no reference to the time complexity of the main algorithm (i.e., the oracle machine). Here, we focus on efficient (i.e., polynomial-time) reductions, which are often called *Cook reductions*. That is, we consider oracle machines (as in Definition 1.11) that run in time polynomial in the length of their input. We stress that the running time of an oracle machine is the number of steps made during its computation, and that the oracle's reply on each query is obtained in a single step.

The key property of efficient reductions is that they allow for the transformation of efficient implementations of the subroutine into efficient implementations of the task reduced to it. That is, as we shall see, if one problem is Cook-reducible to another problem and the latter is polynomial-time solvable then so is the former.

The most popular case is that of reducing decision problems to decision problems, but we will also consider reducing search problems to search problems and reducing search problems to decision problems. Note that when reducing to a decision problem, the oracle is determined as the single valid solver of the decision problem (i.e., the function $f : \{0,1\}^* \to \{0,1\}$ solves the decision problem of membership in S if, for every x, it holds that $f(x) = 1$ if $x \in S$ and $f(x) = 0$ otherwise). In contrast, when reducing to a search problem, there may be many different valid solvers (i.e., each function f that satisfies $(x, f(x)) \in R$ for every $(x,y) \in R$ is a valid solver of the search problem of R). We capture both cases in the following definition.

Definition 2.9 (Cook reduction): *A problem Π is* Cook-reducible *to a problem Π' if there exists a polynomial-time oracle machine M such that for every function f that solves Π' it holds that M^f solves Π, where $M^f(x)$ denotes the output of M on input x when given oracle access to f.*

Note that Π (resp., Π') may be either a search problem or a decision problem (or even a yet undefined type of a problem). At this point the reader should verify that if Π is Cook-reducible to Π' and Π' is solvable in polynomial-time then so is Π. (See Exercise 2.5 for other properties of Cook-reductions.) Also observe that the second part of the proof of Theorem 2.6 is actually a Cook-reduction of the search problem of any R in $\mathcal{PC}$ to a decision problem regarding a related set $S'_R = \{(x,y') : \exists y'' \text{ s.t. } (x,y'y'') \in R\}$, which in $\mathcal{NP}$. Thus, that proof establishes that *any search problem in $\mathcal{PC}$ is Cook-reducible to some decision problem in $\mathcal{NP}$*. We shall see a tighter relation between search and decision problems in Section 2.2.3 (i.e., in some cases, R will be reduced to $S_R = \{x : \exists y \text{ s.t. } (x,y) \in R\}$ rather than to S'_R).

A Karp-reduction is a special case of a reduction (from a decision problem to a decision problem). Specifically, for decision problems S and S', we say that S is Karp-reducible to S' if there is a reduction of S to S' *that operates as follows*: On input x (an instance for S), the reduction computes x', makes query x' to the oracle S' (i.e., invokes the subroutine for S' on input x'), and answers whatever the latter returns. This reduction is often represented by the polynomial-time computable mapping of x to x'; that is, the standard definition of a Karp-reduction is actually as follows.

Definition 2.10 (Karp reduction): *A polynomial-time computable function f is called a* Karp-reduction of S to S' *if, for every x, it holds that $x \in S$ if and only if $f(x) \in S'$.*

Thus, syntactically speaking, a Karp-reduction is not a Cook-reduction, but it trivially gives rise to one (i.e., on input x, the oracle machine makes query $f(x)$, and returns the oracle answer). Being slightly inaccurate but essentially correct, we shall say that Karp-reductions are special cases of Cook-reductions. Needless to say, Karp-reductions constitute a very restricted case of Cook-reductions. Still, this restricted case suffices for many applications (e.g., most importantly for the theory of NP-completeness (when developed for decision problems)), but not for reducing a search problem to a decision problem. Furthermore, whereas each decision problem is Cook-reducible to its complement, some decision problems are *not* Karp-reducible to their complement (see Exercises 2.7 and 2.33).

We comment that Karp-reductions may (and should) be augmented in order to handle reductions of search problems to search problems. Such an augmented Karp-reduction of the search problem of R to the search problem of R' operates as follows: On input x (an instance for R), the reduction computes x', makes query x' to the oracle R' (i.e., invokes the subroutine for searching R' on input x') obtaining y' such that $(x',y') \in R'$, and uses y' to compute a solution y to x (i.e., $y \in R(x)$). Thus, such a reduction can be represented by two polynomial-time computable mappings, f and g, such that $(x, g(x,y')) \in R$ for any y' that solves $f(x)$ (i.e., for y' that satisfies $(f(x),y') \in R'$). (Indeed, in general, unlike in the case of decision problems, the reduction cannot just return y' as an answer to x.) This augmentation is called a Levin-reduction and, analogously to the case of a Karp-reduction, is often represented by the two polynomial-time computable mappings (i.e., of x to x', and of (x,y') to y).

Definition 2.11 (Levin reduction): *A pair of polynomial-time computable functions, f and g, is called a* Levin-reduction of R to R' *if f is a Karp reduction of $S_R = \{x : \exists y \text{ s.t. } (x,y) \in R\}$ to $S_{R'} = \{x' : \exists y' \text{ s.t. } (x',y') \in R'\}$ and for every $x \in S_R$ and $y' \in R'(f(x))$ it holds that $(x, g(x,y')) \in R$, where $R'(x') = \{y' : (x',y') \in R'\}$.*

Indeed, the function f preserves the existence of solutions; that is, for any x, it holds that $R(x) \neq \emptyset$ if and only if $R'(f(x)) \neq \emptyset$. As for the second function (i.e., g), it maps any solution y' for the reduced instance $f(x)$ to a solution for the original instance x (where this mapping may also depend on x). It is natural to consider also a third function, which maps solutions for R to solutions for R' (see Exercise 2.28).

Terminology: In the sequel, whenever we neglect to mention the type of a reduction, we refer to a Cook-reduction. Two additional terms, which will be particularly useful in the advanced chapters, are presented next.

- We say that two problems are computationally equivalent if they are reducible to one another. This means that the two problems are essentially as hard (or as easy). Note that, for various complexity classes (e.g., $\mathcal{NP}$ and $\mathcal{PC}$), computationally equivalent problems need not reside in the same class, since the reductions allowed here are Cook-reductions. For example, as we shall see in Section 2.2.3, there exist many natural $R \in \mathcal{PC}$ such that the search problem of R and the decision problem of $S_R = \{x : \exists y \text{ s.t. } (x,y) \in R\}$ are computationally equivalent.
- We say that a *class of problems, $\mathcal{C}$, is reducible to a problem* Π' *if every problem in $\mathcal{C}$, is reducible to* Π'. We say that the class $\mathcal{C}$ is reducible to the class $\mathcal{C}'$ if for every $\Pi \in \mathcal{C}$ there exists $\Pi' \in \mathcal{C}'$ such that Π is reducible to Π'. For example, recall that $\mathcal{PC}$ is reducible to $\mathcal{NP}$.

The fact that we allow Cook-reductions is essential to various important connections between decision problems and other computational problems. Specifically, as shown in Section 2.2.2, a natural class of optimization problems is reducible to $\mathcal{NP}$. Recall that $\mathcal{PC}$ is reducible to $\mathcal{NP}$ (as shown implicitly in the proof of Theorem 2.6). Furthermore, as shown in Section 2.2.3, many natural search problems in $\mathcal{PC}$ are reducible to a corresponding natural decision problem in $\mathcal{NP}$ (rather than merely to some problem in $\mathcal{NP}$).

2.2.2 Reducing optimization problems to search problems

Many search problems refer to a set of potential solutions, per each problem instance, such that different solutions are assigned different "values" (resp., "costs"). In such a case, one may be interested in finding a solution that has value exceeding some threshold (resp., cost below some threshold), or (even better) finding a solution of maximum value (resp., minimum cost). For simplicity, let us focus on the case of a value that we wish to maximize. Still, there are two different objectives (i.e., exceeding a threshold and optimizing), giving rise to two different (auxiliary) search problems related to the same relation R. Specifically, for a binary relation R and a *value function* $f : \{0,1\}^* \times \{0,1\}^* \to \mathbb{R}$, we consider two search problems.

1. *Exceeding a threshold*: Given a pair (x,v) the task is to find $y \in R(x)$ such that $f(x,y) \geq v$, where $R(x) = \{y : (x,y) \in R\}$. That is, we are actually referring to the search problem of the relation
$$R_f \stackrel{\text{def}}{=} \{(\langle x,v\rangle, y) : (x,y) \in R \wedge f(x,y) \geq v\}, \tag{2.1}$$
where $\langle x,v\rangle$ denotes a string that encodes the pair (x,v).
2. *Maximization*: Given x the task is to find $y \in R(x)$ such that $f(x,y) = v_x$, where v_x is the maximum value of $f(x,y')$ over all $y' \in R(x)$. That is, we are actually referring to the search problem of the relation
$$R'_f \stackrel{\text{def}}{=} \{(x,y) \in R : f(x,y) = \max_{y' \in R(x)} \{f(x,y')\}\}. \tag{2.2}$$

Examples of value functions include the size of a clique in a graph, the amount of flow in a network (with link capacities), etc. The task may be to find a clique of size exceeding a given threshold in a given graph or to find a maximum-size clique in a given graph. Note that, in these examples, the "base" search problem (i.e., the relation R) is quite easy to solve, and the difficulty arises from the auxiliary condition on the value of a solution (presented in R_f and R'_f). Indeed, one may trivialize R (i.e., let $R(x) = \{0,1\}^{\text{poly}(|x|)}$ for every x), and impose all necessary structure by the function f (see Exercise 2.8).

We confine ourselves to the case that f is polynomial-time computable, which in particular means that $f(x,y)$ can be represented by a rational number of length polynomial in $|x|+|y|$. We will show next that, in this case, the two aforementioned search problems (i.e., of R_f and R'_f) are computationally equivalent.

Theorem 2.12 *For any polynomial-time computable $f:\{0,1\}^*\times\{0,1\}^*\to\mathbb{R}$ and a polynomially bounded binary relation R, let R_f and R'_f be as in Eq.* (2.1) *and Eq.* (2.2)*, respectively. Then the search problems of R_f and R'_f are computationally equivalent.*

It follows that, *for $R \in \mathcal{PC}$ and polynomial-time computable f, both the R_f and R'_f are reducible to $\mathcal{NP}$.* We note, however, that, while $R_f \in \mathcal{PC}$ always holds, it is not necessarily the case that $R'_f \in \mathcal{PC}$. See further discussion following the proof.

Proof: The search problem of R_f is reduced to the search problem of R'_f by finding an optimal solution (for the given instance) and comparing its value to the given threshold value. That is, we construct an oracle machine that solves R_f by making a single query to R'_f. Specifically, on input (x,v), the machine issues the query x (to a solver for R'_f), obtaining the optimal solution y (or an indication $\perp$ that $R(x) = \emptyset$), computes $f(x,y)$, and returns y if $f(x,y) \geq v$. Otherwise (i.e., either $y = \perp$ or $f(x,y) < v$), the machine returns an indication that $R_f(x,v) = \emptyset$.

Turning to the opposite direction, we reduce the search problem of R_f to the search problem of R'_f by first finding the optimal value $v_x = \max_{y\in R(x)}\{f(x,y)\}$ (by binary search on its possible values), and next finding a solution of value v_x. In both steps, we use oracle calls to R_f. For simplicity, we assume that f assigns positive integer values, and let $\ell = \text{poly}(|x|)$ be such that $f(x,y) \leq 2^\ell - 1$ for every $y \in R(x)$. Then, on input x, we first find $v_x = \max\{f(x,y) : y \in R(x)\}$, by making oracle calls of the form $\langle x,v\rangle$. The point is that $v_x < v$ if any only if $R_f(\langle x,v\rangle) = \emptyset$, which in turn is indicated by the oracle answer $\perp$ (to the query $\langle x,v\rangle$). Making ℓ queries, we determine v_x (see Exercise 2.9). Note that in case $R(x) = \emptyset$, all answers will indicate that $R_f(\langle x,v\rangle) = \emptyset$, which we treat as if $v_x = 0$. Finally, we make the query (x,v_x), and halt returning the oracle's answer (which is $y \in R(x)$ such that $f(x,y) = v_x$ if $v_x > 0$ and an indication that $R(x) = \emptyset$ otherwise). ■

Proof's digest. Note that the first direction uses the hypothesis that f is polynomial-time computable, whereas the opposite direction only used the fact that the optimal value lies in a finite space of exponential size that can be "efficiently searched". Whereas the first direction can be proved using a Levin-reduction, this seems impossible for the opposite direction (in general).

On the complexity of R_f and R'_f. We focus on the natural case in which $R \in \mathcal{PC}$ and f is polynomial-time computable. In this case, Theorem 2.12 implies that R_f and R'_f are computationally equivalent. A closer look reveals, however, that $R_f \in \mathcal{PC}$ always holds, whereas $R'_f \in \mathcal{PC}$ does *not* necessarily hold. That is, the problem of finding a solution (for a given instance) that exceeds a given threshold is in the class $\mathcal{PC}$, whereas the problem of finding an optimal solution is not necessarily in the class $\mathcal{PC}$. For example, the problem of finding a clique of a given size K in a given graph G is in $\mathcal{PC}$, whereas the problem of finding a maximum size clique in a given graph G is not known (and is quite unlikely) to be in $\mathcal{PC}$ (although it is Cook-reducible to $\mathcal{PC}$). Indeed, the class of problems that are reducible to $\mathcal{PC}$ is a natural and interesting class (see the ending paragraph of Section 3.2.1); indeed, for every $R \in \mathcal{PC}$ and polynomial-time computable f, the former class contains R'_f.

2.2.3 Self-reducibility of search problems

The results presented in this section further justify the focus on decision problems. Loosely speaking, these results show that for many natural relations R, the question of whether or not the search problem of R is efficiently solvable (i.e., is in $\mathcal{PF}$) is determined by the question of whether or not the "decision problem implicit in R" (i.e., $S_R = \{x : \exists y \text{ s.t. } (x,y) \in R\}$) is efficiently solvable (i.e., is in $\mathcal{P}$). Note that the latter decision problem is easily reducible to the search problem of R, and so our focus is on the other direction. That is, we are interested in relations R for which the search problem of R is reducible to the decision problem of S_R.

Teaching note: Our usage of the term self-reducibility differs from the traditional one. Traditionally, a decision problem is called (downwards) self-reducible if it is Cook-reducible to itself via a reduction that on input x only makes queries that are smaller than x (according to some appropriate measure on the size of strings). Under some natural restrictions (i.e., the reduction takes the disjunction of the oracle answers) such reductions yield reductions of search to decision (as discussed in the main text). For further details, see Exercise 2.13.

Definition 2.13 (the decision implicit in a search and self-reducibility): *The* decision problem implicit the search problem of R *is deciding membership in the set* $S_R = \{x : R(x) \neq \emptyset\}$, *where* $R(x) = \{y : (x,y) \in R\}$. *The search problem of R is called* self-reducible *if it can be reduced to the decision problem of* S_R.

Note that the search problem of R and the problem of deciding membership in S_R refer to the same instances: The search problem requires finding an adequate solution (i.e., given x find $y \in R(x)$), whereas the decision problem refers to the question of whether such solutions exist (i.e., given x determine whether or not $R(x)$ is non-empty). Thus, S_R is really the "decision problem implicit in R," because it is a decision problem that one implicitly solves when solving the search problem of R. Indeed, for any R, *the decision problem of S_R is easily reducible to* *the search problem for R* (and if R is in $\mathcal{PC}$ then S_R is in $\mathcal{NP}$).[5] It follows that *if a search problem R is self-reducible then it is computationally equivalent to the decision problem S_R.*

Note that the general notion of a reduction (i.e., Cook-reduction) seems inherent to the notion of self-reducibility. This is the case not only due to syntactic considerations, but rather due to the following inherent reason. An oracle to any decision problem returns a single bit per invocation, while the intractability of a search problem in $\mathcal{PC}$ must be due to lacking more than a single bit (see Exercise 2.10).

We shall see that self-reducibility is a property of many natural search problems (including all NP-complete search problems). This justifies the relevance of decision problems to search problems in a stronger sense than established in Section 2.1.3: Recall that in Section 2.1.3 we showed that the fate of the search problem class $\mathcal{PC}$ (w.r.t $\mathcal{PF}$) is determined by the fate of the decision problem class $\mathcal{NP}$ (w.r.t $\mathcal{P}$). Here we show that, for many natural search problems in $\mathcal{PC}$ (i.e., self-reducible ones), the fate of such a problem R (w.r.t $\mathcal{PF}$) is determined by the fate of the decision problem S_R (w.r.t $\mathcal{P}$), where S_R is the decision problem implicit in R.

We now present a few search problems that are self-reducible. We start with SAT (see Section G.2), the set of satisfiable Boolean formulae (in CNF), and consider the search problem in which given a formula one should provide a truth assignment that satisfies it. The corresponding relation is denoted $R_{\mathtt{SAT}}$; that is, $(\phi,\tau) \in R_{\mathtt{SAT}}$ if τ is a satisfying assignment to the formulae ϕ. The decision problem implicit in $R_{\mathtt{SAT}}$ is indeed SAT. Note that $R_{\mathtt{SAT}}$ is in $\mathcal{PC}$ (i.e., it is polynomially-bounded and membership of (ϕ,τ) in $R_{\mathtt{SAT}}$ is easy to decide (by evaluating a Boolean expression)).

Proposition 2.14 ($R_{\mathtt{SAT}}$ is self-reducible): *The search problem of $R_{\mathtt{SAT}}$ is reducible to* SAT.

Thus, the search problem of $R_{\mathtt{SAT}}$ is computationally equivalent to deciding membership in SAT. Hence, in studying the complexity of SAT, we also address the complexity of the search problem of $R_{\mathtt{SAT}}$.

Proof: We present an oracle machine that solves the search problem of $R_{\mathtt{SAT}}$ by making oracle calls to SAT. Given a formula ϕ, we find a satisfying assignment to ϕ (in case such an assignment exists) as follows. First, we query SAT on ϕ itself, and return an indication that there is no solution if the oracle answer is 0 (indicating $\phi \notin$ SAT). Otherwise, we let τ, initiated to the empty string, denote a prefix of a satisfying assignment of ϕ. We proceed in iterations, where in each iteration we extend τ by one bit. This is done as follows: First we derive a formula, denoted ϕ', by setting the first $|\tau|+1$ variables of ϕ according to the values $\tau 0$. We then query SAT on ϕ' (which means that we ask whether or not $\tau 0$ is a prefix of a satisfying assignment of ϕ). If the answer is positive then we set $\tau \leftarrow \tau 0$ else we set $\tau \leftarrow \tau 1$. This procedure relies on the fact that if τ is a prefix of a satisfying assignment of

[5]For example, the reduction invokes the search oracle and answer 1 if and only if the oracle returns some string (rather than the "no solution" symbol).

