The C Preprocessor

Last revised July 2000

for GCC version 2

Richard M. Stallman

The C Preprocessor

The C preprocessor is a **macro processor** that is used automatically by the C compiler to transform your program before actual compilation. It is called a macro processor because it allows you to define **macros**, which are brief abbreviations for longer constructs.

The C preprocessor is intended only for macro processing of C, C++ and Objective C source files. For macro processing of other files, you are strongly encouraged to use alternatives like M4, which will likely give you better results and avoid many problems. For example, normally the C preprocessor does not preserve arbitrary whitespace verbatim, but instead replaces each sequence with a single space.

For use on C-like source files, the C preprocessor provides four separate facilities that you can use as you see fit:

- Inclusion of header files. These are files of declarations that can be substituted into your program.
- Macro expansion. You can define **macros**, which are abbreviations for arbitrary fragments of C code, and then the C preprocessor will replace the macros with their definitions throughout the program.
- Conditional compilation. Using special preprocessing directives, you can include or exclude parts of the program according to various conditions.
- Line control. If you use a program to combine or rearrange source files into an intermediate file which is then compiled, you can use line control to inform the compiler of where each source line originally came from.

C preprocessors vary in some details. This manual discusses the GNU C preprocessor, which provides a small superset of the features of ISO Standard C.

In its default mode, the GNU C preprocessor does not do a few things required by the standard. These are features which are rarely, if ever, used, and may cause surprising changes to the meaning of a program which does not expect them. To get strict ISO Standard C, you should use the `` `-std=c89' `` or `` `-std=c99' `` options, depending on which version of the standard you want. To get all the mandatory diagnostics, you must also use `` `-pedantic' ``. See section Invoking the C Preprocessor.

Transformations Made Globally

Most C preprocessor features are inactive unless you give specific directives to request their use. (Preprocessing directives are lines starting with a `#' token, possibly preceded by whitespace; see section Preprocessing Directives). However, there are four transformations that the preprocessor always makes on all the input it receives, even in the absence of directives. These are, in order:

1. Trigraphs, if enabled, are replaced with the character they represent.
2. Backslash-newline sequences are deleted, no matter where. This feature allows you to break long lines for cosmetic purposes without changing their meaning. Recently, the non-traditional preprocessor has relaxed its treatment of escaped newlines. Previously, the newline had to immediately follow a backslash. The current implementation allows whitespace in the form of spaces, horizontal and vertical tabs, and form feeds between the backslash and the subsequent newline. The preprocessor issues a warning, but treats it as a valid escaped newline and combines the two lines to form a single logical line. This works within comments and tokens, including multi-line strings, as well as between tokens. Comments are *not* treated as whitespace for the purposes of this relaxation, since they have not yet been replaced with spaces.
3. All comments are replaced with single spaces.
4. Predefined macro names are replaced with their expansions (see section Predefined Macros).

For end-of-line indicators, any of \n, \r\n, \n\r and \r are recognised, and treated as ending a single line. As a result, if you mix these in a single file you might get incorrect line numbering, because the preprocessor would interpret the two-character versions as ending just one line. Previous implementations would only handle UNIX-style \n correctly, so DOS-style \r\n would need to be passed through a filter first.

The first three transformations are done *before* all other parsing and before preprocessing directives are recognized. Thus, for example, you can split a line mechanically with backslash-newline anywhere (except within trigraphs since they are replaced first; see below).

```
/*
*/ # /*
*/ defi\
ne FO\
O 10\
20
```

is equivalent into `#define FOO 1020'.

There is no way to prevent a backslash at the end of a line from being interpreted as a backslash-newline. For example,

```
"foo\\
bar"
```

is equivalent to `"foo\bar"`, not to `"foo\\bar"`. To avoid having to worry about this, do not use the GNU extension which permits multi-line strings. Instead, use string constant concatenation:

```
"foo\\"
"bar"
```

Your program will be more portable this way, too.

There are a few things to note about the above four transformations.

- Comments and predefined macro names (or any macro names, for that matter) are not recognized inside the argument of an `#include' directive, when it is delimited with quotes or with `<' and `>'.
- Comments and predefined macro names are never recognized within a character or string constant.
- ISO "trigraphs" are converted before backslash-newlines are deleted. If you write what looks like a trigraph with a backslash-newline inside, the backslash-newline is deleted as usual, but it is too late to recognize the trigraph. This is relevant only if you use the `-trigraphs' option to enable trigraph processing. See section Invoking the C Preprocessor.

The preprocessor handles null characters embedded in the input file depending upon the context in which the null appears. Note that here we are referring not to the two-character escape sequence "\0", but to the single character ASCII NUL.

There are three different contexts in which a null character may appear:

- Within comments. Here, null characters are silently ignored.
- Within a string or character constant. Here the preprocessor emits a warning, but preserves the null character and passes it through to the output file or compiler front-end.
- In any other context, the preprocessor issues a warning, and discards the null character. The preprocessor treats it like whitespace, combining it with any surrounding whitespace to become a single whitespace block. Representing the null character by "^@", this means that code like
- ```
  #define X^@1
  ```
  ```

is equivalent to

```
#define X 1
```

and X is defined with replacement text "1".

Preprocessing Directives

Most preprocessor features are active only if you use preprocessing directives to request their use.

Preprocessing directives are lines in your program that start with `#'. Whitespace is allowed before and after the `#'. The `#' is followed by an identifier that is the **directive name**. For example, `#define' is the directive that defines a macro.

Since the `#' must be the first token on the line, it cannot come from a macro expansion if you wish it to begin a directive. Also, the directive name is not macro expanded. Thus, if `foo' is defined as a macro expanding to `define', that does not make `#foo' a valid preprocessing directive.

The set of valid directive names is fixed. Programs cannot define new preprocessing directives.

Some directive names require arguments; these make up the rest of the directive line and must be separated from the directive name by whitespace. For example, `#define' must be followed by a macro name and the intended expansion of the macro. See section Object-like Macros.

A preprocessing directive cannot cover more than one line. It may be logically extended with backslash-newline, but that has no effect on its meaning. Comments containing newlines can also divide the directive into multiple lines, but a comment is replaced by a single space before the directive is interpreted.

Header Files

A header file is a file containing C declarations and macro definitions (see section Macros) to be shared between several source files. You request the use of a header file in your program with the C preprocessing directive `#include'.

- Header Uses: What header files are used for.
- Include Syntax: How to write `#include' directives.
- Include Operation: What `#include' does.
- Once-Only: Preventing multiple inclusion of one header file.
- Inheritance: Including one header file in another header file.
- System Headers: Special treatment for some header files.

Uses of Header Files

Header files serve two kinds of purposes.

- System header files declare the interfaces to parts of the operating system. You include them in your program to supply the definitions and declarations you need to invoke system calls and libraries.
- Your own header files contain declarations for interfaces between the source files of your program. Each time you have a group of related declarations and macro definitions all or most of which are needed in several different source files, it is a good idea to create a header file for them.

Including a header file produces the same results in C compilation as copying the header file into each source file that needs it. Such copying would be time-consuming and error-prone. With a header file, the related declarations appear in only one place. If they need to be changed, they can be changed in one place, and programs that include the header file will automatically use the new version when next recompiled. The header file eliminates the labor of finding and changing all the copies as well as the risk that a failure to find one copy will result in inconsistencies within a program.

The usual convention is to give header files names that end with `.h'. Avoid unusual characters in header file names, as they reduce portability.

The `#include' Directive

Both user and system header files are included using the preprocessing directive `#include'. It has three variants:

`#include <`*file*`>`
: This variant is used for system header files. It searches for a file named *file* in a list of directories specified by you, then in a standard list of system directories. You specify directories to search for header files with the command option `-I' (see section Invoking the C Preprocessor). The option `-nostdinc' inhibits searching the standard system directories; in this case only the directories you specify are searched. The first `>' character terminates the file name. The file name may contain a `<' character.

`#include "`*file*`"`
: This variant is used for header files of your own program. It searches for a file named *file* first in the current directory, then in the same directories used for system header files. The current directory is the directory of the current input file. It is tried first because it is presumed to be the location of the files that the current input file refers to. (If the `-I-' option is used, the special treatment of the current directory is inhibited. See section Invoking the C Preprocessor.) The first `"' character terminates the file name. In both these variants, the argument behaves like a string constant in that comments are not recognized, and macro names are not expanded. Thus, in `#include <x/*y>' the `/*' does not start a comment and the directive specifies inclusion of a system header file named `x/*y'. However, in either variant, if backslashes occur within *file*, they are considered ordinary text characters, not escape characters. None of the character escape sequences appropriate to string constants in C are processed. Thus, `#include "x\n\\y"' specifies a filename containing three backslashes.

`#include` *anything else*
: This variant is called a **computed #include**. Any `#include' directive whose argument does not fit the above two forms is a computed include. The text *anything else* is checked for macro calls, which are expanded (see section Macros). When this is done, the result must match one of the above

two variants -- in particular, the expansion must form a string literal token, or a sequence of tokens surrounded by angle braces. See section Implementation-defined Behavior and Implementation Limits. This feature allows you to define a macro which controls the file name to be used at a later point in the program. One application of this is to allow a site-specific configuration file for your program to specify the names of the system include files to be used. This can help in porting the program to various operating systems in which the necessary system header files are found in different places.

How `#include' Works

The `#include' directive works by directing the C preprocessor to scan the specified file as input before continuing with the rest of the current file. The output from the preprocessor contains the output already generated, followed by the output resulting from the included file, followed by the output that comes from the text after the `#include' directive. For example, given a header file `header.h' as follows,

