Covers both C89 and C99

K.N.KING

# C PROGRAMMING

A Modern Approach SECOND EDITION



## 1 Introducing C

When someone says "I want a programming language in which I need only say what I wish done," give him a lollipop.\*

What is C? The simple answer—a widely used programming language developed in the early 1970s at Bell Laboratories—conveys little of C's special flavor. Before we become immersed in the details of the language, let's take a look at where C came from, what it was designed for, and how it has changed over the years (Section 1.1). We'll also discuss C's strengths and weaknesses and see how to get the most out of the language (Section 1.2).

#### 1.1 History of C

Let's take a quick look at C's history, from its origins, to its coming of age as a standardized language, to its influence on recent languages.

#### **Origins**

C is a by-product of the UNIX operating system, which was developed at Bell Laboratories by Ken Thompson, Dennis Ritchie, and others. Thompson single-handedly wrote the original version of UNIX, which ran on the DEC PDP-7 computer, an early minicomputer with only 8K words of main memory (this was 1969, after all!).

Like other operating systems of the time, UNIX was written in assembly language. Programs written in assembly language are usually painful to debug and hard to enhance; UNIX was no exception. Thompson decided that a higher-level

<sup>\*</sup>The epigrams at the beginning of each chapter are from "Epigrams on Programming" by Alan J. Perlis (ACM SIGPLAN Notices (September, 1982): 7–13).

language was needed for the further development of UNIX, so he designed a small language named B. Thompson based B on BCPL, a systems programming language developed in the mid-1960s. BCPL, in turn, traces its ancestry to Algol 60, one of the earliest (and most influential) programming languages.

Ritchie soon joined the UNIX project and began programming in B. In 1970, Bell Labs acquired a PDP-11 for the UNIX project. Once B was up and running on the PDP-11. Thompson rewrote a portion of UNIX in B. By 1971, it became apparent that B was not well-suited to the PDP-11, so Ritchie began to develop an extended version of B. He called his language NB ("New B") at first, and then, as it began to diverge more from B, he changed the name to C. The language was stable enough by 1973 that UNIX could be rewritten in C. The switch to C provided an important benefit: portability. By writing C compilers for other computers at Bell Labs, the team could get UNIX running on those machines as well.

#### Standardization

C continued to evolve during the 1970s, especially between 1977 and 1979. It was during this period that the first book on C appeared. *The C Programming Language*, written by Brian Kernighan and Dennis Ritchie and published in 1978, quickly became the bible of C programmers. In the absence of an official standard for C, this book—known as K&R or the "White Book" to afficionados—served as a de facto standard.

During the 1970s, there were relatively few C programmers, and most of them were UNIX users. By the 1980s, however, C had expanded beyond the narrow confines of the UNIX world. C compilers became available on a variety of machines running under different operating systems. In particular, C began to establish itself on the fast-growing IBM PC platform.

With C's increasing popularity came problems. Programmers who wrote new C compilers relied on K&R as a reference. Unfortunately, K&R was fuzzy about some language features, so compilers often treated these features differently. Also, K&R failed to make a clear distinction between which features belonged to C and which were part of UNIX. To make matters worse, C continued to change after K&R was published, with new features being added and a few older features removed. The need for a thorough, precise, and up-to-date description of the language soon became apparent. Without such a standard, numerous dialects would have arisen, threatening the portability of C programs, one of the language's major strengths.

The development of a U.S. standard for C began in 1983 under the auspices of the American National Standards Institute (ANSI). After many revisions, the standard was completed in 1988 and formally approved in December 1989 as ANSI standard X3.159-1989. In 1990, it was approved by the International Organization for Standardization (ISO) as international standard ISO/IEC 9899:1990. This version of the language is usually referred to as C89 or C90, to distinguish it from the

original version of C, often called K&R C. Appendix C summarizes the major differences between C89 and K&R C.

The language underwent a few changes in 1995 (described in a document known as Amendment 1). More significant changes occurred with the publication of a new standard, ISO/IEC 9899:1999, in 1999. The language described in this standard is commonly known as C99. The terms "ANSI C," "ANSI/ISO C," and "ISO C"—once used to describe C89—are now ambiguous, thanks to the existence of two standards.