ϕ and $\tau 0$ is not a prefix of a satisfying assignment of ϕ then $\tau 1$ must be a prefix of a satisfying assignment of ϕ.

We wish to highlight a key point that has been blurred in the foregoing description. Recall that the formula ϕ' is obtained by replacing some variables by constants, which means that ϕ' *per se* contains Boolean variables as well as Boolean constants. However, the standard definition of SAT disallows Boolean constants in its instances.[6] Nevertheless, ϕ' can be simplified such that the resulting formula contains no Boolean constants. This simplification is performed according to the straightforward Boolean rules: That is, the constant `false` can be omitted from any clause, but if a clause contains only occurrences of the constant `false` then the entire formula simplifies to `false`. Likewise, if the constant `true` appears in a clause then the entire clause can be omitted, but if all clauses are omitted then the entire formula simplifies to `true`. Needless to say, if the simplification process yields a Boolean constant then we may skip the query, and otherwise we just use the simplified form of ϕ' as our query. ■

Reductions analogous to the one used in the proof of Proposition 2.14 can be presented also for other search problems (and not only for NP-complete ones). Two such examples are searching for a 3-coloring of a given graph and searching for an isomorphism between a given pair of graphs (where the first problem is known to be NP-complete and the second problem is believed not to be NP-complete). In both cases, the reduction of the search problem to a decision problem involves extending a prefix of a valid solution by making suitable queries in order to decide which extension to use. Note, however, that in these cases the process of getting rid of constants (representing partial solutions) is more involved. For example, in the case of Graph 3-Colorability (resp., Graph Isomorphism) we need to enforce a partial coloring of a given graph (resp., a partial isomorphism between a given pair of graphs); see Exercises 2.11 and 2.12, respectively.

Reflection: The proof of Proposition 2.14 (as well as the proofs of similar results) consists of two observations.

1. For every relation R in $\mathcal{PC}$, it holds that the search problem of R is reducible to the decision problem of $S'_R = \{(x, y') : \exists y'' \text{ s.t. } (x, y'y'') \in R\}$. Such a reduction is explicit in the proof of Theorem 2.6 and is implicit in the proof of Proposition 2.14.

2. For specific $R \in \mathcal{PC}$ (e.g., S_{SAT}), deciding membership in S'_R is reducible to deciding membership in $S_R = \{x : \exists y \text{ s.t. } (x, y) \in R\}$. This is where the specific structure of SAT was used, allowing for a direct and natural transformation of instances of S'_R to instances of S_R.

   (We comment that if S_R is NP-complete then S'_R, which is always in $\mathcal{NP}$, is reducible to S_R by the mere fact that S_R is NP-complete; this comment is related to the following advanced comment.)

[6]While the problem seems rather technical at the current setting (as it merely amounts to whether or not the definition of SAT allows Boolean constants in its instances), it is far from being so technical in other cases (see Exercises 2.11 and 2.12).

For an arbitrary $R \in \mathcal{PC}$, deciding membership in S'_R is not necessarily reducible to deciding membership in S_R. Furthermore, deciding membership in S'_R is not necessarily reducible to the search problem of R. (See Exercises 2.14, 2.15, and 2.16.)

Teaching note: In the rest of this section, we assume that the students have heard of NP-completeness. Actually, we only need the students to know the definition of NP-completeness (i.e., a set S is $\mathcal{NP}$-complete if $S \in \mathcal{NP}$ and every set in $\mathcal{NP}$ is reducible to S). Yet, the teacher may prefer postponing the presentation of the following material to Section 2.3.1 (or even to a later stage).

Advanced comment: In general, self-reducibility is a property of the search problem and not of the decision problem implicit in it. Under plausible assumptions (e.g., the intractability of factoring), there exists relations $R_1, R_2 \in \mathcal{PC}$ having the same implicit-decision problem (i.e., $\{x : R_1(x) \neq \emptyset\} = \{x : R_2(x) \neq \emptyset\}$) such that R_1 is self-reducible but R_2 is not (see Exercise 2.17). However, this phenomenon does not arise when NP-complete problems are involved; that is, *all search problems that refer to finding NP-witnesses for any NP-complete decision problem are self-reducible.*

Theorem 2.15 *For every R in $\mathcal{PC}$ such that S_R is $\mathcal{NP}$-complete, the search problem of R is reducible to deciding membership in S_R.*

In many cases, as in the proof of Proposition 2.14, the reduction of the search problem to the corresponding decision problem is quite natural. The following proof presents a generic reduction (which may be "unnatural" in some cases).

Proof: In order to reduce the search problem of R to deciding S_R, we compose the following two reductions:

1. A reduction of the search problem of R to deciding membership in $S'_R = \{(x, y') : \exists y'' \text{ s.t. } (x, y'y'') \in R\}$.

   As stated in the foregoing paragraph (titled "reflection"), such a reduction is implicit in the proof of Proposition 2.14 (as well as being explicit in the proof of Theorem 2.6).

2. A reduction of S'_R to S_R.

   This reduction exists by the hypothesis that S_R is $\mathcal{NP}$-complete and the fact that $S'_R \in \mathcal{NP}$. (Note that we do not assume that this reduction is a Karp-reduction, and furthermore it may be a "unnatural" reduction).

The theorem follows. ■

2.3 NP-Completeness

In light of the difficulty of settling the P-vs-NP Question, when faced with a hard problem H in NP, we cannot expect to prove that H is not in P (unconditionally).

The best we can expect is a conditional proof that H is not in P, based on the assumption that NP is different from P. The contrapositive is proving that if H is in P, then so is any problem in NP (i.e., NP equals P). One possible way of proving such an assertion is showing that any problem in NP is polynomial-time reducible to H. This is the essence of the theory of NP-completeness.

Teaching note: Some students heard of NP-completeness before, but we suspect that many have missed important conceptual points. Specifically, we fear that they missed the point that the mere existence of NP-complete problems is amazing (let alone that these problems include natural ones such as SAT). We believe that this situation is a consequence of presenting the detailed proof of Cook's Theorem as the very first thing right after defining NP-completeness.

2.3.1 Definitions

The standard definition of NP-completeness refers to decision problems. Below we will also present a definition of NP-complete (or rather $\mathcal{PC}$-complete) search problems. In both cases, NP-completeness of a problem Π combines two conditions:

1. Π is in the class (i.e., Π being in $\mathcal{NP}$ or $\mathcal{PC}$, depending on whether Π is a decision or a search problem).
2. Each problem in the class is reducible to Π. This condition is called NP-hardness.

Although a perfectly good definition could have allowed arbitrary Cook-reductions (for establishing NP-hardness), it turns out that Karp-reductions (resp., Levin-reductions) suffice for establishing the NP-hardness of all natural NP-complete decision (resp., search) problems. Consequently, NP-completeness is usually defined using this restricted notion of a polynomial-time reduction.

Definition 2.16 (NP-completeness of decision problems, restricted notion): *A set S is $\mathcal{NP}$-complete if it is in $\mathcal{NP}$ and every set in $\mathcal{NP}$ is Karp-reducible to S.*

A set is $\mathcal{NP}$-hard if every set in $\mathcal{NP}$ is Karp-reducible to it. Indeed, there is no reason to insist on Karp-reductions (rather than using arbitrary Cook-reductions), except that the restricted notion suffices for all known demonstrations of NP-completeness and is easier to work with. An analogous definition applies to search problems.

Definition 2.17 (NP-completeness of search problems, restricted notion): *A binary relation R is $\mathcal{PC}$-complete if it is in $\mathcal{PC}$ and every relation in $\mathcal{PC}$ is Levin-reducible to R.*

In the sequel, we will sometimes abuse the terminology and refer to search problems as NP-complete (rather than $\mathcal{PC}$-complete). Likewise, we will say that a search problem is NP-hard (rather than $\mathcal{PC}$-hard) if every relation in $\mathcal{PC}$ is Levin-reducible to it.

We stress that the mere fact that we have defined a property (i.e., NP-completeness) does not mean that there exist objects that satisfy this property. *It is indeed remarkable that NP-complete problems do exist.* Such problems are "universal" in the sense that solving them allows to solve any other (reasonable) problem (i.e., problems in NP).

2.3.2 The existence of NP-complete problems

We suggest not to confuse the mere existence of NP-complete problems, which is remarkable by itself, with the even more remarkable existence of "natural" NP-complete problems. The following proof delivers the first message as well as focuses on the essence of NP-completeness, rather than on more complicated technical details. The essence of NP-completeness is that a single computational problem may "effectively encode" a wide class of seemingly unrelated problems.

Theorem 2.18 *There exist NP-complete relations and sets.*

Proof: The proof (as well as any other NP-completeness proof) is based on the observation that some decision problems in $\mathcal{NP}$ (resp., search problems in $\mathcal{PC}$) are "rich enough" to encode all decision problems in $\mathcal{NP}$ (resp., all search problems in $\mathcal{PC}$). This fact is most obvious for the "generic" decision and search problems, denoted $S_{\mathtt{u}}$ and $R_{\mathtt{u}}$ (and defined next), which are used to derive the simplest proof of the current theorem.

We consider the following relation $R_{\mathtt{u}}$ and the decision problem $S_{\mathtt{u}}$ implicit in $R_{\mathtt{u}}$ (i.e., $S_{\mathtt{u}} = \{\overline{x} : \exists y \text{ s.t. } (\overline{x}, y) \in R_{\mathtt{u}}\}$). Both problems refer to the same type of instances, which in turn have the form $\overline{x} = \langle M, x, 1^t \rangle$, where M is a description of a (deterministic) Turing machine, x is a string, and t is a natural number. The number t is given in unary (rather than in binary) in order to allow various complexity measures, which depend on the instance length, to be polynomial in t (rather than poly-logarithmic in t).

Definition: *The relation $R_{\mathtt{u}}$ consists of pairs $(\langle M, x, 1^t\rangle, y)$ such that M accepts the input pair (x, y) within t steps, where $|y| \leq t$.*[7] The corresponding set $S_{\mathtt{u}} \overset{\text{def}}{=} \{\overline{x} : \exists y \text{ s.t. } (\overline{x}, y) \in R_{\mathtt{u}}\}$ consists of triples $\langle M, x, 1^t \rangle$ such that machine M accepts some input of the form $(x, \cdot)$ within t steps.

It is easy to see that $R_{\mathtt{u}}$ is in $\mathcal{PC}$ and that $S_{\mathtt{u}}$ is in $\mathcal{NP}$. Indeed, $R_{\mathtt{u}}$ is recognizable by a universal Turing machine, which on input $(\langle M, x, 1^t \rangle, y)$ emulates (t steps of) the computation of M on (x,y). (The fact that $S_{\mathtt{u}} \in \mathcal{NP}$ follows similarly.) We comment that u indeed stands for *universal* (i.e., universal machine), and the proof extends to any reasonable model of computation (which has adequate universal machines).

We now turn to show that $R_{\mathtt{u}}$ and $S_{\mathtt{u}}$ are NP-hard in the adequate sense (i.e., $R_{\mathtt{u}}$ is $\mathcal{PC}$-hard and $S_{\mathtt{u}}$ is $\mathcal{NP}$-hard). We first show that any set in $\mathcal{NP}$ is Karp-reducible to $S_{\mathtt{u}}$. Let S be a set in $\mathcal{NP}$ and let us denote its witness relation by

[7]Instead of requiring that $|y| \leq t$, one may require that M is "canonical" in the sense that it reads its entire input before halting.

R; that is, R is in $\mathcal{PC}$ and $x \in S$ if and only if there exists y such that $(x,y) \in R$. Let p_R be a polynomial bounding the length of solutions in R (i.e., $|y| \leq p_R(|x|)$ for every $(x,y) \in R$), let M_R be a polynomial-time machine deciding membership (of alleged (x,y) pairs) in R, and let t_R be a polynomial bounding its running-time. Then, the desired Karp-reduction maps an instance x (for S) to the instance $\langle M_R, x, 1^{t_R(|x|+p_R(|x|))}\rangle$ (for $S_{\mathtt{u}}$); that is,

$$x \mapsto f(x) \overset{\text{def}}{=} \langle M_R, x, 1^{t_R(|x|+p_R(|x|))}\rangle. \tag{2.3}$$

Note that this mapping can be computed in polynomial-time, and that $x \in S$ if and only if $f(x) = \langle M_R, x, 1^{t_R(|x|+p_R(|x|))}\rangle \in S_{\mathtt{u}}$. Details follow.

First, note that the mapping f does depend (of course) on S, and so it may depend on the fixed objects M_R, p_R and T_R (which depend on S). Thus, computing f on input x calls for printing the fixed string M_R, copying x, and printing a number of 1's that is a fixed polynomial in the length of x. Hence, f is polynomial-time computable. Second, recall that $x \in S$ if and only if there exists y such that $|y| \leq p_R(|x|)$ and $(x,y) \in R$. Since M_R accepts $(x,y) \in R$ within $t_R(|x|+|y|)$ steps, it follows that $x \in S$ if and only if there exists y such that $|y| \leq p_R(|x|)$ and M_R accepts (x,y) within $t_R(|x|+|y|)$ steps. It follows that $x \in S$ if and only if $f(x) \in S_{\mathtt{u}}$.

We now turn to the search version. For reducing the search problem of any $R \in \mathcal{PC}$ to the search problem of $R_{\mathtt{u}}$, we use essentially the same reduction. On input an instance x (for R), we make the query $\langle M_R, x, 1^{t_R(|x|+p_R(|x|))}\rangle$ to the search problem of $R_{\mathtt{u}}$ and return whatever the latter returns. Note that if $x \notin S$ then the answer will be "no solution", whereas for every x and y it holds that $(x,y) \in R$ if and only if $(\langle M_R, x, 1^{t_R(|x|+p_R(|x|))}\rangle, y) \in R_{\mathtt{u}}$. Thus, a Levin-reduction of R to $R_{\mathtt{u}}$ consists of the pair of functions (f,g), where f is the foregoing Karp-reduction and $g(x,y) = y$. Note that indeed, for every $(f(x),y) \in R_{\mathtt{u}}$, it holds that $(x, g(x,y)) = (x,y) \in R$. ■

Advanced comment. Note that the role of 1^t in the definition of $R_{\mathtt{u}}$ is to allow placing $R_{\mathtt{u}}$ in $\mathcal{PC}$. In contrast, consider the relation $R'_{\mathtt{u}}$ that consists of pairs $(\langle M,x,t\rangle, y)$ such that M accepts xy within t steps. Indeed, the difference is that in $R_{\mathtt{u}}$ the time-bound t appears in unary notation, whereas in $R'_{\mathtt{u}}$ it appears in binary. Then, as will become obvious in §4.2.1.2, membership in $R'_{\mathtt{u}}$ cannot be decided in polynomial time (even in the special case where x and y are fixed). Going even further, we note that omitting t altogether from the problem instance yields a search problem that is not solvable at all. That is, consider the relation $R_H \overset{\text{def}}{=} \{(\langle M,x\rangle, y) : M(xy) = 1\}$ (which is related to the halting problem). Indeed, the search problem of any relation (an in particular of any relation in $\mathcal{PC}$) is Karp-reducible to the search problem of R_H, but the latter is not solvable at all (i.e., there exists no algorithm that halts on every input and on input $\overline{x} = \langle M,x\rangle$ outputs y such that $(\overline{x},y) \in R_H$ if and only such a y exists).

Bounded Halting and Non-Halting

We note that the problem shown to be NP-complete in the proof of Theorem 2.18 is related to the following two problems, called `Bounded Halting` and `Bounded Non-Halting`. Fixing any programming language, the instance to each of these problems consists of a program π and a time bound t (presented in unary). The decision version of `Bounded Halting` (resp., `Bounded Non-Halting`) consists of determining whether or not *there exists an input* (of length at most t) *on which the program* π *halts in* t *steps* (resp., does *not* halt in t steps), whereas the search problem consists of finding such an input.

The decision version of `Bounded Non-Halting` refers to a fundamental computational problem in the area of program verification; specifically, the question of whether a given program halts within a given time-bound on all inputs of a given length.[8] We mention the `Bounded Halting` problem because it is often referred to in the literature, but we believe that `Bounded Non-Halting` is more relevant to the project of program verification (because one seeks programs that halt on all inputs rather than programs that halt on some input).

It is easy to prove that both problems are NP-complete (see Exercise 2.19). Note that the two (decision) problems are not complementary (i.e., $(M, 1^t)$ may be a yes-instance of both decision problems).[9]

The fact that `Bounded Non-Halting` is probably intractable (i.e., is intractable provided that $\mathcal{P} \neq \mathcal{NP}$) is even more relevant to the project of program verification than the fact that the Halting Problem is undecidable. The reason being that the latter problem (as well as other related undecidable problems) refers to arbitrarily long computations, whereas the former problem refers to computations of explicitly bounded number of steps. Specifically, `Bounded Non-Halting` is concerned with the existence of an input that causes the program to violate a certain condition (i.e., halting) within a given time-bound.

In light of the above, the common practice of bashing Bounded (Non-)Halting as an "unnatural" problem seems very odd at an age in which computer programs plays such a central role. (Nevertheless, we will use the term "natural" in this traditionally and odd sense in the next title, which refers to natural computational problems that seem unrelated to computation.)

[8]The length parameter need not equal the time-bound. Indeed, a more general version of the problem refers to two bounds, ℓ and t, and to whether the given program halts within t steps on each possible ℓ-bit input. It is easy to prove that the problem remains NP-complete also in the case that the instances are restricted to have parameters ℓ and t such that $t = p(\ell)$, for any fixed polynomial p (e.g., $p(n) = n^2$, rather than $p(n) = n$ as used in the main text).

[9]Indeed, $(M, 1^t)$ can not be a no-instance of both decision problems, but this does not make the problems complementary. In fact, the two decision problems yield a three-way partition of the instances $(M, 1^t)$: (1) pairs $(M, 1^t)$ such that for *every input* x (of length at most t) the computation of $M(x)$ halts within t steps, (2) pairs $(M, 1^t)$ for which such halting occurs on *some inputs but not on all inputs*, and (3) pairs $(M, 1^t)$ such that there *exists no input* (of length at most t) on which M halts in t steps. Note that instances of type (1) are no-instances of `Bounded Non-Halting`, whereas instances of type (3) are no-instances of `Bounded Halting`. It follows that recognizing each of these three sets of instances is NP-hard under Cook-reductions.

2.3.3 Some natural NP-complete problems

Having established the mere existence of NP-complete problems, we now turn to prove the existence of NP-complete problems that do not (explicitly) refer to computation in the problem's definition. We stress that thousands of such problems are known (and a list of several hundreds can be found in [81]).