```
char *test ();
```

and a main program called `program.c' that uses the header file, like this,

```
int x;
#include "header.h"

main ()
{
  printf (test ());
}
```

the output generated by the C preprocessor for `program.c' as input would be

```
int x;
char *test ();

main ()
{
  printf (test ());
}
```

Included files are not limited to declarations and macro definitions; those are merely the typical uses. Any fragment of a C program can be included from another file. The include file could even contain the beginning of a statement that is concluded in the containing file, or the end of a statement that was started in the including file. However, a comment or a string or character constant may not start in the included file and finish in the including file. An unterminated comment, string constant or character constant in an included file is considered to end (with an error message) at the end of the file.

It is possible for a header file to begin or end a syntactic unit such as a function definition, but that would be very confusing, so don't do it.

The line following the `#include' directive is always treated as a separate line by the C preprocessor, even if the included file lacks a final newline.

Once-Only Include Files

Very often, one header file includes another. It can easily result that a certain header file is included more than once. This may lead to errors, if the header file defines structure types or typedefs, and is certainly wasteful. Therefore, we often wish to prevent multiple inclusion of a header file.

The standard way to do this is to enclose the entire real contents of the file in a conditional, like this:

```
#ifndef FILE_FOO_SEEN
#define FILE_FOO_SEEN

the entire file

#endif /* FILE_FOO_SEEN */
```

The macro `FILE_FOO_SEEN` indicates that the file has been included once already. In a user header file, the macro name should not begin with `` `_' ``. In a system header file, this name should begin with `` `__' `` to avoid conflicts with user programs. In any kind of header file, the macro name should contain the name of the file and some additional text, to avoid conflicts with other header files.

The GNU C preprocessor is programmed to notice when a header file uses this particular construct and handle it efficiently. If a header file is contained entirely in a `` `#ifndef' `` conditional, modulo whitespace and comments, then it remembers that fact. If a subsequent `` `#include' `` specifies the same file, and the macro in the `` `#ifndef' `` is already defined, then the directive is skipped without processing the specified file at all.

In the Objective C language, there is a variant of `` `#include' `` called `` `#import' `` which includes a file, but does so at most once. If you use `` `#import' `` *instead of* `` `#include' ``, then you don't need the conditionals inside the header file to prevent multiple execution of the contents.

`` `#import' `` is obsolete because it is not a well designed feature. It requires the users of a header file -- the applications programmers --- to know that a certain header file should only be included once. It is much better for the header file's implementor to write the file so that users don't need to know this. Using `` `#ifndef' `` accomplishes this goal.

Inheritance and Header Files

Inheritance is what happens when one object or file derives some of its contents by virtual copying from another object or file. In the case of C header files, inheritance means that one header file includes another header file and then replaces or adds something.

If the inheriting header file and the base header file have different names, then inheritance is straightforward: simply write `` `#include "*base*"' `` in the inheriting file.

Sometimes it is necessary to give the inheriting file the same name as the base file. This is less straightforward.

For example, suppose an application program uses the system header `sys/signal.h', but the version of `/usr/include/sys/signal.h' on a particular system doesn't do what the application program expects. It might be convenient to define a "local" version, perhaps under the name `/usr/local/include/sys/signal.h', to override or add to the one supplied by the system.

You can do this by compiling with the option `-I.', and writing a file `sys/signal.h' that does what the application program expects. Making this file include the standard `sys/signal.h' is not so easy --- writing `#include <sys/signal.h>' in that file doesn't work, because it includes your own version of the file, not the standard system version. Used in that file itself, this leads to an infinite recursion and a fatal error in compilation.

`#include </usr/include/sys/signal.h>' would find the proper file, but that is not clean, since it makes an assumption about where the system header file is found. This is bad for maintenance, since it means that any change in where the system's header files are kept requires a change somewhere else.

The clean way to solve this problem is to use `#include_next', which means, "Include the *next* file with this name." This directive works like `#include' except in searching for the specified file: it starts searching the list of header file directories *after* the directory in which the current file was found.

Suppose you specify `-I /usr/local/include', and the list of directories to search also includes `/usr/include'; and suppose both directories contain `sys/signal.h'. Ordinary `#include <sys/signal.h>' finds the file under `/usr/local/include'. If that file contains `#include_next <sys/signal.h>', it starts searching after that directory, and finds the file in `/usr/include'.

`#include_next' is a GCC extension and should not be used in programs intended to be portable to other compilers.

System Headers

The header files declaring interfaces to the operating system and runtime libraries often cannot be written in strictly conforming C. Therefore, GNU C gives code found in **system headers** special treatment. Certain categories of warnings are suppressed, notably those enabled by `-pedantic'.

Normally, only the headers found in specific directories are considered system headers. The set of these directories is determined when GCC is compiled. There are, however, two ways to add to the set.

The `-isystem' command line option adds its argument to the list of directories to search for headers, just like `-I'. In addition, any headers found in that directory will be considered system headers. Note that unlike `-I', you must put a space between `-isystem' and its argument.

All directories named by `-isystem' are searched **after** all directories named by `-I', no matter what their order was on the command line. If the same directory is named by both `-I' and `-isystem', `-I' wins; it is as if the `-isystem' option had never been specified at all.

There is also a directive, `#pragma GCC system_header', which tells GCC to consider the rest of the current include file a system header, no matter where it was found. Code that comes before the `#pragma' in the file will not be affected.

`#pragma GCC system_header' has no effect in the primary source file.

Macros

A macro is a sort of abbreviation which you can define once and then use later. There are many complicated features associated with macros in the C preprocessor.

- Object-like Macros: Macros that always expand the same way.
- Function-like Macros: Macros that accept arguments that are substituted into the macro expansion.
- Macro Varargs: Macros with variable number of arguments.
- Predefined: Predefined macros that are always available.
- Stringification: Macro arguments converted into string constants.
- Concatenation: Building tokens from parts taken from macro arguments.
- Undefining: Cancelling a macro's definition.
- Redefining: Changing a macro's definition.
- Poisoning: Ensuring a macro is never defined or used.
- Macro Pitfalls: Macros can confuse the unwary. Here we explain several common problems and strange features.

Object-like Macros

An **object-like macro** is a kind of abbreviation. It is a name which stands for a fragment of code. Some people refer to these as **manifest constants**.

Before you can use a macro, you must **define** it explicitly with the `#define' directive. `#define' is followed by the name of the macro and then the token sequence it should be an abbreviation for, which is variously referred to as the macro's **body**, **expansion** or **replacement list**. For example,

```
#define BUFFER_SIZE 1020
```

defines a macro named `BUFFER_SIZE' as an abbreviation for the token `1020'. If somewhere after this `#define' directive there comes a C statement of the form

```
foo = (char *) xmalloc (BUFFER_SIZE);
```

then the C preprocessor will recognize and **expand** the macro `BUFFER_SIZE', resulting in

```
foo = (char *) xmalloc (1020);
```

The use of all upper case for macro names is a standard convention. Programs are easier to read when it is possible to tell at a glance which names are macros.

Normally, a macro definition can only span a single logical line, like all C preprocessing directives. Comments within a macro definition may contain newlines, which make no difference since each comment is replaced by a space regardless of its contents.

Apart from this, there is no restriction on what can go in a macro body provided it decomposes into valid preprocessing tokens. In particular, parentheses need not balance, and the body need not resemble valid C code. (If it does not, you may get error messages from the C compiler when you use the macro.)

The C preprocessor scans your program sequentially, so macro definitions take effect at the place you write them. Therefore, the following input to the C preprocessor

```
foo = X;
#define X 4
bar = X;
```

produces as output

```
foo = X;

bar = 4;
```

When the preprocessor expands a macro name, the macro's expansion replaces the macro invocation, and the result is re-scanned for more macros to expand. For example, after

```
#define BUFSIZE 1020
#define TABLESIZE BUFSIZE
```

the name `TABLESIZE' when used in the program would go through two stages of expansion, resulting ultimately in `1020'.

This is not the same as defining `TABLESIZE' to be `1020'. The `#define' for `TABLESIZE' uses exactly the expansion you specify -- in this case, `BUFSIZE' -- and does not check to see whether it too contains macro names. Only when you *use* `TABLESIZE' is the result of its expansion scanned for more macro names. See section Cascaded Use of Macros.

Macros with Arguments

An object-like macro is always replaced by exactly the same tokens each time it is used. Macros can be made more flexible by taking **arguments**. Arguments are fragments of code that you supply each time the macro is used. These fragments are included in the expansion of the macro according to the directions in the macro definition. A macro that accepts arguments is called a **function-like macro** because the syntax for using it looks like a function call.

To define a macro that uses arguments, you write a `#define' directive with a list of **parameters** in parentheses after the name of the macro. The parameters must be valid C identifiers, separated by commas and optionally whitespace. The `(' must follow the macro name immediately, with no space in between. If

you leave a space, you instead define an object-like macro whose expansion begins with a `(', and often leads to confusing errors at compile time.

As an example, here is a macro that computes the minimum of two numeric values, as it is defined in many C programs:

```
#define min(X, Y)  ((X) < (Y) ? (X) : (Y))
```

(This is not the best way to define a "minimum" macro in GNU C. See section Duplication of Side Effects, for more information.)

To invoke a function-like macro, you write the name of the macro followed by a list of **arguments** in parentheses, separated by commas. The invocation of the macro need not be restricted to a single logical line - it can cross as many lines in the source file as you wish. The number of arguments you give must match the number of parameters in the macro definition; empty arguments are fine. Examples of use of the macro `min' include `min (1, 2)' and `min (x + 28, *p)'.

The expansion text of the macro depends on the arguments you use. Each macro parameter is replaced throughout the macro expansion with the tokens of the corresponding argument. Leading and trailing argument whitespace is dropped, and all whitespace between the tokens of an argument is reduced to a single space. Using the same macro `min' defined above, `min (1, 2)' expands into

```
((1) < (2) ? (1) : (2))
```

where `1' has been substituted for `X' and `2' for `Y'.

Likewise, `min (x + 28, *p)' expands into

```
((x + 28) < (*p) ? (x + 28) : (*p))
```

Parentheses within each argument must balance; a comma within such parentheses does not end the argument. However, there is no requirement for square brackets or braces to balance, and they do not prevent a comma from separating arguments. Thus,

```
macro (array[x = y, x + 1])
```

passes two arguments to `macro`: `array[x = y' and `x + 1]'. If you want to supply `array[x = y, x + 1]' as an argument, you must write it as `array[(x = y, x + 1)]', which is equivalent C code.

After the arguments have been substituted into the macro body, the resulting expansion replaces the macro invocation, and re-scanned for more macro calls. Therefore even arguments can contain calls to other macros, either with or without arguments, and even to the same macro. For example, `min (min (a, b), c)' expands into this text:

```
((((a) < (b) ? (a) : (b))) < (c)
 ? (((a) < (b) ? (a) : (b)))
```

```
 : (c))
```

(Line breaks shown here for clarity would not actually be generated.)

If a macro `foo` takes one argument, and you want to supply an empty argument, simply supply no preprocessing tokens. Since whitespace does not form a preprocessing token, it is optional. For example, `` `foo ()' ``, `` `foo ( )' `` and `` `bar (, arg2)' ``.

Previous GNU preprocessor implementations and documentation were incorrect on this point, insisting that a function-like macro that takes a single argument be passed a space if an empty argument was required.

If you use a macro name followed by something other than a `` `(' `` (after ignoring any whitespace that might follow), it does not form an invocation of the macro, and the preprocessor does not change what you have written. Therefore, it is possible for the same identifier to be a variable or function in your program as well as a macro, and you can choose in each instance whether to refer to the macro (if an actual argument list follows) or the variable or function (if an argument list does not follow). For example,

```
#define foo(X) X
foo bar foo(baz)
```

expands to `` `foo bar baz' ``. Such dual use of one name could be confusing and should be avoided except when the two meanings are effectively synonymous: that is, when the name is both a macro and a function and the two have similar effects. You can think of the name simply as a function; use of the name for purposes other than calling it (such as, to take the address) will refer to the function, while calls will expand the macro and generate better but equivalent code.

For example, you can use a function named `` `min' `` in the same source file that defines the macro. If you write `` `&min' `` with no argument list, you refer to the function. If you write `` `min (x, bb)' ``, with an argument list, the macro is expanded. If you write `` `(min) (a, bb)' ``, where the name `` `min' `` is not followed by an open-parenthesis, the macro is not expanded, so you wind up with a call to the function `` `min' ``.

In the definition of a macro with arguments, the list of argument names must follow the macro name immediately with no space in between. If there is a space after the macro name, the macro is defined as taking no arguments, and all the rest of the line is taken to be the expansion. The reason for this is that it is often useful to define a macro that takes no arguments and whose definition begins with an identifier in parentheses. This rule makes it possible for you to do either this:

```
#define FOO(x) - 1 / (x)
```

(which defines `` `FOO' `` to take an argument and expand into minus the reciprocal of that argument) or this:

```
#define BAR (x) - 1 / (x)
```

(which defines `` `BAR' `` to take no argument and always expand into `` `(x) - 1 / (x)' ``).

Note that the *uses* of a macro with arguments can have spaces before the left parenthesis; it's the *definition* where it matters whether there is a space.

Macros with Variable Numbers of Arguments

In the ISO C standard of 1999, a macro can be declared to accept a variable number of arguments much as a function can. The syntax for defining the macro is similar to that of a function. Here is an example:

```
#define eprintf(...) fprintf (stderr, __VA_ARGS__)
```

Here `...' is a **variable argument**. In the invocation of such a macro, it represents the zero or more tokens until the closing parenthesis that ends the invocation, including any commas. This set of tokens replaces the identifier __VA_ARGS__ in the macro body wherever it appears. Thus, we have this expansion:

```
eprintf ("%s:%d: ", input_file_name, line_number)
==>
fprintf (stderr, "%s:%d: " , input_file_name, line_number)
```

Within a `#define' directive, ISO C mandates that the only place the identifier __VA_ARGS__ can appear is in the replacement list of a variable-argument macro. It may not be used as a macro name, macro argument name, or within a different type of macro. It may also be forbidden in open text; the standard is ambiguous. We recommend you avoid using it except for its defined purpose.

If your macro is complicated, you may want a more descriptive name for the variable argument than __VA_ARGS__. GNU cpp permits this, as an extension. You may write an argument name immediately before the `...'; that name is used for the variable argument. The `eprintf` macro above could be written

```
#define eprintf(args...) fprintf (stderr, args)
```

using this extension. You cannot use __VA_ARGS__ and this extension in the same macro.

We might instead have defined eprintf as follows:

```
#define eprintf(format, ...) fprintf (stderr, format, __VA_ARGS__)
```

This formulation looks more descriptive, but cannot be used as flexibly. There is no way to produce expanded output of

```
fprintf (stderr, "success!\n")
```

because, in standard C, you are not allowed to leave the variable argument out entirely, and passing an empty argument for the variable arguments will not do what you want. Writing

```
eprintf ("success!\n", )
```

produces

```
fprintf (stderr, "success!\n",)
```

where the extra comma originates from the replacement list and not from the arguments to eprintf.

There is another extension in the GNU C preprocessor which deals with this difficulty. First, you are allowed to leave the variable argument out entirely:

```
eprintf ("success!\n")
```

Second, the `##' token paste operator has a special meaning when placed between a comma and a variable argument. If you write

```
#define eprintf(format, ...) fprintf (stderr, format, ##__VA_ARGS__)
```

and the variable argument is left out when the `eprintf' macro is used, then the comma before the `##' will be deleted. This does *not* happen if you pass an empty argument, nor does it happen if the token preceding `##' is anything other than a comma.

Previous versions of the preprocessor implemented this extension much more generally. We have restricted it in order to minimize the difference from the C standard. See section Undefined Behavior and Deprecated Features.

Predefined Macros

Several object-like macros are predefined; you use them without supplying their definitions. They fall into two classes: standard macros and system-specific macros.

- Standard Predefined: Standard predefined macros.
- Nonstandard Predefined: Nonstandard predefined macros.

Standard Predefined Macros

The standard predefined macros are available with the same meanings regardless of the machine or operating system on which you are using GNU C. Their names all start and end with double underscores. Those preceding __GNUC__ in this table are standardized by ISO C; the rest are GNU C extensions.

__FILE__
: This macro expands to the name of the current input file, in the form of a C string constant. The precise name returned is the one that was specified in `#include' or as the input file name argument. For example, `"/usr/local/include/myheader.h"' is a possible expansion of this macro.

__LINE__
: This macro expands to the current input line number, in the form of a decimal integer constant. While we call it a predefined macro, it's a pretty strange macro, since its "definition" changes with each new line of source code. This and `__FILE__' are useful in generating an error message to

report an inconsistency detected by the program; the message can state the source line at which the inconsistency was detected. For example,