Because C99 isn't yet universal, and because of the need to maintain millions (if not billions) of lines of code written in older versions of C. I'll use a special icon (shown in the left margin) to mark discussions of features that were added in C99. A compiler that doesn't recognize these features isn't "C99-compliant." If history is any guide, it will be some years before all C compilers are C99-compliant, if they ever are. Appendix B lists the major differences between C99 and C89.

#### **C-Based Languages**

C has had a huge influence on modern-day programming languages, many of which borrow heavily from it. Of the many C-based languages, several are especially prominent:

- $\blacksquare$  C++ includes all the features of C, but adds classes and other features to support object-oriented programming.
- Java is based on C++ and therefore inherits many C features.
- C# is a more recent language derived from C++ and Java.
- *Perl* was originally a fairly simple scripting language; over time it has grown and adopted many of the features of C.

Considering the popularity of these newer languages, it's logical to ask whether it's worth the trouble to learn C. I think it is, for several reasons. First, learning C can give you greater insight into the features of C++, Java, C#, Perl, and the other C-based languages. Programmers who learn one of these languages first often fail to master basic features that were inherited from C. Second, there are a lot of older C programs around; you may find yourself needing to read and maintain this code. Third, C is still widely used for developing new software, especially in situations where memory or processing power is limited or where the simplicity of C is desired.

If you haven't already used one of the newer C-based languages, you'll find that this book is excellent preparation for learning these languages. It emphasizes data abstraction, information hiding, and other principles that play a large role in object-oriented programming. C++ includes all the features of C, so you'll be able to use everything you learn from this book if you later tackle C++. Many of the features of C can be found in the other C-based languages as well.

## 1.2 Strengths and Weaknesses of C

Like any other programming language, C has strengths and weaknesses. Both stem from the language's original use (writing operating systems and other systems software) and its underlying philosophy:

- **C** is a low-level language. To serve as a suitable language for systems programming, C provides access to machine-level concepts (bytes and addresses, for example) that other programming languages try to hide. C also provides operations that correspond closely to a computer's built-in instructions, so that programs can be fast. Since application programs rely on it for input/ouiput, storage management, and numerous other services, an operating system can't afford to be slow.
- C is a small language. C provides a more limited set of features than many languages. (The reference manual in the second edition of K&R covers the entire language in 49 pages.) To keep the number of features small, C relies heavily on a "library" of standard functions. (A "function" is similar to what other programming languages might call a "procedure," "subroutine," or "method.")
- C is a permissive language. C assumes that you know what you're doing, so it allows you a wider degree of latitude than many languages. Moreover, C doesn't mandate the detailed error-checking found in other languages.

#### Strengths

C's strengths help explain why the language has become so popular:

- Efficiency. Efficiency has been one of C's advantages from the beginning. Because C was intended for applications where assembly language had traditionally been used, it was crucial that C programs could run quickly and in limited amounts of memory.
- Portability. Although program portability wasn't a primary goal of C, it has turned out to be one of the language's strengths. When a program must run on computers ranging from PCs to supercomputers, it is often written in C. One reason for the portability of C programs is that—thanks to C's early association with UNIX and the later ANSI/ISO standards—the language hasn't splintered into incompatible dialects. Another is that C compilers are small and easily written, which has helped make them widely available. Finally, C itself has features that support portability (although there's nothing to prevent programmers from writing nonportable programs).
- Power. C's large collection of data types and operators help make it a power-ful language. In C, it's often possible to accomplish quite a bit with just a few lines of code.

- has no inherent restrictions that limit it to this arena. C is now used for applications of all kinds, from embedded systems to commercial data processing. Moreover, C imposes very few restrictions on the use of its features; operations that would be illegal in other languages are often permitted in C. For example, C allows a character to be added to an integer value (or, for that matter, a floating-point number). This flexibility can make programming easier, although it may allow some bugs to slip through.
- Standard library. One of C's great strengths is its standard library, which contains hundreds of functions for input/output, string handling, storage allocation, and other useful operations.
- Integration with UNIX. C is particularly powerful in combination with UNIX (including the popular variant known as Linux). In fact, some UNIX tools assume that the user knows C.