We will prove that deciding the satisfiability of propositional formulae is NP-complete (i.e., Cook's Theorem), and also present some combinatorial problems that are NP-complete. This presentation is aimed at providing a (small) sample of natural NP-completeness results as well as some tools towards proving NP-completeness of new problems of interest. We start by making a comment regarding the latter issue.

The reduction presented in the proof of Theorem 2.18 is called "generic" because it (explicitly) refers to any (generic) NP-problem. That is, we actually presented a scheme for the design of reductions from any desired NP-problem to the single problem proved to be NP-complete. Indeed, in doing so, we have followed the definition of NP-completeness. However, once we know some NP-complete problems, a different route is open to us. We may establish the NP-completeness of a new problem by reducing a known NP-complete problem to the new problem. This alternative route is indeed a common practice, and it is based on the following simple proposition.

Proposition 2.19 *If an NP-complete problem Π is reducible to some problem Π' in NP then Π' is NP-complete. Furthermore, reducibility via Karp-reductions* (resp., Levin-reductions) *is preserved.*

Proof: The proof boils down to asserting the transitivity of reductions. Specifically, the NP-hardness of Π means that every problem in NP is reducible to Π, which in turn is reducible to Π'. Thus, by transitivity of reduction (see Exercise 2.6), every problem in NP is reducible to Π', which means that Π' is NP-hard and the proposition follows. ■

2.3.3.1 Circuit and formula satisfiability: CSAT and SAT

We consider two related computational problems, CSAT and SAT, which refer (in the decision version) to the satisfiability of Boolean circuits and formulae, respectively. (We refer the reader to the definition of Boolean circuits, formulae and CNF formulae that appear in §1.2.4.1.)

Teaching note: We suggest to establish the NP-completeness of SAT by a reduction from the circuit satisfaction problem (CSAT), after establishing the NP-completeness of the latter. Doing so allows to decouple two important parts of the proof of the NP-completeness of SAT: the emulation of Turing machines by circuits, and the encoding of circuits by formulae with auxiliary variables.

CSAT. Recall that Boolean circuits are directed acyclic graphs with internal vertices, called gates, labeled by Boolean operations (of arity either 2 or 1), and external vertices called terminals that are associated with either inputs or outputs. When setting the inputs of such a circuit, all internal nodes are assigned values in the natural way, and this yields a value to the output(s), called an evaluation of the circuit on the given input. The evaluation of circuit C on input z is denoted $C(z)$. We focus on circuits with a single output, and let CSAT denote the set of satisfiable Boolean circuits (i.e., a circuit C is in CSAT if there exists an input z such that $C(z) = 1$). We also consider the related relation $R_{\mathtt{CSAT}} = \{(C, z) : C(z) = 1\}$.

Theorem 2.20 (NP-completeness of CSAT): *The set* (resp., relation) CSAT (resp., $R_{\mathtt{CSAT}}$) *is $\mathcal{NP}$-complete* (resp., $\mathcal{PC}$-complete).

Proof: As usual it is easy to see that $\mathtt{CSAT} \in \mathcal{NP}$ (resp., $R_{\mathtt{CSAT}} \in \mathcal{PC}$). Thus, we turn to showing that these problems are NP-hard. We will focus on the decision version (but also discuss the search version).

We will present (again, but for the last time in this book) a generic reduction, this time of any NP-problem to CSAT. The reduction is based on the observation, mentioned in §1.2.4.1, that the computation of polynomial-time algorithms can be emulated by polynomial-size circuits. In the current context, we wish to emulate the computation of a fixed machine M on input (x, y), *where x is fixed and y varies* (but $|y| = \text{poly}(|x|)$ and the total number of steps of $M(x, y)$ is polynomial in $|x| + |y|$). Thus, x will be "hard-wired" into the circuit, whereas y will serve as the input to the circuit. The circuit itself, denoted C_x, will consists of "layers" such that each layer represents an instantaneous configuration of the machine M, and the relation between consecutive configurations in a computation of this machine is captured by ("uniform") local gadgets in the circuit. The number of layers will depend on the polynomial that upper-bounds the running-time of M, and an additional gadget will be used to detect whether the last configuration is accepting. Thus, only the first layer of the circuit C_x will depend on x. The punch-line is that determining whether, for a given x, there exists a y ($|y| = \text{poly}(|x|)$) such that $M(x, y) = 1$ (in a given number of steps) reduces to the question of whether there exists a y such that $C_x(y) = 1$. Performing this reduction for any machine M_R that corresponds to any $R \in \mathcal{PC}$ (as in the proof of Theorem 2.18), we establish the fact that CSAT is NP-complete. Details follow.

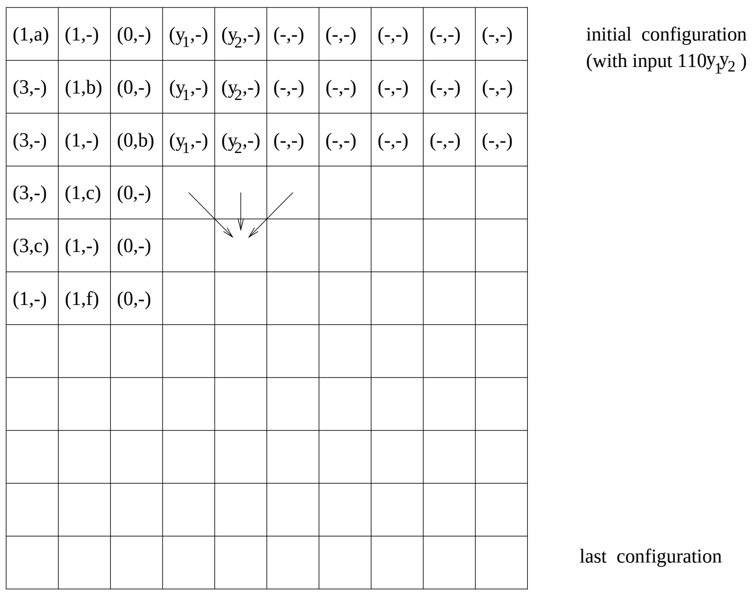
Recall that we wish to reduce an arbitrary set $S \in \mathcal{NP}$ to CSAT. Let R, p_R, M_R and t_R be as in the proof of Theorem 2.18 (i.e., R is the witness relation of S, whereas p_R bounds the length of the NP-witnesses, M_R is the machine deciding membership in R, and t_R is its polynomial time-bound). Without loss of generality (and for simplicity), suppose that M_R is a one-tape Turing machine. We will construct a Karp-reduction that maps an instance x (for S) to a circuit, denoted $f(x) \stackrel{\text{def}}{=} C_x$, such that $C_x(y) = 1$ if and only if M_R accepts the input (x, y) within $t_R(|x| + p_R(|x|))$ steps. Thus, it will follow that $x \in S$ if and only if there exists $y \in \{0,1\}^{p_R(|x|)}$ such that $C_x(y) = 1$ (i.e., if and only if $C_x \in \mathtt{CSAT}$). The circuit C_x will depend on x as well as on M_R, p_R and t_R. (We stress that M_R, p_R and t_R are fixed, whereas x is varies and thus explicit in our notation.)

Before describing the circuit C_x, let us consider a possible computation of M_R on input (x, y), where x is fixed and y represents a generic string of length at most $p_R(|x|)$. Such a computation proceeds for $t = t_R(|x| + p_R(|x|))$ steps, and corresponds to a sequence of $t + 1$ instantaneous configurations, each of length t. Each such configuration can be encoded by t pairs of symbols, where the first symbol in each pair indicates the contents of a cell and the second symbol indicates either a state of the machine or the fact that the machine is not located in this cell. Thus, each pair is a member of $\Sigma \times (Q \cup \{\perp\})$, where Σ is the finite "work alphabet" of M_R, Q is its finite set of internal states, and $\perp$ is an indication that the machine is not present at a cell. The initial configuration includes xy as input, and the decision of $M_R(x, y)$ can be read from (the leftmost cell of) the last configuration.[10] With the exception of the first row, the values of the entries in each row are determined by the entries of the row just above it, where this determination reflects the transition function of M_R. Furthermore, the value of each entry in the said array is determined by the values of (up to) three entries that reside in the row above it (see Exercise 2.20). Thus, the aforementioned computation is represented by a $(t + 1) \times t$ array, where each entry encodes one out of a constant number of possibilities, which in turn can be encoded by a constant-length bit string. See Figure 2.1.

The circuit C_x has a structure that corresponds to the aforementioned array. Each entry in the array is represented by a *constant* number of gates such that when C_x is evaluated at y these gates will be assigned values that encode the contents of the said entry. In particular, the entries of the first row of the array are "encoded" by hard-wiring the reduction's input (i.e., x), and feeding the circuit's input (i.e., y) to the adequate input terminals. That is, the circuit has $p_R(|x|)$ ("real") input terminals, and the hard-wiring of constants to the other $O(t - p_R(|x|))$ gates that represent the first row is done by simple gadgets (as in Figure 1.2). Indeed, additional hard-wiring in the first row corresponds to the other fixed elements of the initial configuration (i.e., the blank symbols, and the encoding of the initial state and of the initial location; cf. Figure 2.1). The entries of subsequent rows will be "encoded" (or rather computed at evaluation time) by using *constant-size* circuits that determine the value of an entry based on the three relevant entries in the row above it. Recall that each entry is encoded by a constant number of gates, and thus these constant-size circuits merely compute the constant-size function described in Exercise 2.20. In addition, the circuit will have a few extra gates that check the values of the entries of the last row in order to determine whether or not it encodes an accepting configuration.[11] Note that the circuit C_x can be constructed in polynomial time from the string x, because we just need to encode x in an appropriate manner as well as generate a "highly uniform" grid-like circuit

[10]We refer to the output convention presented in §1.2.3.1, by which the output is written in the leftmost cells and the machine halts at the cell to its right.

[11]In continuation to Footnote 10, we note that it suffices to check the values of the two leftmost entries of the last row. We assumed here that the circuit propagates a halting configuration to the last row. Alternatively, we may check for the existence of an accepting/halting configuration in the entire array, since this condition is quite simple.


(1,a) (1,-) (0,-) (y1,-) (y2,-) (-,-) (-,-) (-,-) (-,-) (-,-)
(3,-) (1,b) (0,-) (y1,-) (y2,-) (-,-) (-,-) (-,-) (-,-) (-,-)
(3,-) (1,-) (0,b) (y1,-) (y2,-) (-,-) (-,-) (-,-) (-,-) (-,-)
(3,-) (1,c) (0,-)
(3,c) (1,-) (0,-)
(1,-) (1,f) (0,-)
initial configuration (with input 110y1y2)
last configuration

Blank characters as well as the indication that the machine is not present in the cell are marked by a hyphen (-). The three arrows represent the determination of an entry by the three entries that reside above it. The machine underlying this example accepts the input if and only if the input contains a zero.

Figure 2.1: An array representing ten computation steps on input $110y_1y_2$.

of size $O(t_R(|x|+p_R(|x|))^2)$.[12]

Although the foregoing construction of C_x capitalizes on various specific details of the (one-tape) Turing machine model, it can be adapted to any other "reasonable" model of efficient computation.[13] Alternatively, we recall the Cobham-Edmonds Thesis asserting that any problem that is solvable in polynomial-time (on some "reasonable" model) can be solved in polynomial-time by a (one-tape) Turing machine.

Turning back to the circuit C_x, we observe that indeed $C_x(y) = 1$ if and only if M_R accepts the input (x, y) while making at most $t = t_R(|x| + p_R(|x|))$ steps. Recalling that $S = \{x : \exists y \text{ s.t. } |y| \le p_R(|x|) \wedge (x, y) \in R\}$ and that M_R decides membership in R in time t_R, we infer that $x \in S$ if and only if $f(x) = C_x \in \texttt{CSAT}$.

[12]**Advanced comment:** A more efficient construction, which generate almost-linear sized circuits (i.e., circuits of size $\widetilde{O}(t_R(|x| + p_R(|x|)))$) is known; see [170].

[13]**Advanced comment:** Note that it is actually inessential that each entry in each configuration is determined by a constant number of entries in the previous configuration. Any polynomial-time computable transformation of configurations will do, since we can emulate such a transformation by a polynomial-size circuit. Indeed, this emulation will be based on presenting the said transformation in some concrete model of computation, which brings us to the next comment (invoking the Cobham-Edmonds Thesis).

Furthermore, $(x,y) \in R$ if and only if $(f(x),y) \in R_{\texttt{CSAT}}$. It follows that f is a Karp-reduction of S to `CSAT`, and, for $g(x,y) \stackrel{\rm def}{=} y$ it holds that (f,g) is a Levin-reduction of R to $R_{\texttt{CSAT}}$. The theorem follows. ■

SAT. Recall that Boolean formulae are special types of Boolean circuits (i.e., circuits having a tree structure).[14] We further restrict our attention to formulae given in conjunctive normal form (CNF). We denote by `SAT` the set of satisfiable CNF formulae (i.e., a CNF formula ϕ is in `SAT` if there exists an truth assignment τ such that $\phi(\tau) = 1$). We also consider the related relation $R_{\texttt{SAT}} = \{(\phi,\tau) : \phi(\tau) = 1\}$.

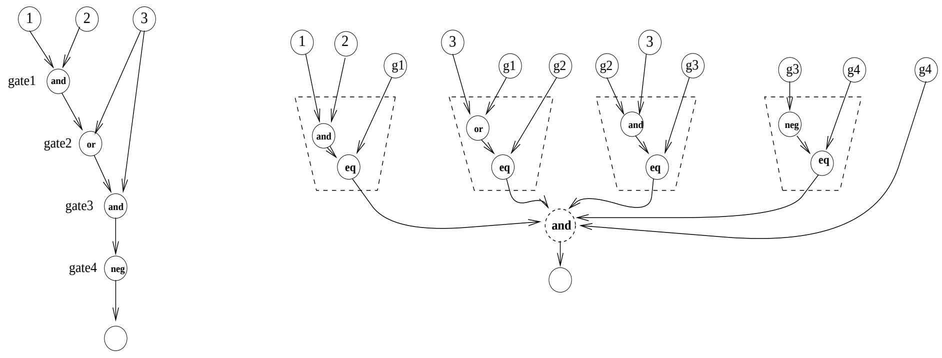
Theorem 2.21 (NP-completeness of SAT): *The set* (resp., relation) `SAT` (resp., $R_{\texttt{SAT}}$) *is* $\mathcal{NP}$*-complete* (resp., $\mathcal{PC}$-complete).

Proof: Since the set of possible instances of SAT is a subset of the set of instances of CSAT, it is clear that $\texttt{SAT} \in \mathcal{NP}$ (resp., $R_{\texttt{SAT}} \in \mathcal{PC}$). To prove that SAT is NP-hard, we reduce CSAT to SAT (and use Proposition 2.19). The reduction boils down to introducing auxiliary variables in order to "cut" the computation of an arbitrary ("deep") circuit into a conjunction of related computations of "shallow" circuits (i.e., depth-2 circuits) of unbounded fan-in, which in turn may be presented as a CNF formula. The aforementioned auxiliary variables hold the *possible* values of the internal gates of the original circuit, and the clauses of the CNF formula enforce the consistency of these values with the corresponding gate operation. For example, if $\texttt{gate}_i$ and $\texttt{gate}_j$ feed into $\texttt{gate}_k$, which is a $\wedge$-gate, then the corresponding auxiliary variables g_i, g_j, g_k should satisfy the Boolean condition $g_k \equiv (g_i \wedge g_j)$, which can be written as a 3CNF with four clauses. Details follow.

We start by Karp-reducing `CSAT` to `SAT`. Given a Boolean circuit C, with n input terminals and m gates, we first construct m *constant-size* formulae on $n+m$ variables, where the first n variables correspond to the input terminals of the circuit, and the other m variables correspond to its gates. The $i^{\rm th}$ formula will depend on the variable that correspond to the $i^{\rm th}$ gate and the 1-2 variables that correspond to the vertices that feed into this gate (i.e., 2 vertices in case of $\wedge$-gate or $\vee$-gate and a single vertex in case of a $\neg$-gate, where these vertices may be either input terminals or other gates). This (constant-size) formula will be satisfied by a truth assignment if and only if this assignment matches the gate's functionality (i.e., feeding this gate with the corresponding values result in the corresponding output value). Note that these *constant-size* formulae can be written as constant-size CNF formulae (in fact, as 3CNF formulae).[15] Taking the conjunction of these m formulae as well as the variable associated with the gate that feeds into the output terminal, we obtain a formula ϕ in CNF (see Figure 2.2, where $n = 3$ and $m = 4$).

[14] For an alternative definition, see Section G.2.

[15] Recall that any Boolean function can be written as a CNF formula having size that is exponential in the length of its input, which in this case is a constant (i.e., either 2 or 3). Futhermore, note that the Boolean functions that we refer to here depends on 2-3 Boolean variables (since they indicate whether or not the corresponding values respect the gate's functionality).



Using auxiliary variables (i.e., the g_i's) to "cut" a depth-5 circuit (into a CNF). The dashed regions will be replaced by equivalent CNF formulae. The dashed cycle representing an unbounded fan-in `and`*-gate is the conjunction of all constant-size circuits (which enforce the functionalities of the original gates) and the variable that represents the gate that feed the output terminal in the original circuit.*

Figure 2.2: The idea underlying the reduction of CSAT to SAT.

Note that ϕ can be constructed in polynomial-time from the circuit C; that is, the mapping of C to $\phi = f(C)$ is polynomial-time computable. We claim that C is in `CSAT` if and only if ϕ is in `SAT`.

1. Suppose that for some string s it holds that $C(s) = 1$. Then, assigning the i^{th} auxiliary variable the value that is assigned to the i^{th} gate of C when evaluated on s, we obtain (together with s) a truth assignment that satisfies ϕ. This is the case because such an assignment satisfies all m constant-size CNFs as well as the variable associated with the output of C.

2. On the other hand, if τ satisfies ϕ then the first n bits in τ correspond to an input on which C evaluates to 1. This is the case because the m constant-size CNFs guarantee that the variables of ϕ are assigned values that correspond to the evaluation of C on the first n bits of τ, while the fact that the variable associated with the output of C has value `true` guarantees that this evaluation of C yields the value 1.

   Note that the latter mapping (of τ to its n-bit prefix) is the second mapping required by the definition of a Levin-reduction.