```
fprintf (stderr, "Internal error: "
                 "negative string length "
                 "%d at %s, line %d.",
         length, __FILE__, __LINE__);
```

A `#include' directive changes the expansions of `__FILE__' and `__LINE__' to correspond to the included file. At the end of that file, when processing resumes on the input file that contained the `#include' directive, the expansions of `__FILE__' and `__LINE__' revert to the values they had before the `#include' (but `__LINE__' is then incremented by one as processing moves to the line after the `#include'). The expansions of both `__FILE__' and `__LINE__' are altered if a `#line' directive is used. See section Combining Source Files.

`__DATE__`
: This macro expands to a string constant that describes the date on which the preprocessor is being run. The string constant contains eleven characters and looks like `"Feb 1 1996"'.

`__TIME__`
: This macro expands to a string constant that describes the time at which the preprocessor is being run. The string constant contains eight characters and looks like `"23:59:01"'.

`__STDC__`
: This macro expands to the constant 1, to signify that this is ISO Standard C. (Whether that is actually true depends on what C compiler will operate on the output from the preprocessor.) On some hosts, system include files use a different convention, where `__STDC__' is normally 0, but is 1 if the user specifies strict conformance to the C Standard. The preprocessor follows the host convention when processing system include files, but when processing user files it follows the usual GNU C convention. This macro is not defined if the `-traditional' option is used.

`__STDC_VERSION__`
: This macro expands to the C Standard's version number, a long integer constant of the form `*yyyymm*L' where *yyyy* and *mm* are the year and month of the Standard version. This signifies which version of the C Standard the preprocessor conforms to. Like `__STDC__', whether this version number is accurate for the entire implementation depends on what C compiler will operate on the output from the preprocessor. This macro is not defined if the `-traditional' option is used.

`__GNUC__`
: This macro is defined if and only if this is GNU C. This macro is defined only when the entire GNU C compiler is in use; if you invoke the preprocessor directly, `__GNUC__' is undefined. The value identifies the major version number of GNU CC (`1' for GNU CC version 1, which is now obsolete, and `2' for version 2).

`__GNUC_MINOR__`
: The macro contains the minor version number of the compiler. This can be used to work around differences between different releases of the compiler (for example, if GCC 2.6.3 is known to support a feature, you can test for `__GNUC__ > 2 || (__GNUC__ == 2 && __GNUC_MINOR__ >= 6)`).

`__GNUC_PATCHLEVEL__`
: This macro contains the patch level of the compiler. This can be used to work around differences between different patch level releases of the compiler (for example, if GCC 2.6.2 is known to contain a bug, whereas GCC 2.6.3 contains a fix, and you have code which can workaround the problem depending on whether the bug is fixed or not, you can test for `__GNUC__ > 2 ||`

`(__GNUC__ == 2 && __GNUC_MINOR__ > 6) || (__GNUC__ == 2 && __GNUC_MINOR__ == 6 && __GNUC_PATCHLEVEL__ > 3)`).

`__GNUG__`
: The GNU C compiler defines this when the compilation language is C++; use `__GNUG__' to distinguish between GNU C and GNU C++.

`__cplusplus`
: The ISO standard for C++ requires predefining this variable. You can use `__cplusplus' to test whether a header is compiled by a C compiler or a C++ compiler. The compiler currently uses a value of `1', instead of the value `199711L', which would indicate full conformance with the standard.

`__STRICT_ANSI__`
: GNU C defines this macro if and only if the `-ansi' switch was specified when GNU C was invoked. Its definition is the null string. This macro exists primarily to direct certain GNU header files not to define certain traditional Unix constructs which are incompatible with ISO C.

`__BASE_FILE__`
: This macro expands to the name of the main input file, in the form of a C string constant. This is the source file that was specified on the command line of the preprocessor or C compiler.

`__INCLUDE_LEVEL__`
: This macro expands to a decimal integer constant that represents the depth of nesting in include files. The value of this macro is incremented on every `#include' directive and decremented at the end of every included file. It starts out at 0, it's value within the base file specified on the command line.

`__VERSION__`
: This macro expands to a string constant which describes the version number of GNU C. The string is normally a sequence of decimal numbers separated by periods, such as `"2.6.0"'.

`__OPTIMIZE__`
: GNU CC defines this macro in optimizing compilations. It causes certain GNU header files to define alternative macro definitions for some system library functions. You should not refer to or test the definition of this macro unless you make very sure that programs will execute with the same effect regardless.

`__CHAR_UNSIGNED__`
: GNU C defines this macro if and only if the data type `char` is unsigned on the target machine. It exists to cause the standard header file `limits.h' to work correctly. You should not refer to this macro yourself; instead, refer to the standard macros defined in `limits.h'. The preprocessor uses this macro to determine whether or not to sign-extend large character constants written in octal; see section The `#if' Directive.

`__REGISTER_PREFIX__`
: This macro expands to a string (not a string constant) describing the prefix applied to CPU registers in assembler code. You can use it to write assembler code that is usable in multiple environments. For example, in the `m68k-aout' environment it expands to the null string, but in the `m68k-coff' environment it expands to the string `%'.

`__USER_LABEL_PREFIX__`
: Similar to `__REGISTER_PREFIX__`, but describes the prefix applied to user generated labels in assembler code. For example, in the `m68k-aout' environment it expands to the string `_', but in the `m68k-coff' environment it expands to the null string. This does not work with the `-mno-underscores' option that the i386 OSF/rose and m88k targets provide nor with the `-mcall*' options of the rs6000 System V Release 4 target.

Nonstandard Predefined Macros

The C preprocessor normally has several predefined macros that vary between machines because their purpose is to indicate what type of system and machine is in use. This manual, being for all systems and machines, cannot tell you exactly what their names are; instead, we offer a list of some typical ones. You can use `cpp -dM` to see the values of predefined macros; see section Invoking the C Preprocessor.

Some nonstandard predefined macros describe the operating system in use, with more or less specificity. For example,

`unix`
: `unix` is normally predefined on all Unix systems.

`BSD`
: `BSD` is predefined on recent versions of Berkeley Unix (perhaps only in version 4.3).

Other nonstandard predefined macros describe the kind of CPU, with more or less specificity. For example,

`vax`
: `vax` is predefined on Vax computers.

`mc68000`
: `mc68000` is predefined on most computers whose CPU is a Motorola 68000, 68010 or 68020.

`m68k`
: `m68k` is also predefined on most computers whose CPU is a 68000, 68010 or 68020; however, some makers use `mc68000` and some use `m68k`. Some predefine both names. What happens in GNU C depends on the system you are using it on.

`M68020`
: `M68020` has been observed to be predefined on some systems that use 68020 CPUs -- in addition to `mc68000` and `m68k`, which are less specific.

`_AM29K`
`_AM29000`
: Both `_AM29K` and `_AM29000` are predefined for the AMD 29000 CPU family.

`ns32000`
: `ns32000` is predefined on computers which use the National Semiconductor 32000 series CPU.

Yet other nonstandard predefined macros describe the manufacturer of the system. For example,

`sun`
: `sun` is predefined on all models of Sun computers.

`pyr`
: `pyr` is predefined on all models of Pyramid computers.

`sequent`
: `sequent` is predefined on all models of Sequent computers.

These predefined symbols are not only nonstandard, they are contrary to the ISO standard because their names do not start with underscores. Therefore, the option `-ansi` inhibits the definition of these symbols.

This tends to make `-ansi' useless, since many programs depend on the customary nonstandard predefined symbols. Even system header files check them and will generate incorrect declarations if they do not find the names that are expected. You might think that the header files supplied for the Uglix computer would not need to test what machine they are running on, because they can simply assume it is the Uglix; but often they do, and they do so using the customary names. As a result, very few C programs will compile with `-ansi'. We intend to avoid such problems on the GNU system.

What, then, should you do in an ISO C program to test the type of machine it will run on?

GNU C offers a parallel series of symbols for this purpose, whose names are made from the customary ones by adding `__' at the beginning and end. Thus, the symbol __vax__ would be available on a Vax, and so on.

The set of nonstandard predefined names in the GNU C preprocessor is controlled (when `cpp` is itself compiled) by the macro `CPP_PREDEFINES', which should be a string containing `-D' options, separated by spaces. For example, on the Sun 3, we use the following definition:

```
#define CPP_PREDEFINES "-Dmc68000 -Dsun -Dunix -Dm68k"
```

This macro is usually specified in `tm.h'.

Stringification

Stringification means turning a sequence of preprocessing tokens into a string literal. For example, stringifying `foo (z)' results in `"foo (z)"'.

In the C preprocessor, stringification is possible when macro arguments are substituted during macro expansion. When a parameter appears preceded by a `#' token in the replacement list of a function-like macro, it indicates that both tokens should be replaced with the stringification of the corresponding argument during expansion. The same argument may be substituted in other places in the definition without stringification if the argument name appears in those places with no preceding `#'.

Here is an example of a macro definition that uses stringification:

```
#define WARN_IF(EXP) \
do { if (EXP) \
        fprintf (stderr, "Warning: " #EXP "\n"); } \
while (0)
```

Here the argument for `EXP' is substituted once, as-is, into the `if' statement, and once, stringified, into the argument to `fprintf'. The `do' and `while (0)' are a kludge to make it possible to write `WARN_IF (*arg*);', which the resemblance of `WARN_IF' to a function would make C programmers want to do; see section Swallowing the Semicolon.

The stringification feature is limited to transforming the tokens of a macro argument into a string constant: there is no way to combine the argument with surrounding text and stringify it all together. The example

above shows how an equivalent result can be obtained in ISO Standard C, using the fact that adjacent string constants are concatenated by the C compiler to form a single string constant. The preprocessor stringifies the actual value of `EXP' into a separate string constant, resulting in text like

```
do { if (x == 0) \
        fprintf (stderr, "Warning: " "x == 0" "\n"); } \
while (0)
```

but the compiler then sees three consecutive string constants and concatenates them into one, producing effectively

```
do { if (x == 0) \
        fprintf (stderr, "Warning: x == 0\n"); } \
while (0)
```

Stringification in C involves more than putting double-quote characters around the fragment. The preprocessor backslash-escapes the surrounding quotes of string literals, and all backslashes within string and character constants, in order to get a valid C string constant with the proper contents. Thus, stringifying `p = "foo\n";' results in `"p = \"foo\\n\";"'. However, backslashes that are not inside string or character constants are not duplicated: `\n' by itself stringifies to `"\n"'.

Whitespace (including comments) in the text being stringified is handled according to precise rules. All leading and trailing whitespace is ignored. Any sequence of whitespace in the middle of the text is converted to a single space in the stringified result.

Concatenation

Concatenation means joining two strings into one. In the context of macro expansion, concatenation refers to joining two preprocessing tokens to form one. In particular, a token of a macro argument can be concatenated with another argument's token or with fixed text to produce a longer name. The longer name might be the name of a function, variable, type, or a C keyword; it might even be the name of another macro, in which case it will be expanded.

When you define a function-like or object-like macro, you request concatenation with the special operator `##' in the macro's replacement list. When the macro is called, any arguments are substituted without performing macro expansion, every `##' operator is deleted, and the two tokens on either side of it are concatenated to form a single token.

Consider a C program that interprets named commands. There probably needs to be a table of commands, perhaps an array of structures declared as follows:

```
struct command
{
  char *name;
  void (*function) ();
};

struct command commands[] =
```

```
{
  { "quit", quit_command},
  { "help", help_command},
  ...
};
```

It would be cleaner not to have to give each command name twice, once in the string constant and once in the function name. A macro which takes the name of a command as an argument can make this unnecessary. The string constant can be created with stringification, and the function name by concatenating the argument with `_command'. Here is how it is done:

```
#define COMMAND(NAME)  { #NAME, NAME ## _command }

struct command commands[] =
{
  COMMAND (quit),
  COMMAND (help),
  ...
};
```

The usual case of concatenation is concatenating two names (or a name and a number) into a longer name. This isn't the only valid case. It is also possible to concatenate two numbers (or a number and a name, such as `1.5' and `e3') into a number. Also, multi-character operators such as `+=' can be formed by concatenation. However, two tokens that don't together form a valid token cannot be concatenated. For example, concatenation of `x' on one side and `+' on the other is not meaningful because those two tokens do not form a valid preprocessing token when concatenated. UNDEFINED

Keep in mind that the C preprocessor converts comments to whitespace before macros are even considered. Therefore, you cannot create a comment by concatenating `/' and `*': the `/*' sequence that starts a comment is not a token, but rather the beginning of a comment. You can freely use comments next to `##' in a macro definition, or in arguments that will be concatenated, because the comments will be converted to spaces at first sight, and concatenation operates on tokens and so ignores whitespace.

Undefining Macros

To **undefine** a macro means to cancel its definition. This is done with the `#undef' directive. `#undef' is followed by the macro name to be undefined.

Like definition, undefinition occurs at a specific point in the source file, and it applies starting from that point. The name ceases to be a macro name, and from that point on it is treated by the preprocessor as if it had never been a macro name.

For example,

```
#define FOO 4
x = FOO;
#undef FOO
x = FOO;
```

expands into

```
x = 4;

x = FOO;
```

In this example, `FOO' had better be a variable or function as well as (temporarily) a macro, in order for the result of the expansion to be valid C code.

The same form of `#undef' directive will cancel definitions with arguments or definitions that don't expect arguments. The `#undef' directive has no effect when used on a name not currently defined as a macro.

Redefining Macros

Redefining a macro means defining (with `#define') a name that is already defined as a macro.

A redefinition is trivial if the new definition is transparently identical to the old one. You probably wouldn't deliberately write a trivial redefinition, but they can happen automatically when a header file is included more than once (see section Header Files), so they are accepted silently and without effect.

Nontrivial redefinition is considered likely to be an error, so it provokes a warning message from the preprocessor. However, sometimes it is useful to change the definition of a macro in mid-compilation. You can inhibit the warning by undefining the macro with `#undef' before the second definition.

In order for a redefinition to be trivial, the parameter names must match and be in the same order, and the new replacement list must exactly match the one already in effect, with two possible exceptions:

- Whitespace may be added or deleted at the beginning or the end of the replacement list. In a sense this is vacuous, since strictly such whitespace doesn't form part of the macro's expansion.
- Between tokens in the expansion, any two forms of whitespace are considered equivalent. In particular, whitespace may not be eliminated entirely, nor may it be added where there previously wasn't any.

Recall that a comment counts as whitespace.

As a particular case of the above, you may not redefine an object-like macro as a function-like macro, and vice-versa.

Poisoning Macros

Sometimes, there is an identifier that you want to remove completely from your program, and make sure that it never creeps back in. To enforce this, the `#pragma GCC poison' directive can be used. `#pragma GCC poison' is followed by a list of identifiers to poison, and takes effect for the rest of the source. You cannot `#undef' a poisoned identifier or test to see if it's defined with `#ifdef'.

For example,

```
#pragma GCC poison printf sprintf fprintf
sprintf(some_string, "hello");
```

will produce an error.

Pitfalls and Subtleties of Macros

In this section we describe some special rules that apply to macros and macro expansion, and point out certain cases in which the rules have counterintuitive consequences that you must watch out for.

- Misnesting: Macros can contain unmatched parentheses.
- Macro Parentheses: Why apparently superfluous parentheses may be necessary to avoid incorrect grouping.
- Swallow Semicolon: Macros that look like functions but expand into compound statements.
- Side Effects: Unsafe macros that cause trouble when arguments contain side effects.
- Self-Reference: Macros whose definitions use the macros' own names.
- Argument Prescan: Arguments are checked for macro calls before they are substituted.
- Cascaded Macros: Macros whose definitions use other macros.
- Newlines in Args: Sometimes line numbers get confused.

Improperly Nested Constructs

Recall that when a macro is called with arguments, the arguments are substituted into the macro body and the result is checked, together with the rest of the input file, for more macro calls.

It is possible to piece together a macro call coming partially from the macro body and partially from the arguments. For example,

```
#define double(x) (2*(x))
#define call_with_1(x) x(1)
```

would expand `` `call_with_1 (double)' `` into `` `(2*(1))' ``.

Macro definitions do not have to have balanced parentheses. By writing an unbalanced open parenthesis in a macro body, it is possible to create a macro call that begins inside the macro body but ends outside of it. For example,

```
#define strange(file) fprintf (file, "%s %d",
...
strange(stderr) p, 35)
```

This bizarre example expands to `` `fprintf (stderr, "%s %d", p, 35)'! ``

Unintended Grouping of Arithmetic

You may have noticed that in most of the macro definition examples shown above, each occurrence of a macro argument name had parentheses around it. In addition, another pair of parentheses usually surround the entire macro definition. Here is why it is best to write macros that way.

Suppose you define a macro as follows,

```
#define ceil_div(x, y) (x + y - 1) / y
```

whose purpose is to divide, rounding up. (One use for this operation is to compute how many `int' objects are needed to hold a certain number of `char' objects.) Then suppose it is used as follows:

```
a = ceil_div (b & c, sizeof (int));
```

This expands into

```
a = (b & c + sizeof (int) - 1) / sizeof (int);
```

which does not do what is intended. The operator-precedence rules of C make it equivalent to this:

```
a = (b & (c + sizeof (int) - 1)) / sizeof (int);
```

What we want is this:

```
a = ((b & c) + sizeof (int) - 1)) / sizeof (int);
```

Defining the macro as

```
#define ceil_div(x, y) ((x) + (y) - 1) / (y)
```

provides the desired result.

Unintended grouping can result in another way. Consider `sizeof ceil_div(1, 2)'. That has the appearance of a C expression that would compute the size of the type of `ceil_div (1, 2)', but in fact it means something very different. Here is what it expands to:

```
sizeof ((1) + (2) - 1) / (2)
```

This would take the size of an integer and divide it by two. The precedence rules have put the division outside the `sizeof' when it was intended to be inside.

Parentheses around the entire macro definition can prevent such problems. Here, then, is the recommended way to define `ceil_div':

```
#define ceil_div(x, y) (((x) + (y) - 1) / (y))
```

Swallowing the Semicolon

Often it is desirable to define a macro that expands into a compound statement. Consider, for example, the following macro, that advances a pointer (the argument `p' says where to find it) across whitespace characters:

```
#define SKIP_SPACES(p, limit)  \
{ register char *lim = (limit); \
  while (p != lim) {            \
    if (*p++ != ' ') {          \
      p--; break; }}}
```

Here backslash-newline is used to split the macro definition, which must be a single logical line, so that it resembles the way such C code would be laid out if not part of a macro definition.

A call to this macro might be `SKIP_SPACES (p, lim)'. Strictly speaking, the call expands to a compound statement, which is a complete statement with no need for a semicolon to end it. However, since it looks like a function call, it minimizes confusion if you can use it like a function call, writing a semicolon afterward, as in `SKIP_SPACES (p, lim);'

This can cause trouble before `else' statements, because the semicolon is actually a null statement. Suppose you write

```
if (*p != 0)
  SKIP_SPACES (p, lim);
else ...
```

The presence of two statements -- the compound statement and a null statement -- in between the `if' condition and the `else' makes invalid C code.

The definition of the macro `SKIP_SPACES' can be altered to solve this problem, using a `do ... while' statement. Here is how:

```
#define SKIP_SPACES(p, limit)     \
do { register char *lim = (limit); \
     while (p != lim) {            \
       if (*p++ != ' ') {          \
         p--; break; }}}           \
while (0)
```

Now `SKIP_SPACES (p, lim);' expands into

```
do {...} while (0);
```

which is one statement.

Duplication of Side Effects

Many C programs define a macro `min', for "minimum", like this:

```
#define min(X, Y)  ((X) < (Y) ? (X) : (Y))
```

When you use this macro with an argument containing a side effect, as shown here,

```
next = min (x + y, foo (z));
```

it expands as follows:

```
next = ((x + y) < (foo (z)) ? (x + y) : (foo (z)));
```

where `x + y' has been substituted for `X' and `foo (z)' for `Y'.

The function `foo' is used only once in the statement as it appears in the program, but the expression `foo (z)' has been substituted twice into the macro expansion. As a result, `foo' might be called two times when the statement is executed. If it has side effects or if it takes a long time to compute, the results might not be what you intended. We say that `min' is an **unsafe** macro.

The best solution to this problem is to define `min' in a way that computes the value of `foo (z)' only once. The C language offers no standard way to do this, but it can be done with GNU C extensions as follows:

```
#define min(X, Y)                       \
({ typeof (X) __x = (X), __y = (Y);   \
   (__x < __y) ? __x : __y; })
```

If you do not wish to use GNU C extensions, the only solution is to be careful when *using* the macro `min'. For example, you can calculate the value of `foo (z)', save it in a variable, and use that variable in `min':