#### Weaknesses

C's weaknesses arise from the same source as many of its strengths: C's closeness to the machine. Here are a few of C's most notorious problems:

- C programs can be error-prone. C's flexibility makes it an error-prone language. Programming mistakes that would be caught in many other languages can't be detected by a C compiler. In this respect, C is a lot like assembly language, where most errors aren't detected until the program is run. To make matters worse, C contains a number of pitfalls for the unwary. In later chapters, we'll see how an extra semicolon can create an infinite loop or a missing & symbol can cause a program crash.
- C programs can be difficult to understand. Although C is a small language by most measures, it has a number of features that aren't found in all programming languages (and that consequently are often misunderstood). These features can be combined in a great variety of ways, many of which—although obvious to the original author of a program—can be hard for others to understand. Another problem is the terse nature of C programs. C was designed at a time when interactive communication with computers was tedious at best. As a result, C was purposefully kept terse to minimize the time required to enter and edit programs. C's flexibility can also be a negative factor; programmers who are too clever for their own good can make programs almost impossible to understand.
- C programs can be difficult to modify. Large programs written in C can be hard to change if they haven't been designed with maintenance in mind. Modern programming languages usually provide features such as classes and packages that support the division of a large program into more manageable pieces. C, unfortunately, lacks such features.

#### Obfuscated C

Even C's most ardent admirers admit that C code can be hard to read. The annual International Obfuscated C Code Contest actually encourages contestants to write the most confusing C programs possible. The winners are truly baffling, as 1990's "Best Small Program" shows:

```
v,i,j,k,l,s,a[99];
main()
{
   for(scanf("%d",&s);*a-s;v=a[j*=v]-a[i],k=i<s,j+=(v=j<s&&
   (!k&&!!printf(2+"\n\n%c"-(!l<<!j)," #Q"[l^v?(l^j)&1:2])&&
++l||a[i]<s&&v&v-i+j&&v+i-j))&&!(l%=s),v||(i==j?a[i+=k]=0:
++a[i])>=s*k&&++a[--i])
   ;
}
```

This program, written by Doron Osovlanski and Baruch Nissenbaum, prints all solutions to the Eight Queens problem (the problem of placing eight queens on a chessboard in such a way that no queen attacks any other queen). In fact, it works for any number of queens between four and 99. For more winning programs, visit www.ioccc.org, the contest's web site.

#### Effective Use of C

Using C effectively requires taking advantage of C's strengths while avoiding its weaknesses. Here are a few suggestions:

- Learn how to avoid C pitfalls. Hints for avoiding pitfalls are scattered throughout this book—just look for the △ symbol. For a more extensive list of pitfalls, see Andrew Koenig's C Traps and Pitfalls (Reading, Mass.: Addison-Wesley, 1989). Modern compilers will detect common pitfalls and issue warnings, but no compiler spots them all.
- Use software tools to make programs more reliable. C programmers are prolific tool builders (and users). One of the most famous C tools is named lint. lint, which is traditionally provided with UNIX, can subject a program to a more extensive error analysis than most C compilers. If lint (or a similar program) is available, it's a good idea to use it. Another useful tool is a debugger. Because of the nature of C, many bugs can't be detected by a C compiler; these show up instead in the form of run-time errors or incorrect output. Consequently, using a good debugger is practically mandatory for C programmers.
- Take advantage of existing code libraries. One of the benefits of using C is that so many other people also use it; it's a good bet that they've written code you can employ in your own programs. C code is often bundled into libraries (collections of functions); obtaining a suitable library is a good way to reduce errors—and save considerable programming effort. Libraries for common

Q&A

tasks, including user-interface development, graphics, communications, database management, and networking, are readily available. Some libraries are in the public domain, some are open source, and some are sold commercially.

- Adopt a sensible set of coding conventions. A coding convention is a style rule that a programmer has decided to adopt even though it's not enforced by the language. Well-chosen conventions help make programs more uniform, easier to read, and easier to modify. Conventions are important when using any programming language, but especially so with C. As noted above, C's highly flexible nature makes it possible for programmers to write code that is all but unreadable. The programming examples in this book follow one set of conventions, but there are other, equally valid, conventions in use. (We'll discuss some of the alternatives from time to time.) Which set you use is less important than adopting some conventions and sticking to them.
- Avoid "tricks" and overly complex code. C encourages programming tricks. There are usually several ways to accomplish a given task in C: programmers are often tempted to choose the method that's most concise. Don't get carried away; the shortest solution is often the hardest to comprehend. In this book, I'll illustrate a style that's reasonably concise but still understandable.
- Stick to the standard. Most C compilers provide language features and library functions that aren't part of the C89 or C99 standards. For portability, it's best to avoid using nonstandard features and libraries unless they're absolutely necessary.