Thus, we have established that f is a Karp-reduction of `CSAT` to `SAT`, and that augmenting f with the aforementioned second mapping yields a Levin-reduction of $R_{\texttt{CSAT}}$ to $R_{\texttt{SAT}}$. ■

Comment. The fact that the second mapping required by the definition of a Levin-reduction is explicit in the proof of the validity of the corresponding Karp-reduction is a fairly common phenomenon. Actually (see Exercise 2.28), typical presentations of Karp-reductions provide two auxiliary polynomial-time computable mappings (in addition to the main mapping of instances from one problem (e.g., CSAT) to instances of another problem (e.g., SAT)): The first auxiliary mapping is of solutions for the preimage instance (e.g., of CSAT) to solutions for the image instance of the reduction (e.g., of SAT), whereas the second mapping goes the other way around. (Note that only the main mapping and the second auxiliary mapping are required in the definition of a Levin-reduction.) For example, the proof of the validity of the Karp-reduction of CSAT to SAT, denoted f, specified two additional mappings h and g such that $(C,s) \in R_{\mathtt{CSAT}}$ implies $(f(C),h(C,s)) \in R_{\mathtt{SAT}}$ and $(f(C),\tau) \in R_{\mathtt{SAT}}$ implies $(C,g(C,\tau)) \in R_{\mathtt{CSAT}}$. Specifically, in the proof of Theorem 2.21, we used $h(C,s) = (s,a_1,...,a_m)$ where a_i is the value assigned to the i^{th} gate in the evaluation of $C(s)$, and $g(C,\tau)$ being the n-bit prefix of τ.

3SAT. Note that the formulae resulting from the Karp-reduction presented in the proof of Theorem 2.21 are in conjunctive normal form (CNF) with each clause referring to at most three variables. Thus, the above reduction actually establishes the NP-completeness of 3SAT (i.e., SAT restricted to CNF formula with up to three variables per clause). Alternatively, one may Karp-reduce SAT (i.e., satisfiability of CNF formula) to 3SAT (i.e., satisfiability of 3CNF formula), by replacing long clauses with conjunctions of three-variable clauses using auxiliary variables (see Exercise 2.21). Either way, we get the following result, where the furthermore part is proved by an additional reduction.

Proposition 2.22 *3SAT is NP-complete. Furthermore, the problem remains NP-complete also if we restrict the instances such that each variable appears in at most three clauses.*

Proof Sketch: The furthermore part is proved by reduction from 3SAT. We just replace each occurrence of each Boolean variable by a new copy of this variable, and add clauses to enforce that all these copies are assigned the same value. Specifically, replacing the variable z by copies $z^{(1)},...,z^{(m)}$, we add the clauses $z^{(i+1)} \vee \neg z^{(i)}$ for $i = 1...,m$ (where $m+1$ is understood as 1). $\square$

Related problems. Note that instances of SAT can be viewed as systems of Boolean conditions over Boolean variables. Such systems can be emulated by various types of systems of arithmetic conditions, implying the NP-hardness of solving the latter types of systems. Examples include systems of *integer* linear inequalities (see Exercise 2.23), and systems of quadratic equalities (see Exercise 2.25).

2.3.3.2 Combinatorics and graph theory

Teaching note: The purpose of this subsection is to expose the students to a sample of NP-completeness results and proof techniques (i.e., the design of reductions among computational problems). The author believes that this traditional material is insightful, but one may skip it in the context of a complexity class.

We present just a few of the many appealing combinatorial problems that are known to be NP-complete. Throughout this section, we focus on the decision versions of the various problems, and adopt a more informal style. Specifically, we will present a typical decision problem as a problem of deciding whether a given instance, which belongs to a set of relevant instances, is a "yes-instance" or a "no-instance" (rather than referring to deciding membership of arbitrary strings in a set of yes-instances). For further discussion of this style and its rigorous formulation, see Section 2.4.1. We will also neglect to show that these decision problems are in NP.

We start with the set cover problem, in which an instance consists of a collection of finite sets $S_1, ..., S_m$ and an integer K and the question (for decision) is whether or not there exist (at most)[16] K sets that cover $\bigcup_{i=1}^m S_i$ (i.e., indices $i_1, ..., i_K$ such that $\bigcup_{j=1}^K S_{i_j} = \bigcup_{i=1}^m S_i$).

Proposition 2.23 `Set Cover` *is NP-complete.*

Proof Sketch: We sketch a reduction of SAT to Set Cover. For a CNF formula ϕ with m clauses and n variables, we consider the sets $S_{1,\mathtt{t}}, S_{1,\mathtt{f}}, .., S_{n,\mathtt{t}}, S_{n,\mathtt{f}} \subseteq \{1, ..., m\}$ such that $S_{i,\mathtt{t}}$ (resp., $S_{i,\mathtt{f}}$) is the set of the indices of the clauses (of ϕ) that are satisfied by setting the i^{th} variable to `true` (resp., `false`). That is, if the i^{th} variable appears unnegated (resp., negated) in the j^{th} clause then $j \in S_{i,\mathtt{t}}$ (resp., $j \in S_{i,\mathtt{f}}$). Note that the union of these $2n$ sets equals $\{1, ..., m\}$. Now, on input ϕ, the reduction outputs the Set Cover instance $f(\phi) \stackrel{\text{def}}{=} ((S_1, .., S_{2n}), n)$, where $S_{2i-1} = S_{i,\mathtt{t}} \cup \{m+i\}$ and $S_{2i} = S_{i,\mathtt{f}} \cup \{m+i\}$ for $i = 1, ..., n$.

Note that f is computable in polynomial-time, and that if ϕ is satisfied by $\tau_1 \cdots \tau_n$ then the collection $\{S_{2i-\tau_i} : i = 1, ..., n\}$ covers $\{1, ..., m+n\}$. Thus, $\phi \in SAT$ implies that $f(\phi)$ is a yes-instance of `Set Cover`. On the other hand, each cover of $\{m+1, ..., m+n\} \subset \{1, ..., m+n\}$ must include either S_{2i-1} or S_{2i} for each i. Thus, a cover of $\{1, ..., m+n\}$ using n of the S_j's must contain, for every i, either S_{2i-1} or S_{2i} but not both. Setting τ_i accordingly (i.e., $\tau_i = 1$ if and only if S_{2i-1} is in the cover) implies that $\{S_{2i-\tau_i} : i = 1, ..., n\}$ covers $\{1, ..., m\}$, which in turn implies that $\tau_1 \cdots \tau_n$ satisfies ϕ. Thus, if $f(\phi)$ is a yes-instance of `Set Cover` then $\phi \in$ `SAT`. □

Exact Cover and `3XC`. The exact cover problem is similar to the set cover problem, except that here the sets that are used in the cover are not allowed to intersect. That is, each element in the universe should be covered by *exactly* one set in the cover. Restricting the set of instances to sequences of subsets each having exactly three elements, we get the restricted problem `3-Exact Cover` (`3XC`), where it is unnecessary to specify the number of sets to be used in the cover. The problem `3XC` is rather technical, but it is quite useful for demonstrating the NP-completeness of other problems (by reducing `3XC` to them).

Proposition 2.24 `3-Exact Cover` *is NP-complete.*

[16]Clearly, in case of `Set Cover`, the two formulations (i.e., asking for exactly K sets or at most K sets) are computationally equivalent.

Indeed, it follows that the `Exact Cover` (in which sets of arbitrary size are allowed) is NP-complete. This follows both for the case that the number of sets in the desired cover is unspecified and for the various cases in which this number is bounded (i.e., upper-bounded or lower-bounded or both).

Proof Sketch: The reduction is obtained by composing three reductions. We first reduce a *restricted case* of `3SAT` to a restricted version of `Set Cover`, denoted `3SC`, in which each set has at most three elements (and an instance consists, as in the general case, of a sequence of finite sets as well as an integer K). Specifically, we refer to `3SAT` instances that are restricted such that each *variable* appears in at most three clauses, and recall that this restricted problem is NP-complete (see Proposition 2.22). Actually, we further reduce this special case of `3SAT` to one in which each *literal* appears in at most two clauses.[17] Now, we reduce the new version of `3SAT` to `3SC` by using the (very same) reduction presented in the proof of Proposition 2.23, and observing that the size of each set in the reduced instance is at most three (i.e., one more than the number of occurrences of the corresponding literal).

Next, we reduce `3SC` to the following restricted case of `Exact Cover`, denoted `3XC`', in which each set has *at most* three elements, an instance consists of a sequence of finite sets as well as an integer K, and the question is whether there exists an exact cover with at most K sets. The reduction maps an instance $((S_1, ..., S_m), K)$ of `3SC` to the instance (C', K) such that C' is a collection of all subsets of each of the sets $S_1, ..., S_m$. Since each S_i has size at most 3, we introduce at most 7 non-empty subsets per each such set, and the reduction can be computed in polynomial-time. The reader may easily verify the validity of this reduction.

Finally, we reduce `3XC`' to `3XC`. Consider an instance $((S_1, ..., S_m), K)$ of `3XC`', and suppose that $\bigcup_{i=1}^m S_i = [n]$. If $n > 3K$ then this is definitely a no-instance, which can be mapped to a dummy no-instance of `3XC`, and so we assume that $x \stackrel{\text{def}}{=} 3K - n \geq 0$. Note that x represents the "excess" covering ability of an exact cover having K sets, each having three elements. Thus, we augment the set system with x new elements, denoted $n + 1, ..., 3K$, and replace each S_i such that $|S_i| < 3$ by a sub-collection of 3-sets that cover S_i as well as arbitrary elements from $\{n + 1, ..., 3K\}$. That is, in case $|S_i| = 2$, the set S_i is replaced by the sub-collection $(S_i \cup \{n+1\}, ..., S_i \cup \{3K\})$, whereas a singleton S_i is replaced by the sets $S_i \cup \{j_1, j_2\}$ for every $j_1 < j_2$ in $\{n + 1, ..., 3K\}$. In addition, we add all possible 3-subsets of $\{n + 1, ..., 3K\}$. This completes the description of the third reduction, the validity of which is left as an exercise. $\square$

[17]This can be done by observing that if all three occurrences of a variable are of the same type (i.e., they are all negated or all non-negated) then this variable can be assigned a value that satisfies all clauses in which it appears, and so the variable and the clauses in which it appear can be omitted from the instance. This yields a reduction of `3SAT` instances in which each variable appears in at most three clauses to `3SAT` instances in which each literal appears in at most two clauses. Actually, a closer look at the proof of Proposition 2.22 reveals the fact that the reduced instances satisfy the latter property anyhow.

Vertex Cover, Independent Set, and Clique. Turning to graph theoretic problems (see Section G.1), we start with the `Vertex Cover` problem, which is a special case of the `Set Cover` problem. The instances consists of pairs (G, K), where $G = (V, E)$ is a simple graph and K is an integer, and the problem is whether or not there exists a set of (at most) K vertices that is incident to all graph edges (i.e., each edge in G has at least one endpoint in this set). Indeed, this instance of `Vertex Cover` can be viewed as an instance of `Set Cover` by considering the collection of sets $(S_v)_{v \in V}$, where S_v denotes the set of edges incident at vertex v (i.e., $S_v = \{e \in E : v \in e\}$). Thus, the NP-hardness of `Set Cover` follows from the NP-hardness of `Vertex Cover` (but this implication is unhelpful for us here: we already know that `Set Cover` is NP-hard and we wish to prove that `Vertex Cover` is NP-hard). We also note that the `Vertex Cover` problem is computationally equivalent to the `Independent Set` and `Clique` problems (see Exercise 2.26), and thus it suffices to establish the NP-hardness of one of these problems.

Proposition 2.25 *The problems* `Vertex Cover`*,* `Independent Set` *and* `Clique` *are NP-complete.*

Teaching note: The following reduction is not the "standard" one (see Exercise 2.27). It is rather adapted from the FGLSS-reduction (see Exercise 9.14), and is used here in anticipation of the latter. Furthermore, although the following reduction tends to create a larger graph, the author finds it more clear than the "standard" reduction.

Proof Sketch: We show a reduction from 3SAT to `Independent Set`. On input a 3CNF formula ϕ with m clauses and n variables, we construct a graph with $7m$ vertices, denoted G_ϕ. The vertices are grouped in m cliques, each corresponding to one of the clauses and containing 7 vertices that correspond to the 7 truth assignments (to the 3 variables in the clause) that *satisfy the clause*. In addition to the internal edges of these m cliques, we add an edge between each pair of vertices that correspond to partial assignments that are *mutually inconsistent*. That is, if a specific (satisfying) assignment to the variables of the i^{th} clause is inconsistent with some (satisfying) assignment to the variables of the j^{th} clause then we connect the corresponding vertices by an edge. (Note that the internal edges of the m cliques may be viewed as a special case of the edges connecting mutually inconsistent partial assignments.) Thus, on input ϕ, the reduction outputs the pair (G_ϕ, m).

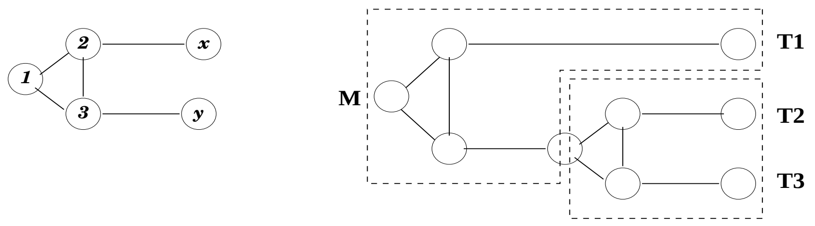
Note that if ϕ is satisfiable by a truth assignment τ then there are no edges between the m vertices that correspond to the partial satisfying assignment derived from τ. (We stress that any truth assignment to ϕ yields an independent set, but only a satisfying assignment guarantees that this independent set contains a vertex from each of the m cliques.) Thus, $\phi \in \texttt{SAT}$ implies that G_ϕ has an independent set of size m. On the other hand, an independent set of size m in G_ϕ must contain exactly one vertex in each of the m cliques, and thus induces a truth assignment that satisfies ϕ. (We stress that each independent set induces a consistent truth assignment to ϕ, because the partial assignments selected in the various cliques must be consistent, and that an independent set containing a vertex from a specific clique induces an assignment that satisfies the corresponding clause.) Thus, if G_ϕ has an independent set of size m then $\phi \in \texttt{SAT}$. □

Graph 3-Colorability (`G3C`). In this problem the instances are graphs and the question is whether or not the graph can be colored using three colors (such that neighboring vertices are not assigned the same color).

Proposition 2.26 `Graph 3-Colorability` *is NP-complete.*

Proof Sketch: We reduce `3SAT` to `G3C` by mapping a 3CNF formula ϕ to the graph G_ϕ, which consists of two special ("designated") vertices, a gadget per each variable of ϕ, a gadget per each clause of ϕ, and edges connecting some of these components.

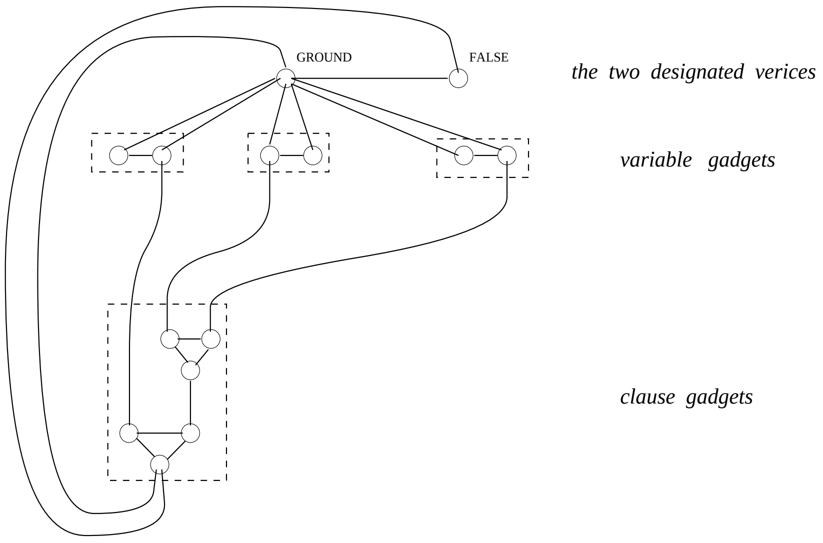
- The two designated vertices are called `ground` and `false`, and are connected by an edge that ensures that they must be given different colors in any 3-coloring of G_ϕ. We will refer to the color assigned to the vertex `ground` (resp., `false`) by the name `ground` (resp., `false`). The third color will be called `true`.
- The gadget associated with variable x is a pair of vertices, associated with the two literals of x (i.e., x and $\neg x$). These vertices are connected by an edge, and each of them is also connected to the vertex `ground`. Thus, in a 3-coloring of G_ϕ one of the vertices associated with the variable is colored `true` and the other is colored `false`.



In a generic 3-coloring of the sub-gadget it must hold that if $x = y$ then $x = y = 1$. Thus, if the three terminals of the gadget are assigned the same color, χ, then M is also assigned the color χ.

Figure 2.3: The reduction to G3C – the clause gadget and its sub-gadget.

- The gadget associated with a clause C is depicted in Figure 2.3. It contains a master vertex, denoted **M**, and three terminal vertices, denoted **T1**, **T2** and **T3**. The master vertex is connected by edges to the vertices `ground` and `false`, and thus in a 3-coloring of G_ϕ the master vertex must be colored `true`. The gadget has the property that it is possible to color the terminals with any combination of the colors `true` and `false`, except for coloring all terminals with `false`. The terminals of the gadget associated with clause C will be *identified* with the vertices that are associated with the corresponding literals appearing in C. This means that the various clause-gadgets are not vertex-disjoint but may rather share some terminals (with the vertex-gadgets as well as among themselves).[18] See Figure 2.4.



A single clause gadget and the relevant variables gadgets.

Figure 2.4: The reduction to G3C – connecting the gadgets.

Verifying the validity of the reduction is left as an exercise. □

2.3.4 NP sets that are neither in P nor NP-complete

As stated in Section 2.3.3, thousands of problems have been shown to be NP-complete (cf., [81, Apdx.], which contains a list of more than three hundreds main entries). Things reached a situation in which people seem to expect any NP-set to be either NP-complete or in $\mathcal{P}$. This naive view is wrong: *Assuming $\mathcal{NP} \neq \mathcal{P}$, there exist sets in $\mathcal{NP}$ that are neither NP-complete nor in $\mathcal{P}$, where here NP-hardness allows also Cook-reductions.*

Theorem 2.27 *Assuming $\mathcal{NP} \neq \mathcal{P}$, there exist a set T in $\mathcal{NP} \setminus \mathcal{P}$ such that some sets in $\mathcal{NP}$ are not Cook-reducible to T.*

Theorem 2.27 asserts that if $\mathcal{NP} \neq \mathcal{P}$ then $\mathcal{NP}$ is partitioned into three non-empty classes: the class $\mathcal{P}$, the class of problems to which $\mathcal{NP}$ is Cook-reducible, and the rest, denote $\mathcal{NPI}$. We already know that the first two classes are not empty,

[18]Alternatively, we may use disjoint gadgets and "connect" each terminal with the corresponding literal (in the corresponding vertex gadget). Such a connection (i.e., an auxiliary gadget) should force the two end-points to have the same color in any 3-coloring of the graph.

and Theorem 2.27 establishes the non-emptiness of $\mathcal{NPI}$ under the condition that $\mathcal{NP} \neq \mathcal{P}$, which is actually a necessary condition (because if $\mathcal{NP} = \mathcal{P}$ then every set in $\mathcal{NP}$ is Cook-reducible to any other set in $\mathcal{NP}$).