```
#define min(X, Y)  ((X) < (Y) ? (X) : (Y))
...
{
  int tem = foo (z);
  next = min (x + y, tem);
}
```

(where we assume that `foo' returns type `int').

Self-Referential Macros

A **self-referential** macro is one whose name appears in its definition. A special feature of ISO Standard C is that the self-reference is not considered a macro call. It is passed into the preprocessor output unchanged.

Let's consider an example:

```
#define foo (4 + foo)
```

where `foo' is also a variable in your program.

Following the ordinary rules, each reference to `foo' will expand into `(4 + foo)'; then this will be rescanned and will expand into `(4 + (4 + foo))'; and so on until it causes a fatal error (memory full) in the preprocessor.

However, the special rule about self-reference cuts this process short after one step, at `(4 + foo)'. Therefore, this macro definition has the possibly useful effect of causing the program to add 4 to the value of `foo' wherever `foo' is referred to.

In most cases, it is a bad idea to take advantage of this feature. A person reading the program who sees that `foo' is a variable will not expect that it is a macro as well. The reader will come across the identifier `foo' in the program and think its value should be that of the variable `foo', whereas in fact the value is four greater.

The special rule for self-reference applies also to **indirect** self-reference. This is the case where a macro *x* expands to use a macro `y', and the expansion of `y' refers to the macro `x'. The resulting reference to `x' comes indirectly from the expansion of `x', so it is a self-reference and is not further expanded. Thus, after

```
#define x (4 + y)
#define y (2 * x)
```

`x' would expand into `(4 + (2 * x))'. Clear?

Suppose `y' is used elsewhere, not from the definition of `x'. Then the use of `x' in the expansion of `y' is not a self-reference because `x' is not "in progress". So it does expand. However, the expansion of `x' contains a reference to `y', and that is an indirect self-reference now because `y' is "in progress". The result is that `y' expands to `(2 * (4 + y))'.

This behavior is specified by the ISO C standard, so you may need to understand it.

Separate Expansion of Macro Arguments

We have explained that the expansion of a macro, including the substituted arguments, is re-scanned for macro calls to be expanded.

What really happens is more subtle: first each argument is scanned separately for macro calls. Then the resulting tokens are substituted into the macro body to produce the macro expansion, and the macro expansion is scanned again for macros to expand.

The result is that the arguments are scanned *twice* to expand macro calls in them.

Most of the time, this has no effect. If the argument contained any macro calls, they are expanded during the first scan. The result therefore contains no macro calls, so the second scan does not change it. If the argument were substituted as given, with no prescan, the single remaining scan would find the same macro calls and produce the same results.

You might expect the double scan to change the results when a self-referential macro is used in an argument of another macro (see section Self-Referential Macros): the self-referential macro would be expanded once in the first scan, and a second time in the second scan. However, this is not what happens. The self-references that do not expand in the first scan are marked so that they will not expand in the second scan either.

The prescan is not done when an argument is stringified or concatenated. Thus,

```
#define str(s) #s
#define foo 4
str (foo)
```

expands to `` `"foo"' ``. Once more, prescan has been prevented from having any noticeable effect.

More precisely, stringification and concatenation use the argument tokens as given without initially scanning for macros. The same argument would be used in expanded form if it is substituted elsewhere without stringification or concatenation.

```
#define str(s) #s lose(s)
#define foo 4
str (foo)
```

expands to `` `"foo" lose(4)' ``.

You might now ask, "Why mention the prescan, if it makes no difference? And why not skip it and make the preprocessor faster?" The answer is that the prescan does make a difference in three special cases:

- Nested calls to a macro.
- Macros that call other macros that stringify or concatenate.
- Macros whose expansions contain unshielded commas.

We say that **nested** calls to a macro occur when a macro's argument contains a call to that very macro. For example, if `` `f' `` is a macro that expects one argument, `` `f (f (1))' `` is a nested pair of calls to `` `f' ``. The desired expansion is made by expanding `` `f (1)' `` and substituting that into the definition of `` `f' ``. The prescan causes the expected result to happen. Without the prescan, `` `f (1)' `` itself would be substituted as an argument, and the inner use of `` `f' `` would appear during the main scan as an indirect self-reference and would not be expanded. Here, the prescan cancels an undesirable side effect (in the medical, not computational, sense of the term) of the special rule for self-referential macros.

Prescan causes trouble in certain other cases of nested macro calls. Here is an example:

```
#define foo  a,b
```

```
#define bar(x) lose(x)
#define lose(x) (1 + (x))

bar(foo)
```

We would like `bar(foo)' to turn into `(1 + (foo))', which would then turn into `(1 + (a,b))'. Instead, `bar(foo)' expands into `lose(a,b)', and you get an error because `lose' requires a single argument. In this case, the problem is easily solved by the same parentheses that ought to be used to prevent misnesting of arithmetic operations:

```
#define foo (a,b)
#define bar(x) lose((x))
```

The problem is more serious when the operands of the macro are not expressions; for example, when they are statements. Then parentheses are unacceptable because they would make for invalid C code:

```
#define foo { int a, b; ... }
```

In GNU C you can shield the commas using the `({...})' construct which turns a compound statement into an expression:

```
#define foo ({ int a, b; ... })
```

Or you can rewrite the macro definition to avoid such commas:

```
#define foo { int a; int b; ... }
```

There is also one case where prescan is useful. It is possible to use prescan to expand an argument and then stringify it -- if you use two levels of macros. Let's add a new macro `xstr' to the example shown above:

```
#define xstr(s) str(s)
#define str(s) #s
#define foo 4
xstr (foo)
```

This expands into `"4"', not `"foo"'. The reason for the difference is that the argument of `xstr' is expanded at prescan (because `xstr' does not specify stringification or concatenation of the argument). The result of prescan then forms the argument for `str'. `str' uses its argument without prescan because it performs stringification; but it cannot prevent or undo the prescanning already done by `xstr'.

Cascaded Use of Macros

A **cascade** of macros is when one macro's body contains a reference to another macro. This is very common practice. For example,

```
#define BUFSIZE 1020
#define TABLESIZE BUFSIZE
```

This is not at all the same as defining `TABLESIZE' to be `1020'. The `#define' for `TABLESIZE' uses exactly the body you specify -- in this case, `BUFSIZE' -- and does not check to see whether it too is the name of a macro.

It's only when you *use* `TABLESIZE' that the result of its expansion is checked for more macro names.

This makes a difference if you change the definition of `BUFSIZE' at some point in the source file. `TABLESIZE', defined as shown, will always expand using the definition of `BUFSIZE' that is currently in effect:

```
#define BUFSIZE 1020
#define TABLESIZE BUFSIZE
#undef BUFSIZE
#define BUFSIZE 37
```

Now `TABLESIZE' expands (in two stages) to `37'. (The `#undef' is to prevent any warning about the nontrivial redefinition of `BUFSIZE'.)

Newlines in Macro Arguments

The invocation of a function-like macro can extend over many logical lines. The ISO C standard requires that newlines within a macro invocation be treated as ordinary whitespace. This means that when the expansion of a function-like macro replaces its invocation, it appears on the same line as the macro name did. Thus line numbers emitted by the compiler or debugger refer to the line the invocation started on, which might be different to the line containing the argument causing the problem.

Here is an example illustrating this:

```
#define ignore_second_arg(a,b,c) a; c

ignore_second_arg (foo (),
                   ignored (),
                   syntax error);
```

The syntax error triggered by the tokens `syntax error' results in an error message citing line three -- the line of ignore_second_arg --- even though the problematic code comes from line five.

Conditionals

In a macro processor, a **conditional** is a directive that allows a part of the program to be ignored during compilation, on some conditions. In the C preprocessor, a conditional can test either an arithmetic expression or whether a name is defined as a macro.

A conditional in the C preprocessor resembles in some ways an `if' statement in C, but it is important to understand the difference between them. The condition in an `if' statement is tested during the execution of your program. Its purpose is to allow your program to behave differently from run to run, depending on the data it is operating on. The condition in a preprocessing conditional directive is tested when your

program is compiled. Its purpose is to allow different code to be included in the program depending on the situation at the time of compilation.

- Uses: What conditionals are for.
- Syntax: How conditionals are written.
- Deletion: Making code into a comment.
- Macros: Why conditionals are used with macros.
- Errors: Detecting inconsistent compilation parameters.

Why Conditionals are Used

Generally there are three kinds of reason to use a conditional.

- A program may need to use different code depending on the machine or operating system it is to run on. In some cases the code for one operating system may be erroneous on another operating system; for example, it might refer to library routines that do not exist on the other system. When this happens, it is not enough to avoid executing the invalid code: merely having it in the program makes it impossible to link the program and run it. With a preprocessing conditional, the offending code can be effectively excised from the program when it is not valid.
- You may want to be able to compile the same source file into two different programs. Sometimes the difference between the programs is that one makes frequent time-consuming consistency checks on its intermediate data, or prints the values of those data for debugging, while the other does not.
- A conditional whose condition is always false is a good way to exclude code from the program but keep it as a sort of comment for future reference.

Most simple programs that are intended to run on only one machine will not need to use preprocessing conditionals.

Syntax of Conditionals

A conditional in the C preprocessor begins with a **conditional directive**: `#if', `#ifdef' or `#ifndef'. See section Conditionals and Macros, for information on `#ifdef' and `#ifndef'; only `#if' is explained here.

- If: Basic conditionals using `#if' and `#endif'.
- Else: Including some text if the condition fails.
- Elif: Testing several alternative possibilities.

The `#if' Directive

The `#if' directive in its simplest form consists of