#### Q & A

#### Q: What is this Q&A section anyway?

A: Glad you asked. The Q&A section, which appears at the end of each chapter, serves several purposes.

The primary purpose of Q&A is to tackle questions that are frequently asked by students learning C. Readers can participate in a dialogue (more or less) with the author, much the same as if they were attending one of my C classes.

Another purpose of Q&A is to provide additional information about topics covered in the chapter. Readers of this book will likely have widely varying backgrounds. Some will be experienced in other programming languages, whereas others will be learning to program for the first time. Readers with experience in a variety of languages may be satisfied with a brief explanation and a couple of examples, but readers with less experience may need more. The bottom line: If you find the coverage of a topic to be sketchy, check Q&A for more details.

On occasion, Q&A will discuss common differences among C compilers. For example, we'll cover some frequently used (but nonstandard) features that are provided by particular compilers.

#### Q: What does lint do? [p. 6]

A: lint checks a C program for a host of potential errors, including—but not limited to—suspicious combinations of types, unused variables, unreachable code, and nonportable code. It produces a list of diagnostic messages, which the programmer must then sift through. The advantage of using lint is that it can detect errors that are missed by the compiler. On the other hand, you've got to remember to use lint; it's all too easy to forget about it. Worse still, lint can produce messages by the hundreds, of which only a fraction refer to actual errors.

#### Q: Where did lint get its name?

A: Unlike the names of many other UNIX tools, lint isn't an acronym; it got its name from the way it picks up pieces of "fluff" from a program.

#### Q: How do I get a copy of lint?

A: lint is a standard UNIX utility; if you rely on another operating system, then you probably don't have lint. Fortunately, versions of lint are available from third parties. An enhanced version of lint known as splint (Secure Programming Lint) is included in many Linux distributions and can be downloaded for free from www.splint.org.

#### Q: Is there some way to force a compiler to do a more thorough job of errorchecking, without having to use lint?

A: Yes. Most compilers will do a more thorough check of a program if asked to. In addition to checking for errors (undisputed violations of the rules of C), most compilers also produce warning messages, indicating potential trouble spots. Some compilers have more than one "warning level"; selecting a higher level causes the compiler to check for more problems than choosing a lower level. If your compiler supports warning levels, it's a good idea to select the highest level, causing the compiler to perform the most thorough job of checking that it's capable of. Error-checking options for the GCC compiler, which is distributed with Linux, are discussed in the Q&A section at the end of Chapter 2.

## \*Q: I'm interested in making my program as reliable as possible. Are there any other tools available besides lint and debuggers?

A: Yes. Other common tools include "bounds-checkers" and "leak-finders." C doesn't require that array subscripts be checked; a bounds-checker adds this capability. A leak-finder helps locate "memory leaks": blocks of memory that are dynamically allocated but never deallocated.

<sup>\*</sup>Starred questions cover material too advanced or too esoteric to interest the average reader, and often refer to topics covered in later chapters. Curious readers with a fair bit of programming experience may wish to delve into these questions immediately; others should definitely skip them on a first reading.

## 2 C Fundamentals

One man's constant is another man's variable.

This chapter introduces several basic concepts, including preprocessing directives, functions, variables, and statements, that we'll need in order to write even the simplest programs. Later chapters will cover these topics in much greater detail.

To start off, Section 2.1 presents a small C program and describes how to compile and link it. Section 2.2 then discusses how to generalize the program, and Section 2.3 shows how to add explanatory remarks, known as comments. Section 2.4 introduces variables, which store data that may change during the execution of a program, and Section 2.5 shows how to use the scanf function to read data into variables. Constants—data that won't change during program execution—can be given names, as Section 2.6 shows. Finally, Section 2.7 explains C's rules for creating names (identifiers) and Section 2.8 gives the rules for laying out a program.