The following proof of Theorem 2.27 presents an unnatural decision problem in $\mathcal{NPI}$. We mention that some natural problems (e.g., factoring) are conjectured to be neither solvable in polynomial-time nor NP-hard. In particular, assuming that factoring is intractable, there exist rather natural decision problems in $\mathcal{NPI}$. Furthermore, if $\mathcal{NP} \neq \text{co}\mathcal{NP}$, where $\text{co}\mathcal{NP} = \{\{0,1\}^* \setminus S : S \in \mathcal{NP}\}$, then $\Delta \stackrel{\rm def}{=} \mathcal{NP} \cap \text{co}\mathcal{NP} \subseteq \mathcal{P} \cup \mathcal{NPI}$ holds (as a corolloary to Theorem 2.33). In other words, if $\mathcal{NP} \neq \text{co}\mathcal{NP}$ then $\Delta \setminus \mathcal{P}$ is a (natural) subset of $\mathcal{NPI}$, and the non-emptiness of $\mathcal{NPI}$ follows provided that $\Delta \neq \mathcal{P}$. Recall that Theorem 2.27 establishes the non-emptiness of $\mathcal{NPI}$ under the seemingly weaker assumption that $\mathcal{NP} \neq \mathcal{P}$.

Teaching note: We recommend either stating Theorem 2.27 without a proof or merely providing the proof idea.

Proof Sketch: The basic idea is modifying an arbitrary set in $\mathcal{NP} \setminus \mathcal{P}$ so as to fail all possible reductions (from $\mathcal{NP}$ to the modified set) as well as all possible polynomial-time decision procedures (for the modified set). Specifically, starting with $S \in \mathcal{NP} \setminus \mathcal{P}$, we derive $S' \subset S$ such that on one hand there is no polynomial-time reduction of S to S' while on the other hand $S' \in \mathcal{NP} \setminus \mathcal{P}$. The process of modifying S into S' proceeds in iterations, alternatively failing a potential reduction (by dropping sufficiently many strings from the rest of S) and failing a potential decision procedure (by including sufficiently many strings from the rest of S). Specifically, each potential reduction of S to S' can be failed by dropping finitely many elements from the current S', whereas each potential decision procedure can be failed by keeping finitely many elements of the current S'. These two assertions are based on the following two corresponding facts:

1. Any polynomial-time reduction (of any set not in $\mathcal{P}$) to any finite set (e.g., a finite subset of S) must fail, because only sets in $\mathcal{P}$ are Cook-reducible to a finite set. Thus, for any finite set F and any potential reduction (i.e., a polynomial-time oracle machine), there exists an input x on which this reduction to F fails.

   We stress that the aforementioined reduction fails while the only queries that are answered positively are those residing in F. Furthermore, the aforementioined failure is due to a finite set of queries (i.e., the set of all queries made by the reduction when invoked on an input that is smaller or equal to x). Thus, for every finite set $F \subset S' \subseteq S$, any reduction of S to S' can be failed by dropping a finite number of elements from S' and without dropping elements of F.

2. For every finite set F, any polynomial-time decision procedure for $S \setminus F$ must fail, because S is (trivially) Cook-reducible to $S \setminus F$. Thus, for any potential decision procedure (i.e., a polynomial-time algorithm), there exists an input x on which this procedure fails.

> We stress that this failure is due to a finite "prefix" of $S \setminus F$ (i.e., the set $\{z \in S \setminus F : z \leq x\}$). Thus, for every finite set F, any polynomial-time decision procedure for $S \setminus F$ can be failed by keeping a finite subset of $S \setminus F$.

As stated, the process of modifying S into S' proceeds in iterations, alternatively failing a potential reduction (by dropping finitely many strings from the rest of S) and failing a potential decision procedure (by including finitely many strings from the rest of S). This can be done efficiently because *it is inessential to determine the first possible points of alternation* (in which sufficiently many strings were dropped (resp., included) to fail the next potential reduction (resp., decision procedure)). It suffices to guarantee that adequate points of alternation (albeit highly non-optimal ones) can be efficiently determined. Thus, S' is the intersection of S and some set in $\mathcal{P}$, which implies that $S' \in \mathcal{NP}$. Following are some comments regarding the implementation of the foregoing idea.

The first issue is that the foregoing plan calls for an ("effective") enumeration of all polynomial-time oracle machines (resp., polynomial-time algorithms). However, none of these sets can be enumerated (by an algorithm). Instead, we enumerate all corresponding machines along with all possible polynomials, and for each pair (M,p) we consider executions of machine M with time bound specified by the polynomial p. That is, we use the machine M_p obtained from the pair (M,p) by suspending the execution of M on input x after $p(|x|)$ steps. We stress that we do not know whether machine M runs in polynomial-time, but the computations of any polynomial-time machine is "covered" by some pair (M,p).

Next, let us clarify the process in which reductions and decision procedures are ruled out. We present a construction of a "filter" set F in $\mathcal{P}$ such that the final set S' will equal $S \cap F$. Recall that we need to select F such that each polynomial-time reduction of S to $S \cap F$ fails, and each polynomial-time procedure for deciding $S \cap F$ fails. The key observation is that for every finite F each polynomial-time reduction of S to $S \cap F$ fails, whereas for every co-finite F (i.e., finite $\{0,1\}^* \setminus F$) each polynomial-time procedure for deciding $S \cap F$ fails. Furthermore, each of these failures occur on some input, and this input is determined by finite portions of S and F. Thus, we alternate between failing possible reductions and decision procedures, while not trying to determine the "optimal" points of alternation but rather determining points of alternation in a way that allows for efficiently deciding membership in F. Specifically, we let $F = \{x : f(|x|) \equiv 0 \bmod 2\}$, where $f : \mathbb{N} \to \{0\} \cup \mathbb{N}$ is defined next such that $f(n)$ can be computed in time poly(n).

The value of $f(n)$ is defined by the the following experiment that consists of exactly n^3 computation steps (where cubic time is selected to allow for some non-trivial manipulations of data as conducted next). For $i = 0, 1, ...$, we scan all inputs in lexicographic order trying to find an input on which the $i+1^{\text{st}}$ (modified) machine fails (where this machine is an oracle machine if i is even and a standard machine otherwise). In order to determine whether or not a failure occurs on a particular input x, we may need to know whether or not x is in the set $S' = S \cap F$ as well as whether some other strings (which may appear as queries) are in S'. Deciding membership in $S \in \mathcal{NP}$ can be done in exponential-time (by using the exhaustive search algorithm that tries all possible NP-witnesses). Indeed, this means that when computing $f(n)$ we may only complete the treatment of inputs that are of logarithmic (in n) length, but anyhow in n^3 steps we can not hope to reach (in our lexicographic scanning) strings of length $3\log_2 n$. As for deciding membership in F, this requires ability to compute f on adequate integers. That is, we may need to compute the value of $f(n')$ for various integers n', but as noted n' will be much smaller than n. Thus, the value of $f(n')$ is just computed recursively (while counting the recursive steps in our total number of steps).[19] The point is that, when considering an input x, we may need the values of f only on $\{1,...,p_{i+1}(|x|)\}$, where p_{i+1} is the polynomial bounding the running-time of the $i+1^{\rm st}$ (modified) machine, and obtaining such a value takes at most $p_{i+1}(|x|)^3$ steps. Finally, if we detect a failure of the $i+1^{\rm st}$ machine, then we increase i and proceed to the next iteration. When we reach the allowed number of steps (i.e., n^3 steps), we halt outputting the current value of i (i.e., the current i is output as the value of $f(n)$).

As hinted in the foregoing, it is most likely that we will complete n^3 steps much before examining inputs of length $3\log_2 n$, but this does not matter. What matters is that f *is monotonically non-decreasing* (because more steps allow to fail at least as many machines) and that f *is unbounded* (see Exercise 2.34). Furthermore, by construction, $f(n)$ is computed in poly(n) time. □

Comment: The proof of Theorem 2.27 actually establishes that *for every $S \notin \mathcal{P}$ there exists $S' \notin \mathcal{P}$ such that S' is Karp-reducible to S but S is not Cook-reducible to S'*.[20] Thus, if $\mathcal{P} \neq \mathcal{NP}$ then there exists an infinite sequence of sets $S_1, S_2, ...$ in $\mathcal{NP} \setminus \mathcal{P}$ such that S_{i+1} is Karp-reducible to S_i but S_i is not Cook-reducible to S_{i+1}. That is, there exists an infinite hierarchy of problems (albeit unnatural ones), all in $\mathcal{NP}$, such that each problem is "easier" than the previous ones (in the sense that it can be reduced to the previous problems while these problems cannot be reduced to it).

2.4 Three relatively advanced topics

In this section we discuss three relatively advanced topics. The first topic, which was eluded to in previous sections, is the notion of promise problems (Section 2.4.1). Next we present an optimal search algorithm for NP (Section 2.4.2), and discuss the class (coNP) of sets that are complements of sets in NP.

Teaching note: These topics are typically not mentioned in a basic course on complexity. Still, pending on time constraints, we suggest discussing them at least at a high level.

[19] We do not bother to present a more efficient implementation of this process. That is, we may afford to recompute $f(n')$ every time we need it (rather than store it for later use).

[20] The said Karp-reduction (of S' to S) maps x to itself if $x \in F$ and otherwise maps x to a fixed no-instance of S.

2.4.1 Promise Problems

Promise problems are a natural generalization of search and decision problems, where one explicitly considers a set of legitimate instances (rather than considering any string as a legitimate instance). As noted before, this provides a more adequate formulation of natural computational problems (and indeed this formulation is used in all informal discussions). For example, in §2.3.3.2 we presented such problems using phrases like "given a graph and an integer..." (or "given a collection of sets..."). In other words, we assumed that the input instance has a certain format (or rather we "promised the solver" that this is the case). Indeed, we claimed that in these cases the assumption can be removed without affecting the complexity of the problem, but we avoided providing a formal treatment of this issue, which is done next.

Teaching note: The notion of promise problems was originally introduced in the context of decision problems, and is typically used only in that context. However, we believe that promise problems are as natural in the context of search problems, and present things accordingly.

2.4.1.1 Definitions

In the context of search problems, a promise problem is a relaxation in which one is only required to find solutions to instances in a predetermined set, called the promise. The requirement regarding efficient checkability of solutions is adapted in an analogous manner.

Definition 2.28 (search problems with a promise): *A* search problem with a promise *consists of a binary relation $R \subseteq \{0,1\}^* \times \{0,1\}^*$ and a* promise *set P. Such a problem is also referred to as the* search problem R with promise P.

- *The search problem R with promise P is* solved by algorithm A *if for every $x \in P$ it holds that $(x, A(x)) \in R$ if $x \in S_R = \{x : R(x) \neq \emptyset\}$ and $A(x) = \perp$ otherwise, where $R(x) = \{y : (x,y) \in R\}$.*

  The time complexity of A on inputs in P *is defined as $T_{A|P}(n) \stackrel{\text{def}}{=} \max_{x \in P \cap \{0,1\}^n}\{t_A(x)\}$, where $t_A(x)$ is the running time of $A(x)$ and $T_{A|P}(n) = 0$ if $P \cap \{0,1\}^n = \emptyset$.*

- *The search problem R with promise P is in the* promise problem extension of $\mathcal{PF}$ *if there exists a polynomial-time algorithm that solves this problem.*[21]

- *The search problem R with promise P is in the* promise problem extension of $\mathcal{PC}$ *if there exists a polynomial T and an algorithm A such that, for every $x \in P$ and $y \in \{0,1\}^*$, algorithm A makes at most $T(|x|)$ steps and it holds that $A(x,y) = 1$ if and only if $(x,y) \in R$.*

[21]In this case it does not matter whether the time complexity of A is defined on inputs in P or on all possible strings. Suppose that A has (polynomial) time complexity T on inputs in P, then we can modify A to halt on any input x after at most $T(|x|)$ steps. This modification may only effects the output of A on inputs not in P (which is OK by us). The modification can be implemented in polynomial-time by computing $t = T(|x|)$ and emulating the execution of $A(x)$ for t steps. A similar comment applies to the definition of $\mathcal{PC}$, $\mathcal{P}$ and $\mathcal{NP}$.

We stress that nothing is required of the solver in the case that the input violates the promise (i.e., $x \notin P$); in particular, in such a case the algorithm may halt with a wrong output. (Indeed, the standard formulation of search problems is obtained by considering the trivial promise $P = \{0,1\}^*$.)[22] In addition to the foregoing motivation for promise problems, we mention one natural class of search problems with a promise. These are search problem in which the promise is that the instance has a solution (i.e., in terms of the foregoing notation $P = S_R$). We refer to such search problems by the name candid search problems.

Definition 2.29 (candid search problems): *An algorithm A solves the* candid search problem of the binary relation *R if for every $x \in S_R \stackrel{\rm def}{=} \{x : \exists y \text{ s.t. } (x,y) \in R\}$ it holds that $(x, A(x)) \in R$. The time complexity of such an algorithm is defined as $T_{A|S_R}(n) \stackrel{\rm def}{=} \max_{x \in P \cap \{0,1\}^n}\{t_A(x)\}$, where $t_A(x)$ is the running time of $A(x)$ and $T_{A|S_R}(n) = 0$ if $P \cap \{0,1\}^n = \emptyset$.*

Note that nothing is required when $x \notin S_R$: In particular, algorithm A may either output a wrong solution (although no solutions exist) or run for more than $T_{A|S_R}(|x|)$ steps. The first case can be essentially eliminated whenever $R \in \mathcal{PC}$. Furthermore, for $R \in \mathcal{PC}$, *if we "know" the time complexity of algorithm A* (e.g., if we can compute $T_{A|S_R}(n)$ in poly(n)-time), then we may modify A into an algorithm A' that solves the (general) search problem of R (i.e., halts with a correct output on each input) in time $T_{A'}(n) = T_{A|S_R}(n) + \text{poly}(n)$. However, as we shall see in Section 2.4.2, the naive assumption by which we always know the running-time of an algorithm that we design is not necessarily valid.

Decision problems with a promise. In the context of decision problems, a promise problem is a relaxation in which one is only required to determine the status of instances that belong to a predetermined set, called the promise. The requirement of efficient verification is adapted in an analogous manner. In view of the standard usage of the term, we refer to *decision problems with a promise* by the name *promise problems*. Formally, promise problems refer to a three-way partition of the set of all strings into yes-instances, no-instances and instances that violate the promise. Standard decision problems are obtained as a special case by insisting that all inputs are allowed (i.e., the promise is trivial).

Definition 2.30 (promise problems): *A* promise problem *consists of a pair of non-intersecting sets of strings, denoted $(S_{\rm yes}, S_{\rm no})$, and $S_{\rm yes} \cup S_{\rm no}$ is called the* promise.

- *The promise problem $(S_{\rm yes}, S_{\rm no})$ is* solved by algorithm *A if for every $x \in S_{\rm yes}$ it holds that $A(x) = 1$ and for every $x \in S_{\rm no}$ it holds that $A(x) = 0$. The promise problem is in the* promise problem extension of *$\mathcal{P}$ if there exists a polynomial-time algorithm that solves it.*
- *The promise problem $(S_{\rm yes}, S_{\rm no})$ is in the* promise problem extension of *$\mathcal{NP}$ if there exists a polynomial p and a polynomial-time algorithm V such that the following two conditions hold:*

[22]Here we refer to the formulation presented in Section 2.1.6.

1. Completeness: *For every $x \in S_{\rm yes}$, there exists y of length at most $p(|x|)$ such that $V(x,y) = 1$.*
2. Soundness: *For every $x \in S_{\rm no}$ and every y, it holds that $V(x,y) = 0$.*

We stress that for algorithms of polynomial-time complexity, it does not matter whether the time complexity is defined only on inputs that satisfy the promise or on all strings (see Footnote 21). Thus, the extended classes $\mathcal{P}$ and $\mathcal{NP}$ (like $\mathcal{PF}$ and $\mathcal{PC}$) are invariant under this choice.

Reducibility among promise problems. The notion of a Cook-reduction extend naturally to promise problems, when postulating that a query that violates the promise (of the problem at the target of the reduction) may be answered arbitrarily.[23] That is, the oracle machine should solve the original problem no matter how queries that violate the promise are answered. The latter requirement is consistent with the conceptual meaning of reductions and promise problems. Recall that reductions captures procedures that make subroutine calls to an arbitrary procedure that solves the reduced problem. But, in the case of promise problems, such a solver may behave arbitrarily on instances that violate the promise. We stress that the main property of a reduction is preserved (see Exercise 2.35): *if the promise problem Π is Cook-reducible to a promise problem that is solvable in polynomial-time, then Π is solvable in polynomial-time.*

We warn that the extension of a complexity class to promise problems does not necessarily inherit the "structural" properties of the standard class. For example, in contrast to Theorem 2.33, there exists promise problems in $\mathcal{NP} \cap {\rm co}\mathcal{NP}$ such that every set in $\mathcal{NP}$ can be Cook-reduced to them: see Exercise 2.36. Needless to say, $\mathcal{NP} = {\rm co}\mathcal{NP}$ does not seem to follow from Exercise 2.36.

2.4.1.2 Discussion

The following discussion refers both to the decision and search versions of promise problems. Recall that promise problems offer the most direct way of capturing natural computational problems (e.g., when referring to computational problems regarding graphs the promise mandates that the input is a graph).

Restricting a computational problem. In addition to the foregoing motivation to promise problems, we mention their use in formulating the natural notion of a *restriction of a computational problem to a subset of the instances.* Specifically, such a restriction means that the promise set of the restricted problem is a subset of the promise set of the unrestricted problem. For example, when we say that 3SAT is a restriction of SAT, we refer to the fact that the set of allowed instances is now restricted to 3CNF formulae (rather than to arbitrary CNF formulae). In both

[23]It follows that Karp-reductions among promise problems are not allowed to make queries that violate the promise. Specifically, we say that the promise problem $\Pi = (\Pi_{\rm yes}, \Pi_{\rm no})$ is Karp-reducible to the promise problem $\Pi' = (\Pi'_{\rm yes}, \Pi'_{\rm no})$ if there exists a polynomial-time mapping f such that for every $x \in \Pi_{\rm yes}$ (resp., $x \in \Pi_{\rm no}$) it holds that $f(x) \in \Pi'_{\rm yes}$ (resp., $f(x) \in \Pi'_{\rm no}$).

cases, the natural computational problem is to determine satisfiability (or to find a satisfying assignment), but the set of instances (i.e., the promise set) is further restricted in the case of 3SAT. The fact that a restricted problem is never harder than the original problem is captured by the fact that the restricted problem is reducible to the original one (via the identity mapping).

The standard convention of avoiding promise problems. Recall that, although promise problems provide a good framework for presenting natural computational problems, we managed to avoid this formulation in previous sections. This was done by relying on the fact that for all the (natural) problems considered in the previous sections, it is easy to decide whether or not a given instance satisfies the promise. For example, given a formula it is easy to decide whether or not it is in CNF (or 3CNF). Actually, the issue arises already when talking about formulae: What we are actually given is a string that is supposed to encode a formula (under some predetermined encoding scheme), and so the promise (which is easy to decide for natural encoding schemes) is that the input string is a valid encoding of some formula. In any case, if the promise is efficiently recognizable (i.e., membership in it can be decided in polynomial-time) then we may avoid mentioning the promise by using one of the following two "nasty" conventions:

1. *Extending the set of instances to the set of all possible strings* (and allowing trivial solutions for the corresponding dummy instances). For example, in the case of a search problem, we may either define all instance that violate the promise to have no solution or define them to have a trivial solution (e.g., be a solution for themselves); that is, for a search problem R with promise P, we may consider the (standard) search problem of R where R is modified such that $R(x) = \emptyset$ for every $x \notin P$ (or, say, $R(x) = \{x\}$ for every $x \notin P$). In the case of a promise (decision) problem $(S_{\rm yes}, S_{\rm no})$, we may consider the problem of deciding membership in $S_{\rm yes}$, which means that instances that violate the promise are considered as no-instances.

2. *Considering every string as a valid encoding of an object that satisfies the promise.* That is, fixing any string x_0 that satisfies the promise, we consider every string that violates the promise as if it were x_0. In the case of a search problem R with promise P, this means considering the (standard) search problem of R where R is modified such that $R(x) = R(x_0)$ for every $x \notin P$. Similarly, in the case of a promise (decision) problem $(S_{\rm yes}, S_{\rm no})$, we consider the problem of deciding membership in $S_{\rm yes}$ (provided $x_0 \in S_{\rm no}$ and otherwise we consider the problem of deciding membership in $\{0,1\}^* \setminus S_{\rm no}$).

We stress that *in the case that the promise is efficiently recognizable* the aforementioned conventions (or modifications) do not effect the complexity of the relevant (search or decision) problem. That is, rather that considering the original promise problem, we consider a (search or decision) problem (without a promise) that is computational equivalent to the original one. Thus, in some sense we loss nothing by studying the latter problem rather than the original one. On the other hand, even in the case that these two problems are computationally equivalent, it is useful to have a formulation that allows to distinguish between them (as we do distinguish between the different NP-complete problems although they are all computationally equivalent). This conceptual concern becomes of crucial importance in the case (to be discussed next) that the promise is *not* efficiently recognizable.

The foregoing transformations of promise problems into computationally equivalent standard (decision and search) problems does not necessarily preserve the complexity of the problem in the case that the promise is not efficiently recognizable. In this case, the terminology of promise problems is unavoidable. Consider, for example, the problem of deciding whether a Hamiltonian graph is 3-colorable. On the face of it, such a problem may have fundamentally different complexity than the problem of deciding whether a given graph is both Hamiltonian and 3-colorable.

The notion of a promise problem provides an adequate formulation for a variety of computational complexity notions and results. Examples include the notion of "unique solutions" (see Section 6.2.3) and the formulation of "gap problems" as capturing various approximation tasks (see Section 10.1).

2.4.1.3 The common convention

In spite of the foregoing opinions, we adopt the common convention of focusing on standard decision and search problems. That is, by default, all complexity classes refer to standard decision and search problems, and the exceptions in which we refer to promise problems are stated explicitly as such. Such exceptions appear in Sections 2.4.2, 6.1.2, 6.2.3, and 10.1.

2.4.2 Optimal search algorithms for NP

We refer to the candid search problem of any relation in $\mathcal{PC}$. Recall that $\mathcal{PC}$ is the class of search problems that allow for efficient checking of the correctness of candidate solutions (see Definition 2.3), and that the candid search problem is a search problem in which the solver is promised that the given instance has a solution (see Definition 2.29).

We claim the existence of an *optimal algorithm for solving the candid search problem of any relation in* $\mathcal{PC}$. Furthermore, we will explicitly present such an algorithm, and prove that it is optimal in a very strong sense: for any algorithm solving the candid search problem of $R \in \mathcal{PC}$, our algorithm solves the same problem in time that is at most a constant factor slower (ignoring a fixed additive polynomial term, which may be disregarded in the case that the problem is not solvable in polynomial-time). Needless to say, we do not know the time complexity of the aforementioned optimal algorithm (indeed if we knew it then we would have resolved the P-vs-NP Question). In fact, the P-vs-NP Question boils down to determining the time complexity of a single explicitly presented algorithm (i.e., the optimal algorithm claimed in Theorem 2.31).

Theorem 2.31 *For every binary relation $R \in \mathcal{PC}$ there exists an algorithm A that satisfies the following:*

1. *A solves the candid search problem of R.*

2. *There exists a polynomial p such that for every algorithm A' that solves the candid search problem of R and for every $x \in S_R \overset{\text{def}}{=} \{x : R(x) \neq \emptyset\}$ it holds that $t_A(x) = O(t_{A'}(x) + p(|x|))$, where t_A* (resp., *$t_{A'}$*) *denotes the number of steps taken by A* (resp., *A'*) *on input x.*

Interestingly, we establish the optimality of A without knowing what its (optimal) running-time is. Furthermore, the optimality claim is "point-wise" (i.e., it refers to any input) rather than "global" (i.e., referring to the (worst case) time complexity as a function of the input length).

We stress that the hidden constant in the O-notation depends only on A', but in the following proof the dependence is exponential in the length of the description of algorithm A' (and it is not known whether a better dependence can be achieved). Indeed, this dependence as well as the idea underlying it constitute one negative aspect of this otherwise amazing result. Another negative aspect is that the optimality of algorithm A refers only to inputs that have a solution (i.e., inputs in S_R). Finally, we note that the theorem as stated refers only to models of computation that have machines that can emulate a given number of steps of other machines with a constant overhead. We mention that in most natural models the overhead of such emulation is at most poly-logarithmic in the number of steps, in which case it holds that $t_A(x) = \widetilde{O}(t_{A'}(x) + p(|x|))$.

Proof Sketch: Fixing R, we let M be a polynomial-time algorithm that decides membership in R, and let p be a polynomial bounding the running-time of M (as a function of the length of the first element in the input pair). We present the following algorithm A that merely emulates all possible search algorithms "in parallel" and checks the result provided by each of them (using M), halting whenever it obtains a correct solution.

Since there are infinitely many possible algorithms, it may not be clear what we mean by the expression "emulating all possible algorithms in parallel." What we mean is emulating them at different "rates" such that the infinite sum of these rates converges to 1 (or to any other constant). Specifically, we will emulate the i^{th} possible algorithm at rate $1/(i+1)^2$, which means emulating a single step of this algorithm per $(i+1)^2$ emulation steps (performed for all algorithms). Note that a straightforward implementation of this idea may create a significant overhead, involved in switching frequently from the emulation of one machine to the emulation of another. Instead, we present an alternative implementation that proceeds in iterations.

In the j^{th} iteration, for $i = 1, ..., 2^{j/2} - 1$, algorithm A emulates $2^j/(i+1)^2$ steps of the i^{th} machine (where the machines are ordered according to the lexicographic order of their descriptions). Each of these emulations is conducted in one chunk, and thus the overhead of switching between the various emulations is insignificant (in comparison to the total number of steps being emulated). In the case that some of these emulations halts with output y, algorithm A invokes M on input (x, y) and output y if and only if $M(x, y) = 1$. Furthermore, the verification of a solution provided by a candidate algorithm is also emulated at the expense of its step-count. (Put in other words, we augment each algorithm with a canonical procedure (i.e.,

M) that checks the validity of the solution offered by the algorithm.)

By its construction, whenever $A(x)$ outputs a string y (i.e., $y \neq \perp$) it must hold that $(x, y) \in R$. To show the optimality of A, we consider an arbitrary algorithm A' that solves the candid search problem of R. Our aim is to show that A is not much slower than A'. Intuitively, this is the case because the overhead of A results from emulating other algorithms (in addition to A'), but the total number of emulation steps wasted (due to these algorithms) is inversely proportional to the rate of algorithm A', which in turn is exponentially related to the length of the description of A'. The punch-line is that since A' is fixed, the length of its description is a constant. Details follow.

For every x, let us denote by $t'(x)$ the number of steps taken by A' on input x, where $t'(x)$ also accounts for the running time of $M(x, \cdot)$; that is, $t'(x) = t_{A'}(x) + p(|x|)$, where $t_{A'}(x)$ is the number of steps taken by $A'(x)$. Then, the emulation of $t'(x)$ steps of A' on input x is "covered" by the j^{th} iteration of A, provided that $2^j/(2^{|A'|+1})^2 \geq t'(x)$ where $|A'|$ denotes the length of the description of A'. (Indeed, we use the fact that the algorithms are emulated in lexicographic order, and note that there are at most $2^{|A'|+1} - 2$ algorithms that precede A' in lexicographic order.) Thus, on input x, algorithm A halts after at most $j_{A'}(x)$ iterations, where $j_{A'}(x) = 2(|A'| + 1) + \log_2(t_{A'}(x) + p(|x|))$, after emulating a total number of steps that is at most

$$t(x) \stackrel{\text{def}}{=} \sum_{j=1}^{j_{A'}(x)} \sum_{i=1}^{2^{j/2}-1} \frac{2^j}{(i+1)^2} < 2^{j_{A'}(x)+1} = 2^{2|A'|+3} \cdot (t_{A'}(x) + p(|x|)).$$

The question of how much time is required for emulating these many steps depends on the specific model of computation. In many models of computation, the emulation of t steps of one machine by another machine requires $\widetilde{O}(t)$ steps of the emulating machines, and in some models this emulation can even be performed with constant overhead. The theorem follows. $\square$

Comment: By construction, the foregoing algorithm A does not halt on input $x \notin S_R$. This can be easily rectified by letting A emulate a straightforward exhaustive search for a solution, and halt with output $\perp$ if this this exhaustive search indicates that there is no solution to the current input. This extra emulation can be performed in parallel to all other emulations (e.g., at a rate of one step for the extra emulation per each step of everything else).

2.4.3 The class coNP and its intersection with NP

By prepending the name of a complexity class (of decision problems) with the prefix "co" we mean the class of complement sets; that is,

$$\text{co}\mathcal{C} \stackrel{\text{def}}{=} \{\{0,1\}^* \setminus S : S \in \mathcal{C}\}$$

Specifically, $\text{co}\mathcal{NP} = \{\{0,1\}^* \setminus S : S \in \mathcal{NP}\}$ is the class of sets that are complements of sets in $\mathcal{NP}$.

Recalling that sets in $\mathcal{NP}$ are characterized by their witness relations such that $x \in S$ if and only if there exists an adequate NP-witness, it follows that their complement sets consists of all instances for which there are no NP-witness (i.e., $x \in \{0,1\}^* \setminus S$ if there is no NP-witness for x being in S). For example, $\texttt{SAT} \in \mathcal{NP}$ implies that the set of unsatisfiable CNF formulae is in co$\mathcal{NP}$. Likewise, the set of graphs that are not 3-colorable is in co$\mathcal{NP}$. (Jumping ahead, we mention that it is widely believed that these sets are not in $\mathcal{NP}$.)

Another perspective on co$\mathcal{NP}$ is obtained by considering the search problems in $\mathcal{PC}$. Recall that for such $R \in \mathcal{PC}$, the set of instances having a solution (i.e., $S_R = \{x : \exists y \text{ s.t. } (x,y) \in R\}$) is in $\mathcal{NP}$. It follows that the set of instances having no solution (i.e., $\{0,1\}^* \setminus S_R = \{x : \forall y\ (x,y) \notin R\}$) is in co$\mathcal{NP}$.

It is widely believed that $\mathcal{NP} \neq$ co$\mathcal{NP}$ (which means that $\mathcal{NP}$ is not closed under complementation). Indeed, this conjecture implies $\mathcal{P} \neq \mathcal{NP}$ (because $\mathcal{P}$ is closed under complementation). The conjecture $\mathcal{NP} \neq$ co$\mathcal{NP}$ means that some sets in co$\mathcal{NP}$ do not have NP-proof systems (because $\mathcal{NP}$ is the class of sets having NP-proof systems). As we will show next, under this conjecture, the complements of NP-complete sets do not have NP-proof systems; for example, there exists no NP-proof system for proving that a given CNF formula is not satisfiable. We first establish this fact for NP-completeness in the standard sense (i.e., under Karp-reductions, as in Definition 2.16).

Proposition 2.32 *Suppose that $\mathcal{NP} \neq$ co$\mathcal{NP}$ and let $S \in \mathcal{NP}$ such that every set in $\mathcal{NP}$ is Karp-reducible to S. Then $\overline{S} \stackrel{\text{def}}{=} \{0,1\}^* \setminus S$ is not in $\mathcal{NP}$.*

Proof Sketch: We first observe that the fact that every set in $\mathcal{NP}$ is Karp-reducible to S implies that every set in co$\mathcal{NP}$ is Karp-reducible to $\overline{S}$. We next claim that *if S' is in $\mathcal{NP}$ then every set that is Karp-reducible to S' is also in $\mathcal{NP}$.* Applying the claim to $S' = \overline{S}$, we conclude that $\overline{S} \in \mathcal{NP}$ implies co$\mathcal{NP} \subseteq \mathcal{NP}$, which in turn implies $\mathcal{NP} =$ co$\mathcal{NP}$ in contradiction to the main hypothesis.

We now turn to prove the foregoing claim; that is, we prove that if S' has an NP-proof system and S'' is Karp-reducible to S' then S'' has an NP-proof system. Let V' be the verification procedure associated with S', and let f be a Karp-reduction of S'' to S'. Then, we define the verification procedure V'' (for membership in S'') by $V''(x,y) = V'(f(x),y)$. That is, any NP-witness that $f(x) \in S'$ serves as an NP-witness for $x \in S''$ (and these are the only NP-witnesses for $x \in S''$). This may not be a "natural" proof system (for S''), but it is definitely an NP-proof system for S''. $\square$

Assuming that $\mathcal{NP} \neq$ co$\mathcal{NP}$, Proposition 2.32 implies that sets in $\mathcal{NP} \cap$ co$\mathcal{NP}$ cannot be NP-complete with respect to Karp-reductions. In light of other limitations of Karp-reductions (see, e.g., Exercise 2.7), one may wonder whether or not the exclusion of NP-complete sets from the class $\mathcal{NP} \cap$ co$\mathcal{NP}$ is due to the use of a restricted notion of reductions (i.e., Karp-reductions). The following theorem asserts that this is not the case: *some sets in $\mathcal{NP}$ cannot be reduced to sets in the intersection $\mathcal{NP} \cap$ co$\mathcal{NP}$ even under general reductions* (i.e., Cook-reductions).

Theorem 2.33 *If every set in $\mathcal{NP}$ can be Cook-reduced to some set in $\mathcal{NP} \cap \text{co}\mathcal{NP}$ then $\mathcal{NP} = \text{co}\mathcal{NP}$.*

In particular, assuming $\mathcal{NP} \neq \text{co}\mathcal{NP}$, no set in $\mathcal{NP} \cap \text{co}\mathcal{NP}$ can be NP-complete, even when NP-completeness is defined with respect to Cook-reductions. Since $\mathcal{NP} \cap \text{co}\mathcal{NP}$ is conjectured to be a proper superset of $\mathcal{P}$, it follows (assuming $\mathcal{NP} \neq \text{co}\mathcal{NP}$) that there are decision problems in $\mathcal{NP}$ that are neither in $\mathcal{P}$ nor NP-hard (i.e., specifically, the decision problems in $(\mathcal{NP} \cap \text{co}\mathcal{NP}) \setminus \mathcal{P}$). We stress that Theorem 2.33 refers to standard decision problems and not to promise problems (see Section 2.4.1 and Exercise 2.36).

Proof: Analogously to the proof of Proposition 2.32 , the current proof boils down to proving that *if S is Cook-reducible to a set in $\mathcal{NP} \cap \text{co}\mathcal{NP}$ then $S \in \mathcal{NP} \cap \text{co}\mathcal{NP}$.* Using this claim, the theorem's hypothesis implies that $\mathcal{NP} \subseteq \mathcal{NP} \cap \text{co}\mathcal{NP}$, which in turn implies $\mathcal{NP} \subseteq \text{co}\mathcal{NP}$ and $\mathcal{NP} = \text{co}\mathcal{NP}$.

Fixing any S and $S' \in \mathcal{NP} \cap \text{co}\mathcal{NP}$ such that S is Cook-reducible to S', we prove that $S \in \mathcal{NP}$ (and the proof that $S \in \text{co}\mathcal{NP}$ is similar).[24] Let us denote by M the oracle machine reducing S to S'. That is, on input x, machine M makes queries and decides whether or not to accept x, and its decision is correct provided all queries are answered according to S'. To show that $S \in \mathcal{NP}$, we will present an NP-proof system for S. This proof system (or rather its verification procedure), denoted V, accepts a pair of the form $(x, ((z_1, \sigma_1, w_1), ..., (z_t, \sigma_t, w_t))$ if the following two conditions hold:

1. On input x, machine M accepts after making the queries $z_1, ..., z_t$, and obtaining the corresponding answers $\sigma_1, ..., \sigma_t$.

   That is, V check that, on input x, after obtaining the answers $\sigma_1, ..., \sigma_{i-1}$ to the first $i-1$ queries, the i^{th} query made by M equals z_i. In addition, V checks that M outputs 1 (indicating acceptance), while making the queries $z_1, ..., z_t$ and receiving the answers $\sigma_1, ..., \sigma_t$, respectively.

2. For every i, it holds that if $\sigma_i = 1$ then w_i is an NP-witness for $z_i \in S'$, whereas if $\sigma_i = 0$ then w_i is an NP-witness for $z_i \in \{0,1\}^* \setminus S'$.

   Thus, if this condition holds then it is the case that each σ_i indicates the correct status of z_i with respect to S' (i.e., $\sigma_i = 1$ if and only if $z_i \in S'$).

We stress that we use the fact that both S' and $\overline{S}' \stackrel{\text{def}}{=} \{0,1\}^* \setminus S$ have NP-proof systems, and refer to the corresponding NP-witnesses.