### 2.1 Writing a Simple Program

In contrast to programs written in some languages, C programs require little "boilerplate"—a complete program can be as short as a few lines.

#### PROGRAM Printing a Pun

The first program in Kernighan and Ritchie's classic *The C Programming Language* is extremely short; it does nothing but write the message hello, world. Unlike other C authors, I won't use this program as my first example. I will, however, uphold another C tradition: the bad pun. Here's the pun:

To C, or not to C: that is the question.

The following program, which we'll name pun.c, displays this message each time it is run.

```
pun.c #include <stdio.h>
int main(void)
{
    printf("To C, or not to C: that is the question.\n");
    return 0;
}
```

Section 2.2 explains the form of this program in some detail. For now, I'll just make a few brief observations. The line

```
#include <stdio.h>
```

is necessary to "include" information about C's standard I/O (input/output) library. The program's executable code goes inside main, which represents the "main" program. The only line inside main is a command to display the desired message. printf is a function from the standard I/O library that can produce nicely formatted output. The \n code tells printf to advance to the next line after printing the message. The line

```
return 0;
```

indicates that the program "returns" the value 0 to the operating system when it terminates.

#### **Compiling and Linking**

Despite its brevity, getting pun.c to run is more involved than you might expect. First, we need to create a file named pun.c containing the program (any text editor will do). The name of the file doesn't matter, but the .c extension is often required by compilers.

Next, we've got to convert the program to a form that the machine can execute. For a C program, that usually involves three steps:

- *Preprocessing*. The program is first given to a *preprocessor*, which obeys commands that begin with # (known as *directives*). A preprocessor is a bit like an editor; it can add things to the program and make modifications.
- Compiling. The modified program now goes to a compiler, which translates it into machine instructions (object code). The program isn't quite ready to run yet, however.
- Linking. In the final step, a linker combines the object code produced by the compiler with any additional code needed to yield a complete executable program. This additional code includes library functions (like printf) that are used in the program.

Fortunately, this process is often automated, so you won't find it too onerous. In fact, the preprocessor is usually integrated with the compiler, so you probably won't even notice it at work.

The commands necessary to compile and link vary, depending on the compiler and operating system. Under UNIX, the C compiler is usually named cc. To compile and link the pun.c program, enter the following command in a terminal or command-line window:

```
% cc pun.c
```

(The % character is the UNIX prompt, not something that you need to enter.) Linking is automatic when using cc: no separate link command is necessary.

After compiling and linking the program, cc leaves the executable program in a file named a .out by default. cc has many options; one of them (the -o option) allows us to choose the name of the file containing the executable program. For example, if we want the executable version of pun. c to be named pun, we would enter the following command:

```
% cc -o pun pun.c
```

#### The GCC Compiler

One of the most popular C compilers is the GCC compiler, which is supplied with Linux but is available for many other platforms as well. Using this compiler is similar to using the traditional UNIX cc compiler. For example, to compile the pun.c program, we would use the following command:

```
% gcc -o pun pun.c
```

Q&A

The Q&A section at the end of the chapter provides more information about GCC.

#### **Integrated Development Environments**

So far, we've assumed the use of a "command-line" compiler that's invoked by entering a command in a special window provided by the operating system. The alternative is to use an *integrated development environment (IDE)*, a software package that allows us to edit, compile, link, execute, and even debug a program without leaving the environment. The components of an IDE are designed to work together. For example, when the compiler detects an error in a program, it can arrange for the editor to highlight the line that contains the error. There's a great deal of variation among IDEs, so I won't discuss them further in this book. However, I would recommend checking to see which IDEs are available for your platform.

### 2.2 The General Form of a Simple Program

Let's take a closer look at pun.c and see how we can generalize it a bit. Simple C programs have the form

```
directives
int main(void)
{
    statements
}
```

In this template, and in similar templates elsewhere in this book, items printed in Courier would appear in a C program exactly as shown; items in *italics* represent text to be supplied by the programmer.

Notice how the braces show where main begins and ends. C uses { and } in much the same way that some other languages use words like begin and end. This illustrates a general point about C: it relies heavily on abbreviations and special symbols, one reason that C programs are concise (or—less charitably—cryptic).