Note that V is indeed an NP-proof system for S. Firstly, the length of the corresponding witnesses is bounded by the running-time of the reduction (and the length of the NP-witnesses supplied for the various queries). Next note that V runs in polynomial time (i.e., verifying the first condition requires an emulation of the polynomial-time execution of M on input x when using the σ_i's to emulate the

[24]Alternatively, we show that $S \in \text{co}\mathcal{NP}$ by applying the following argument to $\overline{S} \stackrel{\text{def}}{=} \{0,1\}^* \setminus S$ and noting that $\overline{S}$ is Cook-reducible to S' (via S, or alternatively that $\overline{S}$ is Cook-reducible to $\{0,1\}^* \setminus S' \in \mathcal{NP} \cap \text{co}\mathcal{NP}$).

oracle, whereas verifying the second condition is done by invoking the relevant NP-proof systems). Finally, observe that $x \in S$ if and only if there exists a sequence $y \stackrel{\text{def}}{=} ((z_1, \sigma_1, w_1), ..., (z_t, \sigma_t, w_t))$ such that $V(x, y) = 1$. In particular, $V(x, y) = 1$ holds only if y contains a valid sequence of queries and answers made by $M(x)$ and answered by the oracle S', and M accepts based on that sequence. ■

The world view – a digest. Recall that on top of the $\mathcal{P} \neq \mathcal{NP}$ conjecture, we mentioned two other conjectures (which clearly imply $\mathcal{P} \neq \mathcal{NP}$):

1. The conjecture that $\mathcal{NP} \neq \text{co}\mathcal{NP}$ (equivalently, $\mathcal{NP} \cap \text{co}\mathcal{NP} \neq \mathcal{NP}$).

   This conjecture is equivalent to the conjecture that CNF formulae have no short proofs of unsatisfiability (i.e., the set $\{0,1\}^* \setminus \texttt{SAT}$ has no NP-proof system).

2. The conjecture that $\mathcal{NP} \cap \text{co}\mathcal{NP} \neq \mathcal{P}$.

   Notable candidates for the class $\mathcal{NP} \cap \text{co}\mathcal{NP} \neq \mathcal{P}$ include decision problems that are computationally equivalent to the integer factorization problem (i.e., the search problem (in $\mathcal{PC}$) in which, given a composite number, the task is to find its prime factors).

Combining these conjectures, we get the world view depicted in Figure 2.5, which also shows the class of co$\mathcal{NP}$-complete sets (defined next).

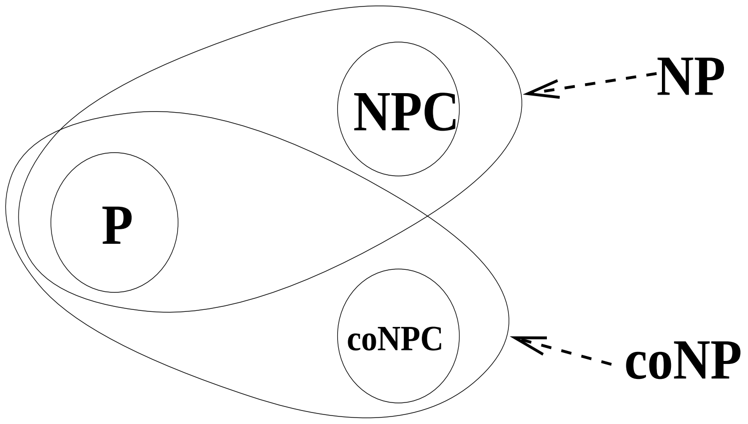


Figure 2.5: The world view under $\mathcal{P} \neq \text{co}\mathcal{NP} \cap \mathcal{NP} \neq \mathcal{NP}$.

Definition 2.34 *A set S is called* co$\mathcal{NP}$-hard *if every set in* co$\mathcal{NP}$ *is Karp-reducible to S. A set is called* co$\mathcal{NP}$-complete *if it is both in* co$\mathcal{NP}$ *and* co$\mathcal{NP}$*-hard.*

Indeed, insisting on Karp-reductions is essential for a distinction between $\mathcal{NP}$-hardness and co$\mathcal{NP}$-hardness.

Chapter Notes

Many sources provide historical accounts of the developments that led to the formulation of the *P vs NP Problem* and to the discovery of the theory of NP-completeness (see, e.g., [81, Sec. 1.5] and [210]). Still, we feel that we should not refrain from offering our own impressions, which are based on the texts of the original papers.

Nowadays, the theory of NP-completeness is commonly attributed to Cook [55], Karp [130], and Levin [145]. It seems that Cook's starting point was his interest in theorem proving procedures for propositional calculus [55, P. 151]. Trying to provide evidence to the difficulty of deciding whether or not a given formula is a tautology, he identified $\mathcal{NP}$ as a class containing "many apparently difficult problems" (cf, e.g., [55, P. 151]), and showed that any problem in $\mathcal{NP}$ is reducible to deciding membership in the set of 3DNF tautologies. In particular, Cook emphasized the importance of the concept of polynomial-time reductions and the complexity class $\mathcal{NP}$ (both explicitly defined for the first time in his paper). He also showed that `CLIQUE` is computationally equivalent to `SAT`, and envisioned a class of problems of the same nature.

Karp's paper [130] can be viewed as fulfilling Cook's prophecy: Stimulated by Cook's work, Karp demonstrated that a "large number of classic difficult computational problems, arising in fields such as mathematical programming, graph theory, combinatorics, computational logic and switching theory, are [NP-]complete (and thus equivalent)" [130, P. 86]. Specifically, his list of twenty-one NP-complete problems includes Integer Linear Programming, Hamilton Circuit, Chromatic Number, Exact Set Cover, Steiner Tree, Knapsack, Job Scheduling, and Max Cut. Interestingly, Karp defined $\mathcal{NP}$ in terms of verification procedures (i.e., Definition 2.5), pointed to its relation to "backtrack search of polynomial bounded depth", and viewed $\mathcal{NP}$ as the residence of a "wide range of important computational problems" (which are not in $\mathcal{P}$).

Independently of these developments, while being in the USSR, Levin proved the existence of "universal search problems" (where universality meant NP-completeness). The starting point of Levin's work [145] was his interest in the "perebor" conjecture asserting the inherent need for brute-force in some search problems that have efficiently checkable solutions (i.e., problems in $\mathcal{PC}$). Levin emphasized the implication of polynomial-time reductions on the relation between the time complexity of the related problem (for any growth rate of the time complexity), asserted the NP-completeness of six "classical search problems", and claimed that the underlying method "provides a mean for readily obtaining" similar results for "many other important search problems."

It is interesting to note that although the works of Cook [55], Karp [130], and Levin [145] were received with different levels of enthusiasm, none of the contemporaries realized the depth of the discovery and the difficulty of the question posed (i.e., the P-vs-NP Question). This fact is evident in every account from the early 1970's, and may explain the frustration of the corresponding generation of researchers, which expected the P-vs-NP Question to be resolved in their life-time (if not in a matter of years). Needless to say, the author's opinion is that there was absolutely no justification for these expectations, and that one should have actually expected quite the opposite.

We mention that the three "founding papers" of the theory of NP-completeness (i.e., Cook [55], Karp [130], and Levin [145]) use the three different types of reductions used in this chapter. Specifically, Cook uses the general notion of polynomial-time reduction [55], often referred to as Cook-reductions (Definition 2.9). The notion of Karp-reductions (Definition 2.10) originates from Karp's paper [130], whereas its augmentation to search problems (i.e., Definition 2.11) originates from Levin's paper [145]. It is worth noting that unlike Cook and Karp's works, which treat decision problems, Levin's work is stated in terms of search problems.

The reductions presented in §2.3.3.2 are not necessarily the original ones. Most notably, the reduction establishing the NP-hardness of the Independent Set problem (i.e., Proposition 2.25) is adapted from [70] (see also Exercise 9.14). In contrast, the reductions presented in §2.3.3.1 are merely a re-interpretation of the original reduction as presented in [55]. The equivalence of the two definitions of $\mathcal{NP}$ (i.e., Theorem 2.8) was proved in [130].

The existence of NP-sets that are neither in P nor NP-complete (i.e., Theorem 2.27) was proven by Ladner [142], Theorem 2.33 was proven by Selman [188], and the existence of optimal search algorithms for NP-relations (i.e., Theorem 2.31) was proven by Levin [145]. (Interestingly, the latter result was proved in the same paper in which Levin presented the discovery of NP-completeness, independently of Cook and Karp.) Promise problems were explicitly introduced by Even, Selman and Yacobi [68]; see [91] for a survey of their numerous applications.

We mention that the standard reductions used to establish natural NP-completeness results have several additional properties or can be modified to have such properties. These properties include an efficient transformation of solutions in the direction of the reduction (see Exercise 2.28), the preservation of the number of solutions (see Exercise 2.29), being computable by a log-space algorithm (see Section 5.2.2), and being invertible in polynomial-time (see Exercise 2.30).

Exercises

Exercise 2.1 ($\mathcal{PF}$ contains problems that are not in $\mathcal{PC}$) Show that $\mathcal{PF}$ contains some (unnatural) problems that are not in $\mathcal{PC}$.

Guideline: Consider the relation $R = \{(x, 1) : x \in \{0,1\}^*\} \cup \{(x, 0) : x \in S\}$, where S is some undecidable set. Note that R is the disjoint union of two binary relations, denoted R_1 and R_2, where R_1 is in $\mathcal{PF}$ whereas R_2 is not in $\mathcal{PC}$. Furthermore, for every x it holds that $R_1(x) \neq \emptyset$.

Exercise 2.2 Show that any $S \in \mathcal{NP}$ has many different NP-proof systems (i.e., verification procedures $V_1, V_2, \ldots$ such that $V_i(x, y) = 1$ does not imply $V_j(x, y) = 1$ for $i \neq j$).

Guideline: For V and p be as in Definition 2.5, define $V_i(x, y) = 1$ if $|y| = p(|x|) + i$ and there exists a prefix y' of y such that $V(x, y') = 1$.

Exercise 2.3 Relying on the fact that primality is decidable in polynomial-time and assuming that there is no polynomial-time factorization algorithm, present two "natural but fundamentally different" NP-proof systems for the set of composite numbers.

Guideline: Consider the following verification procedures V_1 and V_2 for the set of composite numbers. Let $V_1(n, y) = 1$ if and only if $y = n$ and n is not a prime, and $V_2(n, m) = 1$ if and only if m is a non-trivial divisor of n. Show that valid proofs with respect to V_1 are easy to find, whereas valid proofs with respect to V_2 are hard to find.

Exercise 2.4 Regarding Definition 2.7, show that if S is accepted by some non-deterministic machine of time complexity t then it is accepted by a non-deterministic machine of time complexity $O(t)$ that has a transition function that maps each possible symbol-state pair to exactly two triples.

Exercise 2.5 Verify the following properties of Cook-reductions:

1. If Π is Cook-reducible to Π' and Π' is solvable in polynomial-time then so is Π.
2. Cook-reductions are transitive (i.e., if Π is Cook-reducible to Π' and Π' is Cook-reducible to Π'' then Π is Cook-reducible to Π'').
3. If Π is solvable in polynomial-time then it is Cook-reducible to any problem Π'.

In continuation to the last item, show that a problem Π is solvable in polynomial-time if and only if it is Cook-reducible to a trivial problem (e.g., deciding membership in the empty set).

Exercise 2.6 Show that Karp-reductions (and Levin-reductions) are transitive.

Exercise 2.7 Show that some decision problems are not Karp-reducible to their complement (e.g., the empty set is not Karp-reducible to $\{0,1\}^*$).
A popular exercise of dubious nature is showing that any decision problem in $\mathcal{P}$ is Karp-reducible to any *non-trivial* decision problem, where the decision problem regarding a set S is called non-trivial if $S \neq \emptyset$ and $S \neq \{0,1\}^*$. It follows that every non-trivial set in $\mathcal{P}$ is Karp-reducible to its complement.

Exercise 2.8 (reducing search problems to optimization problems) For every polynomially bounded relation R (resp., $R \in \mathcal{PC}$), present a function f (resp., a polynomial-time computable function f) such that the search problem of R is computationally equivalent to the search problem in which given (x, v) one has to find a $y \in \{0,1\}^{\text{poly}(|x|)}$ such that $f(x, y) \geq v$.
(Hint: use a Boolean function.)

Exercise 2.9 (binary search) Show that using ℓ binary queries of the form "is $z \geq v$" it is possible to determine the value of an integer z that is a priori known to reside in the interval $[0, 2^\ell - 1]$.

Guideline: Consider a process that iteratively halves the interval in which z is known to reside in.

Exercise 2.10 Show that if $R \in \mathcal{PC} \setminus \mathcal{PF}$ is self-reducible then the relevant Cook-reduction makes more than a logarithmic number of queries to S_R. More generally, show that if $R \in \mathcal{PC} \setminus \mathcal{PF}$ is Cook-reducible to any decision problem, then this reduction makes more than a logarithmic number of queries.

Guideline: Note that the oracle answers can be emulated by trying all possibilities, and that the correctness of the output of the oracle machine can be efficiently tested.

Exercise 2.11 Show that the standard search problem of Graph 3-Colorability is self-reducible, where this search problem consists of finding a 3-coloring for a given input graph.
(Hint: Iteratively extend the current prefix of a 3-coloring of the graph by making adequate oracle calls to the decision problem of Graph 3-Colorability. Specifically, encode the question of whether or not $(\chi_1, ..., \chi_t) \in \{1,2,3\}^t$ is a prefix of a 3-coloring of the graph G as a query regarding the 3-colorability of an auxiliary graph G'.)[25]

Exercise 2.12 Show that the standard search problem of Graph Isomorphism is self-reducible, where this search problem consists of finding an isomorphism between a given pair of graphs.
(Hint: Iteratively extend the current prefix of an isomorphism between the two N-vertex graphs by making adequate oracle calls to the decision problem of Graph Isomorphism. Specifically, encode the question of whether or not $(\pi_1, ..., \pi_t) \in [N]^t$ is a prefix of an isomorphism between $G_1 = ([N], E_1)$ and $G_2 = ([N], E_2)$ as a query regarding isomorphism between two auxiliary graphs G'_1 and G'_2.)[26]

Exercise 2.13 (downwards self-reducibility) We say that S is downwards self-reducible if there exists a Cook-reduction of S to itself that only makes queries that are each shorter than the reduction's input (i.e., if on input x the reduction makes the query q then $|q| < |x|$).[27]

1. Show that SAT is downwards self-reducible with respect to a natural encoding of CNF formulae. Note that this encoding should have the propery that instantiating a variable in a formula results in a shorter formula.

   A harder exercise consists of showing that Graph 3-Colorability is downwards self-reducible with respect to some reasonable encoding of graphs. Note that this encoding has to be selected carefully (if it is to work for a process analogous to the one used in Exercise 2.11).

[25]Note that we merely need to check whether G has a 3-coloring in which the equalities and inequalities induced by $(\chi_1, ..., \chi_t)$ hold. This can be done by adequate gadgets (e.g., inequality is enforced by an edge between the corresponding vertices, whereas equality is enforced by an adequate subgraph that includes the relevant vertices as well as auxiliary vertices). For Part 1 of Exercise 2.13, equality is better enforced by combining the two vertices.

[26]This can be done by attaching adequate gadgets to pairs of vertices that we wish to be mapped to one another (by the isomorphism). For example, we may connect the vertices in the i^{th} pair to an auxiliary star consisting of $(N + i)$ vertices.

[27]Note that on some instances the reduction may make no queries at all. (This prevent a possible non-viability of the definition due to very short instances.)

2. Suppose that S is downwards self-reducible *by a reduction that outputs the disjunction of the oracle answers.* (Note that this is the case for SAT.) Show that in this case, S is characterized by a witness relation $R \in \mathcal{PC}$ (i.e., $S = \{x : R(x) \neq \emptyset\}$) that is self-reducible (i.e., the search problem of R is Cook-reducible to S). Needless to say, it follows that $S \in \mathcal{NP}$.

   Guideline: Include $(x_0, \langle x_1, ..., x_t\rangle)$ in R if $x_t \in S \cap \{0,1\}^{O(1)}$ and, for every $i \in \{0, 1, ..., t-1\}$, on input x_i the self-reduction makes a set of queries that contains x_{i+1}. Prove that, indeed, $R \in \mathcal{PC}$ and $S = \{x : R(x) \neq \emptyset\}$.

Note that the notion of downwards self-reducibility may be generalized in some natural ways. For example, we may say that S is downwards self-reducible also in case it is computationally equivalent to some set that is downwards self-reducible (in the foregoing strict sense). Note that Part 2 still holds.

Exercise 2.14 (NP problems that are not self-reducible) Assuming that $\mathcal{P} \neq \mathcal{NP} \cap \text{co}\mathcal{NP}$, show that there exists a search problem R in $\mathcal{PC}$ that is not self-reducible (i.e., the search problem of R is not Cook-reducible to the decision problem S_R implicit in R). Prove that it follows that $S'_R = \{(x, y') : \exists y'' \text{ s.t. } (x, y'y'') \in R\}$ is not Cook-reducible to $S_R = \{x : \exists y \text{ s.t. } (x, y) \in R\}$.

Guideline: Given $S \in \mathcal{NP} \cap \text{co}\mathcal{NP} \setminus \mathcal{P}$, present relations $R_1, R_2 \in \mathcal{PC}$ such that $S = \{x : R_1(x) \neq \emptyset\} = \{x : R_2(x) = \emptyset\}$. Then, consider the relation $R = \{(x, 1y) : (x, y) \in R_1\} \cup \{(x, 0y) : (x, y) \in R_2\}$, and prove that $R \notin \mathcal{PF}$ but $S_R = \{0,1\}^*$.

Exercise 2.15 In continuation to Exercise 2.14 and assuming that $\mathcal{P} \neq \mathcal{NP}$, present a search problem R in $\mathcal{PC}$ such that deciding S'_R is not reducible to the search problem of R.

Guideline: Consider the relation $R = \{(x, 0x) : x \in \{0,1\}^*\} \cup \{(x, 1y) : (x, y) \in R'\}$, where R' is an arbitrary relation in $\mathcal{PC} \setminus \mathcal{PF}$, and prove that $R \in \mathcal{PF}$ but $S'_R \notin \mathcal{P}$.

Exercise 2.16 In continuation to Exercise 2.14, present a natural search problem R in $\mathcal{PC}$ such that if factoring integers is intractable then the search problem R (and so also S'_R) is not reducible to S_R.

Guideline: Consider the relation R such that $(N, Q) \in R$ if the integer Q is a non-trivial divisor of the integer N. Use the fact that S_R is in $\mathcal{P}$.

Exercise 2.17 In continuation to Exercises 2.14 and 2.16, show that under suitable assumptions there exists relations $R_1, R_2 \in \mathcal{PC}$ having the same implicit-decision problem (i.e., $\{x : R_1(x) \neq \emptyset\} = \{x : R_2(x) \neq \emptyset\}$) such that R_1 is self-reducible but R_2 is not.

Exercise 2.18 Provide an alternative proof of Theorem 2.15 without referring to the set $S'_R = \{(x, y') : \exists y'' \text{ s.t. } (x, y'y'') \in R\}$. Hint: use Proposition 2.14.

Guideline: Reduce the search problem of R to the search problem of $R_{\texttt{SAT}}$, next reduce $R_{\texttt{SAT}}$ to SAT, and finally reduce SAT to S_R. Justify the existence of each of these three reductions.

Exercise 2.19 Prove that `Bounded Halting` and `Bounded Non-Halting` are NP-complete, where the problems are defined as follows. The instance consists of a pair $(M, 1^t)$, where M is a Turing machine and t is an integer. The decision version of `Bounded Halting` (resp., `Bounded Non-Halting`) consists of determining whether or not there exists an input (of length at most t) on which M halts (resp., does *not* halt) in t steps, whereas the search problem consists of finding such an input.
(Hint: Either modify the proof of Theorem 2.18 or present a reduction of (say) the search problem of $R_{\mathbf{u}}$ to the search problem of Bounded (Non-)Halting. Indeed, the exercise is more straightforward in the case of Bounded Halting.)

Exercise 2.20 In the proof of Theorem 2.20, we claimed that the value of each entry in the "array of configurations" of a machine M is determined by the values of the three entries that reside in the row above it (as in Figure 2.1). Present a function $f_M : \Gamma^3 \to \Gamma$, where $\Gamma = \Sigma \times (Q \cup \{\perp\})$, that substantiates this claim.

Guideline: For example, for every $\sigma_1, \sigma_2, \sigma_3 \in \Sigma$, it holds that $f_M((\sigma_1, \perp), (\sigma_2, \perp), (\sigma_3, \perp)) = (\sigma_2, \perp)$. More interestingly, if the transition function of M maps (σ, q) to $(\tau, p, +1)$ then, for every $\sigma_1, \sigma_2, \sigma_3 \in Q$, it holds that $f_M((\sigma, q), (\sigma_2, \perp), (\sigma_3, \perp)) = (\sigma_2, p)$ and $f_M((\sigma_1, \perp), (\sigma, q), (\sigma_3, \perp)) = (\tau, \perp)$.

Exercise 2.21 Present and analyze a reduction of `SAT` to `3SAT`.

Guideline: For a clause C, consider auxiliary variables such that the i^{th} variable indicates whether one of the first i literals is satisfied, and replace C by a 3CNF that uses the original variables of C as well as the auxiliary variables. For example, the clause $\vee_{i=1}^t x_i$ is replaced by the conjunction of 3CNFs that are logically equivalent to the formulae $(y_2 \equiv (x_1 \vee x_2))$, $(y_i \equiv (y_{i-1} \vee x_i))$ for $i = 3, ..., t$, and y_t. We comment that this is not the standard reduction, but we find it more appealing conceptually.[28]

Exercise 2.22 (efficient solvability of 2SAT) In contrast to Exercise 2.21, prove that `2SAT` (i.e., the satisfiability of 2CNF formulae) is in $\mathcal{P}$.

Guideline: Consider the following "forcing process" for CNF formulae. If the formula contains a singleton clause (i.e., a clause having a single literal), then the corresponding variable is assigned the only value that satisfies the clause, and the formula is simplified accordingly (possibly yielding a constant, which is either `true` or `false`). The process is repeated until the formula is either a constant or contains only 2-literal clauses. Note that a formula ϕ is satisfiable if and only if the formula obtained from ϕ by the forcing process is satisfiable.

1. Prove that a 2CNF formula is unsatisfiable if and only if there exists a variable such that any truth assignment to this variable yields a formula that the forcing process maps to the constant `false`.

   (Extra hint: Applying the forcing process to a 2CNF formula we obtain a sub-formula of it; that is, each clause of the resulting formula is a clause (rather than a sub-clause) of the original formula.)

2. Using Part 1, present a polynomial-time algorithm for solving the search problem of `2SAT`.

[28]The standard reduction replaces the clause $\vee_{i=1}^t x_i$ by the conjunction of the 3CNFs $(x_1 \vee x_2 \vee z_2)$, $((\neg z_{i-1}) \vee x_i \vee z_i)$ for $i = 3, ..., t$, and $\neg z_t$.

Exercise 2.23 (Integer Linear Programming) Prove that the following problem is NP-complete. An instance of the problem is a systems of linear inequalities (say with integer constants), and the problem is to determine whether the system has an integer solution. For example, is there an integer solution to the following system

$$\begin{aligned} x+2y-z &\geq 3 \\ -3x-z &\geq -5 \\ x &\geq 0 \\ -x &\geq -1 \end{aligned}$$

Guideline: Reduce from SAT. Specifically, consider an arithmetization of the input CNF by replacing $\vee$ with addition and $\neg x$ by $1-x$. Thus, each clause gives rize to an inequality (e.g., the clause $x \vee \neg y$ is replaced by the inequality $x+(1-y) \geq 1$, which simplifies to $x-y \geq 2$). Enforce a 0-1 solution by introducing inequalities of the form $x \geq 0$ and $-x \geq -1$, for every variable x.

Exercise 2.24 (Maximum Satisfiability of Linear Systems over GF(2)**)** Prove that the following problem is NP-complete. An instance of the problem consists of a systems of linear equations over GF(2) and an integer k, and the problem is to determine whether there exists an assignment that satisfies at least k equations. (Note that the problem of determining whether there exists an assignment that satisfies all the equations is in $\mathcal{P}$.)

Guideline: Reduce from 3SAT, using an arithetization similar to the one in Exercise 2.23. Specifically, replace each clause that contains $t \leq 3$ literals by 2^t-1 linear GF(2) equations that correspond to the different non-empty subsets of these literals, and assert that their sum (modulo 2) equals one; for example, the clause $x \vee \neg y$ is replaced by the equations $x+(1-y)=1$, $x=1$, and $1-y=1$. Identifying $\{\texttt{false}, \texttt{true}\}$ with $\{0,1\}$, prove that if the original clause is satisfied by a Boolean assignment $\overline{v}$ then exactly 2^{t-1} of the corresponding equations are satisfied by $\overline{v}$, whereas if the original clause is unsatisfied by $\overline{v}$ then none of the corresponding equations is satisfied by $\overline{v}$.

Exercise 2.25 (Satisfiability of Quadratic Systems over GF(2)**)** Prove that the following problem is NP-complete. An instance of the problem consists of a system of quadratic equations over GF(2), and the problem is to determine whether there exists an assignment that satisfies all the equations. Note that the result holds also for systems of quadratic equations over the reals (by adding conditions that enforce a value in $\{0,1\}$).

Guideline: Start by showing that the corresponding problem for cubic equations is NP-complete, by a reduction from 3SAT that maps the clause $x \vee \neg y \vee z$ to the equation $(1-x)\cdot y \cdot (1-z) = 0$. Reduce the problem for cubic equations to the problem for quadaric equations by introducing auxiliary variables; that is, given an instance with variables $x_1, ..., x_n$, introduce the auxiliary variables $x_{i,j}$'s and add equations of the form $x_{i,j} = x_i \cdot x_j$.

Exercise 2.26 (Clique and Independent Set) The instance of the `Independent Set` problem consists of a pair (G, K), where G is a graph and K is an integer, and the question is whether or not the graph G contains an independent set (i.e., a set with no edges between its members) of size (at least) K. The `Clique` problem is analogous. Prove that both problems are computationally equivalent to the `Vertex Cover` problem.

Exercise 2.27 (an alternative proof of Proposition 2.25) Consider the following sketch of a reduction of 3SAT to `Independent Set`. On input a 3CNF formula ϕ with m clauses and n variables, we construct a graph G_ϕ consisting of m triangles (corresponding to the m clauses) augmented with edges that link conflicting literals. That is, if x appears as the $i_1^{\rm th}$ literal of the $j_1^{\rm th}$ clause and $\neg x$ appears as the $i_2^{\rm th}$ literal of the $j_2^{\rm th}$ clause, then we draw an edge between the $i_1^{\rm th}$ vertex of the $j_1^{\rm th}$ triangle and the $i_2^{\rm th}$ vertex of the $j_2^{\rm th}$ triangle. Prove that $\phi \in$ `3SAT` if and only if G_ϕ has an independent set of size m.

Exercise 2.28 (additional properties of standard reductions) In continuation to the discussion in the main text, consider the following augmented form of Karp-reductions. Such a reduction of R to R' consists of three polynomial-time mappings (f, h, g) such that f is a Karp-reduction of S_R to $S_{R'}$ and the following two conditions hold:

1. For every $(x, y) \in R$ it holds that $(f(x), h(x, y)) \in R'$.
2. For every $(f(x), y') \in R'$ it holds that $(x, g(x, y')) \in R$.

(We note that this definition is actually the one used by Levin in [145], except that he restricted h and g to only depend on their second argument.)
Prove that such a reduction implies both a Karp-reduction and a Levin-Reduction, and show that all reductions presented in this chapter satisfy this augmented requirement. Furthermore, prove that in all these cases the main mapping (i.e., f) is 1-1 and polynomial-time invertible.

Exercise 2.29 (parsimonious reductions) Let $R, R' \in \mathcal{PC}$ and let f be a Karp-reduction of $S_R = \{x : R(x) \neq \emptyset\}$ to $S_{R'} = \{x : R'(x) \neq \emptyset\}$. We say that f is parsimonious (with respect to R and R') if for every x it holds that $|R(x)| = |R'(f(x))|$. For each of the reductions presented in this chapter, checked whether or not it is parsimonious. For the reductions that are not parsimonious, find alternative reductions that are parsimonious (cf. [81, Sec. 7.3]).

Exercise 2.30 (on polynomial-time invertible reductions (following [35])) We say that a set S is markable if there exists a polynomial-time (marking) algorithm M such that

1. For every $x, \alpha \in \{0,1\}^*$ it holds that
   (a) $M(x, \alpha) \in S$ if and only if $x \in S$.
   (b) $|M(x, \alpha)| > |x|$.

2. There exists a polynomial-time (de-marking) algorithm D such that, for every $x, \alpha \in \{0,1\}^*$, it holds that $D(M(x,\alpha)) = \alpha$.

Note that all natural NP-sets (e.g., those considered in this chapter) are markable. Prove that *if S' is Karp-reducible to S and S is markable then S' is Karp-reducible to S by a length-increasing, one-to-one, and polynomial-time invertable mapping.*[29] Infer that for any natural NP-complete problem S, any set in $\mathcal{NP}$ is Karp-reducible to S by a length-increasing, one-to-one, and polynomial-time invertable mapping.

Guideline: Let f be a Karp-reduction of S' to S, and let M be the guaranteed marking algorithm. Consider the reduction that maps x to $M(f(x), x)$.

Exercise 2.31 (on the isomorphism of NP-complete sets (following [35])) Suppose that S and T are Karp-reducible to one another by length-increasing, one-to-one, and polynomial-time invertable mappings, denoted f and g respectively. Using the following guidelines, prove that S and T are "effectively" *isomorphic*; that is, present a polynomial-time computable and invertable one-to-one mapping ϕ such that $T = \phi(S) \stackrel{\rm def}{=} \{\phi(x) : x \in S\}$.

1. Let $F \stackrel{\rm def}{=} \{f(x) : x \in \{0,1\}^*\}$ and $G \stackrel{\rm def}{=} \{g(x) : x \in \{0,1\}^*\}$. Using the length-preserving condition of f (resp., g), prove that F (resp., G) is a proper subset of $\{0,1\}^*$. Prove that for every $y \in \{0,1\}^*$ there exists a unique triple $(j,x,i) \in \{1,2\} \times \{0,1\}^* \times \mathbb{N}$ that satisfies one of the following two conditions:

   (a) $j = 1$, $x \in \overline{G} \stackrel{\rm def}{=} \{0,1\}^* \setminus G$, and $y = (g \circ f)^i(x)$;

   (b) $j = 2$, $x \in \overline{F} \stackrel{\rm def}{=} \{0,1\}^* \setminus F$, and $y = (g \circ f)^i(g(x))$.

   (In both cases $i = 0$ is allowed, $h^0(z) = z$, $h^i(z) = h(h^{i-1}(z))$, and $(g \circ f)(z) = g(f(z))$. Hint: starting with y consider the maximal sequence of inverse operations $g^{-1}, f^{-1}, g^{-1}, \ldots$, and note that each inverse shrinks the current string.)

2. Let $U_1 \stackrel{\rm def}{=} \{(g \circ f)^i(x) : x \in \overline{G} \wedge i \geq 0\}$ and $U_2 \stackrel{\rm def}{=} \{(g \circ f)^i(g(x)) : x \in \overline{F} \wedge i \geq 0\}$. Prove that (U_1, U_2) is a partition of $\{0,1\}^*$. Using the fact that f and g are length increasing and polynomial-time invertible, present a polynomial-time procedure for deciding membership in the set U_1.

   Prove the same for the sets $V_1 = \{(f \circ g)^i(x) : x \in \overline{F} \wedge i \geq 0\}$ and $V_2 = \{(f \circ g)^i(f(x)) : x \in \overline{G} \wedge i \geq 0\}$.

3. Note that $U_2 \subseteq G$, and define $\phi(x) \stackrel{\rm def}{=} f(x)$ if $x \in U_1$ and $\phi(x) \stackrel{\rm def}{=} g^{-1}(x)$ otherwise.

   (a) Prove that ϕ is a Karp-reduction of S to T.

[29] When given a string that is not in the image of the mapping, the inverting algorithm returns a special symbol.

(b) Note that ϕ maps U_1 (resp., U_2) to $f(U_1) = \{f(x) : x \in U_1\} = V_2$ (resp., $g^{-1}(U_2) = \{g^{-1}(x) : x \in U_2\} = V_1$). Prove that ϕ is one-to-one and onto.

Observe that $\phi^{-1}(x) = f^{-1}(x)$ if $x \in f(U_1)$ and $\phi^{-1}(x) = g(x)$ otherwise. Prove that ϕ^{-1} is a Karp-reduction of T to S. Infer that $\phi(S) = T$.

Using Exercise 2.30, infer that all natural NP-complete sets are isomorphic.

Exercise 2.32 Prove that a set S is Karp-reducible to some set in $\mathcal{NP}$ if and only if S is in $\mathcal{NP}$.

(Hint: For the non-trivial direction, see the proof of Proposition 2.32.)

Exercise 2.33 Recall that the empty set is not Karp-reducible to $\{0,1\}^*$, whereas any set is Cook-reducible to its complement. Thus our focus here is on the *Karp-reducibility of non-trivial sets to their complements*, where a set is non-trivial if it is neither empty nor contains all strings. Furthermore, since any non-trivial set in $\mathcal{P}$ is Karp-reducible to its complement (see Exercise 2.7), we assume that $\mathcal{P} \neq \mathcal{NP}$ and focus on sets in $\mathcal{NP} \setminus \mathcal{P}$.

1. Prove that $\mathcal{NP} = \text{co}\mathcal{NP}$ implies that some sets in $\mathcal{NP} \setminus \mathcal{P}$ are Karp-reducible to their complements.
2. Prove that $\mathcal{NP} \neq \text{co}\mathcal{NP}$ implies that some sets in $\mathcal{NP} \setminus \mathcal{P}$ are not Karp-reducible to their complements.

(Hint: Use NP-complete sets in both parts, and Exercise 2.32 in the second part.)

Exercise 2.34 Referring to the proof of Theorem 2.27, prove that the function f is unbounded (i.e., for every i there exists an n such that n^3 steps of the process defined in the proof allow for failing the $i + 1^{\text{st}}$ machine).

Guideline: Assume, towards the contradiction that f is bounded. Let $j = \max_{n \in \mathbb{N}}\{f(n)\}$ and n' be the smallest integer such that $f(n') = j$. If j is even then the set F determined by f is co-finite (because $F = \{x : f(|x|) \equiv 0 \pmod 2\} \supseteq \{x : |x| \geq n'\}$). In this case, the j^{th} machine tries to decide $S \cap F$ (which differs from S on finitely many strings), and must fail on some x. Derive a contradiction by showing that the number of steps taken till reaching and considering this x is at most $\exp(\text{poly}(|x|))$, which is smaller than n^3 for some sufficiently large n. A similar argument applies to the case that j is odd, where we use the fact that $F \subseteq \{x : |x| < n'\}$ is finite and so the relevant reduction of S to $S \cap F$ must fail on some input x.

Exercise 2.35 Prove that if the promise problem Π is Cook-reducible to a promise problem that is solvable in polynomial-time, then Π is solvable in polynomial-time. Note that the solver may not halt on inputs that violate the promise.

Guideline: Any polynomial-time algorithm solving any promise problem can be modified such that it halts on all inputs.

Exercise 2.36 (NP-complete promise problems in coNP (following [68])) Consider the promise problem `xSAT`, having instances that are pairs of CNF formulae. The yes-instances consists of pairs (ϕ_1, ϕ_2) such that ϕ_1 is satisfiable and ϕ_2 is unsatisfiable, whereas the no-instances consists of pairs such that ϕ_1 is unsatisfiable and ϕ_2 is satisfiable.

1. Show that `xSAT` is in the intersection of (the promise problem classes that are analogous to) $\mathcal{NP}$ and co$\mathcal{NP}$.
2. Prove that any promise problem in $\mathcal{NP}$ is Cook-reducible to `xSAT`. In designing the reduction, recall that queries that violate the promise may be answered arbitrarily.

   Guideline: Show a reduction of `SAT` to `xSAT`. Specifically, show that the search problem associated with `SAT` is Cook-reducible to `xSAT`, by following the ideas of the proof of Proposition 2.14. Actually, we need a more careful implementation of the search process. Suppose that we know (or assume) that τ is a prefix of a satisfying assignment to ϕ, and we wish to extend τ by one bit. Then, for each $\sigma \in \{0,1\}$, we construct a formula, denoted ϕ'_σ, by setting the first $|\tau|+1$ variables of ϕ according to the values $\tau\sigma$. We query the oracle about the pair (ϕ'_1, ϕ'_0), and extend τ accordingly (i.e., we extend τ by the value 1 if and only if the answer is positive). Note that if both ϕ'_1 and ϕ'_0 are satisfiable then it does not matter which bit we use in the extension, whereas if exactly one formula is satisfiable then the oracle answer is reliable.
3. Pinpoint the source of failure of the proof of Theorem 2.33 when applied to the reduction provided in the previous item.