Q&A

Even the simplest C programs rely on three key language features: directives (editing commands that modify the program prior to compilation), functions (named blocks of executable code, of which main is an example), and statements (commands to be performed when the program is run). We'll take a closer look at these features now.

#### **Directives**

Before a C program is compiled, it is first edited by a preprocessor. Commands intended for the preprocessor are called directives. Chapters 14 and 15 discuss directives in detail. For now, we're interested only in the #include directive.

The pun.c program begins with the line

```
#include <stdio.h>
```

headers ➤ 15.2

This directive states that the information in <stdio.h> is to be "included" into the program before it is compiled. <stdio.h> contains information about C's standard I/O library. C has a number of *headers* like <stdio.h>; each contains information about some part of the standard library. The reason we're including <stdio.h> is that C, unlike some programming languages, has no built-in "read" and "write" commands. The ability to perform input and output is provided instead by functions in the standard library.

Directives always begin with a # character, which distinguishes them from other items in a C program. By default, directives are one line long; there's no semicolon or other special marker at the end of a directive.

#### **Functions**

Functions are like "procedures" or "subroutines" in other programming languages—they're the building blocks from which programs are constructed. In fact, a C program is little more than a collection of functions. Functions fall into two categories: those written by the programmer and those provided as part of the C implementation. I'll refer to the latter as *library functions*, since they belong to a "library" of functions that are supplied with the compiler.

The term "function" comes from mathematics, where a function is a rule for computing a value when given one or more arguments:

$$f(x) = x + 1$$
$$g(y, z) = y^2 - z^2$$

C uses the term "function" more loosely. In C, a function is simply a series of statements that have been grouped together and given a name. Some functions compute a value; some don't. A function that computes a value uses the return statement to specify what value it "returns." For example, a function that adds I to its argument might execute the statement

```
return x + 1;
```

while a function that computes the difference of the squares of its arguments might execute the statement

```
return y * y - z * z;
```

Although a C program may consist of many functions, only the main function is mandatory. main is special: it gets called automatically when the program is executed. Until Chapter 9, where we'll learn how to write other functions, main will be the only function in our programs.



The name main is critical; it can't be begin or start or even MAIN.

If main is a function, does it return a value? Yes: it returns a status code that is given to the operating system when the program terminates. Let's take another look at the pun.c program:

```
#include <stdio.h>
int main(void)
{
  printf("To C, or not to C: that is the question.\n");
  return 0;
}
```

The word int just before main indicates that the main function returns an integer value. The word void in parentheses indicates that main has no arguments.

The statement

return 0;

return value of main ▶9.5



has two effects: it causes the main function to terminate (thus ending the program) and it indicates that the main function returns a value of 0. We'll have more to say about main's return value in a later chapter. For now, we'll always have main return the value 0, which indicates normal program termination.

If there's no return statement at the end of the main function, the program will still terminate. However, many compilers will produce a warning message (because the function was supposed to return an integer but failed to).

Q&A

Statements

A statement is a command to be executed when the program runs. We'll explore statements later in the book, primarily in Chapters 5 and 6. The pun.c program uses only two kinds of statements. One is the return statement; the other is the function call. Asking a function to perform its assigned task is known as calling the function. The pun.c program, for example, calls the printf function to display a string on the screen:

```
printf("To C, or not to C: that is the question.\n");
```

compound statement ➤ 5.2

C requires that each statement end with a semicolon. (As with any good rule, there's one exception: the compound statement, which we'll encounter later.) The semicolon shows the compiler where the statement ends; since statements can continue over several lines, it's not always obvious where they end. Directives, on the other hand, are normally one line long, and they *don't* end with a semicolon.

#### **Printing Strings**

printf is a powerful function that we'll examine in Chapter 3. So far, we've only used printf to display a *string literal*—a series of characters enclosed in double quotation marks. When printf displays a string literal, it doesn't show the quotation marks.

printf doesn't automatically advance to the next output line when it finishes printing. To instruct printf to advance one line, we must include \n (the new-line character) in the string to be printed. Writing a new-line character terminates the current output line; subsequent output goes onto the next line. To illustrate this point, consider the effect of replacing the statement

```
printf("To C, or not to C: that is the question.\n");
by two calls of printf:
printf("To C, or not to C: ");
printf("that is the question.\n");
```