```
#if expression
controlled text
#endif /* expression */
```

The comment following the `#endif' is not required, but it is a good practice because it helps people match the `#endif' to the corresponding `#if'. Such comments should always be used, except in short conditionals that are not nested. In fact, you can put anything at all after the `#endif' and it will be ignored by the GNU C preprocessor, but only comments are acceptable in ISO Standard C.

expression is a C expression of integer type, subject to stringent restrictions. It may contain

- Integer constants, which are all regarded as `long` or `unsigned long`.
- Character constants, which are interpreted according to the character set and conventions of the machine and operating system on which the preprocessor is running. The GNU C preprocessor uses the C data type `char' for these character constants; therefore, whether some character codes are negative is determined by the C compiler used to compile the preprocessor. If it treats `char' as signed, then character codes large enough to set the sign bit will be considered negative; otherwise, no character code is considered negative.
- Arithmetic operators for addition, subtraction, multiplication, division, bitwise operations, shifts, comparisons, and logical operations (`&&' and `||'). The latter two obey the usual short-circuiting rules of standard C.
- Identifiers that are not macros, which are all treated as zero(!).
- Macro calls. All macro calls in the expression are expanded before actual computation of the expression's value begins.

Note that `sizeof' operators and `enum`-type values are not allowed. `enum`-type values, like all other identifiers that are not taken as macro calls and expanded, are treated as zero.

The *controlled text* inside of a conditional can include preprocessing directives. Then the directives inside the conditional are obeyed only if that branch of the conditional succeeds. The text can also contain other conditional groups. However, the `#if' and `#endif' directives must balance.

The `#else' Directive

The `#else' directive can be added to a conditional to provide alternative text to be used if the condition is false. This is what it looks like:

```
#if expression
text-if-true
#else /* Not expression */
text-if-false
#endif /* Not expression */
```

If *expression* is nonzero, and thus the *text-if-true* is active, then `#else' acts like a failing conditional and the *text-if-false* is ignored. Conversely, if the `#if' conditional fails, the *text-if-false* is considered included.

The `#elif' Directive

One common case of nested conditionals is used to check for more than two possible alternatives. For example, you might have

```
#if X == 1
...
#else /* X != 1 */
#if X == 2
...
#else /* X != 2 */
...
#endif /* X != 2 */
#endif /* X != 1 */
```

Another conditional directive, `#elif', allows this to be abbreviated as follows:

```
#if X == 1
...
#elif X == 2
...
#else /* X != 2 and X != 1*/
...
#endif /* X != 2 and X != 1*/
```

`#elif' stands for "else if". Like `#else', it goes in the middle of a `#if'-`#endif' pair and subdivides it; it does not require a matching `#endif' of its own. Like `#if', the `#elif' directive includes an expression to be tested.

The text following the `#elif' is processed only if the original `#if'-condition failed and the `#elif' condition succeeds. More than one `#elif' can go in the same `#if'-`#endif' group. Then the text after each `#elif' is processed only if the `#elif' condition succeeds after the original `#if' and any previous `#elif' directives within it have failed. `#else' is equivalent to `#elif 1', and `#else' is allowed after any number of `#elif' directives, but `#elif' may not follow `#else'.

Keeping Deleted Code for Future Reference

If you replace or delete a part of the program but want to keep the old code around as a comment for future reference, the easy way to do this is to put `#if 0' before it and `#endif' after it. This is better than using comment delimiters `/*' and `*/' since those won't work if the code already contains comments (C comments do not nest).

This works even if the code being turned off contains conditionals, but they must be entire conditionals (balanced `#if' and `#endif').

Conversely, do not use `#if 0' for comments which are not C code. Use the comment delimiters `/*' and `*/' instead. The interior of `#if 0' must consist of complete tokens; in particular, single-quote

characters must balance. Comments often contain unbalanced single-quote characters (known in English as apostrophes). These confuse `#if 0'. They do not confuse `/*'.

Conditionals and Macros

Conditionals are useful in connection with macros or assertions, because those are the only ways that an expression's value can vary from one compilation to another. A `#if' directive whose expression uses no macros or assertions is equivalent to `#if 1' or `#if 0'; you might as well determine which one, by computing the value of the expression yourself, and then simplify the program.

For example, here is a conditional that tests the expression `BUFSIZE == 1020', where `BUFSIZE' must be a macro.

```
#if BUFSIZE == 1020
  printf ("Large buffers!\n");
#endif /* BUFSIZE is large */
```

(Programmers often wish they could test the size of a variable or data type in `#if', but this does not work. The preprocessor does not understand `sizeof', or typedef names, or even the type keywords such as `int'.)

The special operator `defined' is used in `#if' and `#elif' expressions to test whether a certain name is defined as a macro. Either `defined *name*' or `defined (*name*)' is an expression whose value is 1 if *name* is defined as macro at the current point in the program, and 0 otherwise. To the `defined' operator it makes no difference what the definition of the macro is; all that matters is whether there is a definition. Thus, for example,

```
#if defined (vax) || defined (ns16000)
```

would succeed if either of the names `vax' and `ns16000' is defined as a macro. You can test the same condition using assertions (see section Assertions), like this:

```
#if #cpu (vax) || #cpu (ns16000)
```

If a macro is defined and later undefined with `#undef', subsequent use of the `defined' operator returns 0, because the name is no longer defined. If the macro is defined again with another `#define', `defined' will recommence returning 1.

If the `defined' operator appears as a result of a macro expansion, the C standard says the behavior is undefined. GNU cpp treats it as a genuine `defined' operator and evaluates it normally. It will warn wherever your code uses this feature if you use the command-line option `-pedantic', since other compilers may handle it differently.

Conditionals that test whether a single macro is defined are very common, so there are two special short conditional directives for this case.

`#ifdef` *name*
: is equivalent to `` `#if defined (name)' ``.

`#ifndef` *name*
: is equivalent to `` `#if ! defined (name)' ``.

Macro definitions can vary between compilations for several reasons.

- Some macros are predefined on each kind of machine. For example, on a Vax, the name `` `vax' `` is a predefined macro. On other machines, it would not be defined.
- Many more macros are defined by system header files. Different systems and machines define different macros, or give them different values. It is useful to test these macros with conditionals to avoid using a system feature on a machine where it is not implemented.
- Macros are a common way of allowing users to customize a program for different machines or applications. For example, the macro `` `BUFSIZE' `` might be defined in a configuration file for your program that is included as a header file in each source file. You would use `` `BUFSIZE' `` in a preprocessing conditional in order to generate different code depending on the chosen configuration.
- Macros can be defined or undefined with `` `-D' `` and `` `-U' `` command options when you compile the program. You can arrange to compile the same source file into two different programs by choosing a macro name to specify which program you want, writing conditionals to test whether or how this macro is defined, and then controlling the state of the macro with compiler command options. See section Invoking the C Preprocessor.

The `#error' and `#warning' Directives

The directive `` `#error' `` causes the preprocessor to report a fatal error. The tokens forming the rest of the line following `` `#error' `` are used as the error message, and not macro-expanded. Internal whitespace sequences are each replaced with a single space. The line must consist of complete tokens.

You would use `` `#error' `` inside of a conditional that detects a combination of parameters which you know the program does not properly support. For example, if you know that the program will not run properly on a Vax, you might write

```
#ifdef __vax__
#error "Won't work on Vaxen.  See comments at get_last_object."
#endif
```

See section Nonstandard Predefined Macros, for why this works.

If you have several configuration parameters that must be set up by the installation in a consistent way, you can use conditionals to detect an inconsistency and report it with `` `#error' ``. For example,

```
#if HASH_TABLE_SIZE % 2 == 0 || HASH_TABLE_SIZE % 3 == 0 \
    || HASH_TABLE_SIZE % 5 == 0
#error HASH_TABLE_SIZE should not be divisible by a small prime
#endif
```

The directive `#warning' is like the directive `#error', but causes the preprocessor to issue a warning and continue preprocessing. The tokens following `#warning' are used as the warning message, and not macro-expanded.

You might use `#warning' in obsolete header files, with a message directing the user to the header file which should be used instead.

Assertions

Assertions are a more systematic alternative to macros in writing conditionals to test what sort of computer or system the compiled program will run on. Assertions are usually predefined, but you can define them with preprocessing directives or command-line options.

The macros traditionally used to describe the type of target are not classified in any way according to which question they answer; they may indicate a hardware architecture, a particular hardware model, an operating system, a particular version of an operating system, or specific configuration options. These are jumbled together in a single namespace. In contrast, each assertion consists of a named question and an answer. The question is usually called the **predicate**. An assertion looks like this:

```
#predicate (answer)
```

You must use a properly formed identifier for *predicate*. The value of *answer* can be any sequence of words; all characters are significant except for leading and trailing whitespace, and differences in internal whitespace sequences are ignored. (This is similar to the rules governing macro redefinition.) Thus, `x + y' is different from `x+y' but equivalent to ` x + y '. `)' is not allowed in an answer.

Here is a conditional to test whether the answer *answer* is asserted for the predicate *predicate*:

```
#if #predicate (answer)
```

There may be more than one answer asserted for a given predicate. If you omit the answer, you can test whether *any* answer is asserted for *predicate*:

```
#if #predicate
```

Most of the time, the assertions you test will be predefined assertions. GNU C provides three predefined predicates: `system`, `cpu`, and `machine`. `system` is for assertions about the type of software, `cpu` describes the type of computer architecture, and `machine` gives more information about the computer. For example, on a GNU system, the following assertions would be true:

```
#system (gnu)
#system (mach)
#system (mach 3)
#system (mach 3.subversion)
#system (hurd)
#system (hurd version)
```

and perhaps others. The alternatives with more or less version information let you ask more or less detailed questions about the type of system software.

On a Unix system, you would find `#system (unix)` and perhaps one of: `#system (aix)`, `#system (bsd)`, `#system (hpux)`, `#system (lynx)`, `#system (mach)`, `#system (posix)`, `#system (svr3)`, `#system (svr4)`, or `#system (xpg4)` with possible version numbers following.

Other values for `system` are `#system (mvs)` and `#system (vms)`.

Portability note: Many Unix C compilers provide only one answer for the `system` assertion: `#system (unix)`, if they support assertions at all. This is less than useful.

An assertion with a multi-word answer is completely different from several assertions with individual single-word answers. For example, the presence of `system (mach 3.0)` does not mean that `system (3.0)` is true. It also does not directly imply `system (mach)`, but in GNU C, that last will normally be asserted as well.

The current list of possible assertion values for `cpu` is: `#cpu (a29k)`, `#cpu (alpha)`, `#cpu (arm)`, `#cpu (clipper)`, `#cpu (convex)`, `#cpu (elxsi)`, `#cpu (tron)`, `#cpu (h8300)`, `#cpu (i370)`, `#cpu (i386)`, `#cpu (i860)`, `#cpu (i960)`, `#cpu (m68k)`, `#cpu (m88k)`, `#cpu (mips)`, `#cpu (ns32k)`, `#cpu (hppa)`, `#cpu (pyr)`, `#cpu (ibm032)`, `#cpu (rs6000)`, `#cpu (sh)`, `#cpu (sparc)`, `#cpu (spur)`, `#cpu (tahoe)`, `#cpu (vax)`, `#cpu (we32000)`.

You can create assertions within a C program using `#assert', like this:

```
#assert predicate (answer)
```

(Note the absence of a `#' before *predicate*.)

Each time you do this, you assert a new true answer for *predicate*. Asserting one answer does not invalidate previously asserted answers; they all remain true. The only way to remove an answer is with `#unassert'. `#unassert' has the same syntax as `#assert'. You can also remove all answers to a *predicate* like this:

```
#unassert predicate
```

You can also add or cancel assertions using command options when you run `gcc` or `cpp`. See section Invoking the C Preprocessor.

Combining Source Files

One of the jobs of the C preprocessor is to inform the C compiler of where each line of C code came from: which source file and which line number.

C code can come from multiple source files if you use `#include'; both `#include' and the use of conditionals and macros can cause the line number of a line in the preprocessor output to be different from the line's number in the original source file. You will appreciate the value of making both the C compiler (in error messages) and symbolic debuggers such as GDB use the line numbers in your source file.

The C preprocessor builds on this feature by offering a directive by which you can control the feature explicitly. This is useful when a file for input to the C preprocessor is the output from another program such as the `bison` parser generator, which operates on another file that is the true source file. Parts of the output from `bison` are generated from scratch, other parts come from a standard parser file. The rest are copied nearly verbatim from the source file, but their line numbers in the `bison` output are not the same as their original line numbers. Naturally you would like compiler error messages and symbolic debuggers to know the original source file and line number of each line in the `bison` input.

`bison` arranges this by writing `#line' directives into the output file. `#line' is a directive that specifies the original line number and source file name for subsequent input in the current preprocessor input file. `#line' has three variants:

`#line` *linenum*
: Here *linenum* is a decimal integer constant. This specifies that the line number of the following line of input, in its original source file, was *linenum*.

`#line` *linenum filename*
: Here *linenum* is a decimal integer constant and *filename* is a string constant. This specifies that the following line of input came originally from source file *filename* and its line number there was *linenum*. Keep in mind that *filename* is not just a file name; it is surrounded by double-quote characters so that it looks like a string constant.

`#line` *anything else*
: *anything else* is checked for macro calls, which are expanded. The result should be a decimal integer constant followed optionally by a string constant, as described above.

`#line' directives alter the results of the `__FILE__' and `__LINE__' predefined macros from that point on. See section Standard Predefined Macros.

The output of the preprocessor (which is the input for the rest of the compiler) contains directives that look much like `#line' directives. They start with just `#' instead of `#line', but this is followed by a line number and file name as in `#line'. See section C Preprocessor Output.

Miscellaneous Preprocessing Directives

This section describes some additional, rarely used, preprocessing directives.

The ISO standard specifies that the effect of the `#pragma' directive is implementation-defined. The GNU C preprocessor recognizes some pragmas, and passes unrecognized ones through to the preprocessor output, so they are available to the compilation pass.

In line with the C99 standard, which introduces a STDC namespace for C99 pragmas, the preprocessor introduces a GCC name space for GCC pragmas. Supported GCC preprocessor pragmas are of the form `#pragma GCC ...'. For backwards compatibility previously supported pragmas are also recognized without the `GCC' prefix, however that use is deprecated. Pragmas that are already deprecated are not recognized with a `GCC' prefix.

The `#pragma GCC dependency' allows you to check the relative dates of the current file and another file. If the other file is more recent than the current file, a warning is issued. This is useful if the include file is derived from the other file, and should be regenerated. The other file is searched for using the normal include search path. Optional trailing text can be used to give more information in the warning message.

```
#pragma GCC dependency "parse.y"
#pragma GCC dependency "/usr/include/time.h" rerun /path/to/fixincludes
```

The C99 standard also introduces the `_Pragma' operator. The syntax is `_Pragma (string-literal)', where `string-literal' can be either a normal or wide-character string literal. It is destringized, by replacing all `\\' with a single `\' and all `\"' with a `"'. The result is then processed as if it had appeared as the right hand side of a `#pragma' directive. For example,

```
_Pragma ("GCC dependency \"parse.y\"")
```

has the same effect as `#pragma GCC dependency "parse.y"'. The same effect could be achieved using macros, for example

```
#define DO_PRAGMA(x) _Pragma (#x)
DO_PRAGMA (GCC dependency "parse.y")
```

The standard is unclear on where a `_Pragma' operator can appear. The preprocessor accepts it even within a preprocessing conditional directive like `#if'. To be safe, you are probably best keeping it out of directives other than `#define', and putting it on a line of its own.

The `#ident' directive is supported for compatibility with certain other systems. It is followed by a line of text. On some systems, the text is copied into a special place in the object file; on most systems, the text is ignored and this directive has no effect. Typically `#ident' is only used in header files supplied with those systems where it is meaningful.

The **null directive** consists of a `#' followed by a newline, with only whitespace (including comments) in between. A null directive is understood as a preprocessing directive but has no effect on the preprocessor output. The primary significance of the existence of the null directive is that an input line consisting of just a `#' will produce no output, rather than a line of output containing just a `#'. Supposedly some old C programs contain such lines.

C Preprocessor Output

The output from the C preprocessor looks much like the input, except that all preprocessing directive lines have been replaced with blank lines and all comments with spaces.

The ISO standard specifies that it is implementation defined whether a preprocessor preserves whitespace between tokens, or replaces it with e.g. a single space. In the GNU C preprocessor, whitespace between tokens is collapsed to become a single space, with the exception that the first token on a non-directive line is preceded with sufficient spaces that it appears in the same column in the preprocessed output that it appeared in in the original source file. This is so the output is easy to read. See section Undefined Behavior and Deprecated Features.

Source file name and line number information is conveyed by lines of the form

```
# linenum filename flags
```

which are inserted as needed into the output (but never within a string or character constant), and in place of long sequences of empty lines. Such a line means that the following line originated in file *filename* at line *linenum*.

After the file name comes zero or more flags, which are `` `1' ``, `` `2' ``, `` `3' ``, or `` `4' ``. If there are multiple flags, spaces separate them. Here is what the flags mean:

`` `1' ``
: This indicates the start of a new file.

`` `2' ``
: This indicates returning to a file (after having included another file).

`` `3' ``
: This indicates that the following text comes from a system header file, so certain warnings should be suppressed.

`` `4' ``
: This indicates that the following text should be treated as C.

Implementation-defined Behavior and Implementation Limits

The ISO C standard mandates that implementations document various aspects of preprocessor behavior. You should try to avoid undue reliance on behaviour described here, as it is possible that it will change subtly in future implementations.

- The mapping of physical source file multi-byte characters to the execution character set. Currently, GNU cpp only supports character sets that are strict supersets of ASCII, and performs no translation of characters.
- Non-empty sequences of whitespace characters. Each whitespace sequence is not preserved, but collapsed to a single space. For aesthetic reasons, the first token on each non-directive line of output is preceded with sufficient spaces that it appears in the same column as it did in the original source file.
- The numeric value of character constants in preprocessor expressions. The preprocessor interprets character constants in preprocessing directives on the host machine. Expressions outside

preprocessing directives are compiled to be interpreted on the target machine. In the normal case of a native compiler, these two environments are the same and so character constants will be evaluated identically in both cases. However, in the case of a cross compiler, the values may be different. Multi-character character constants are interpreted a character at a time, shifting the previous result left by the number of bits per character on the host, and adding the new character. For example, 'ab' on an 8-bit host would be interpreted as 'a' * 256 + 'b'. If there are more characters in the constant than can fit in the widest native integer type on the host, usually a `long', the behavior is undefined. Evaluation of wide character constants is not properly implemented yet.

- Source file inclusion. For a discussion on how the preprocessor locates header files, see section How `#include' Works.
- Interpretation of the filename resulting from a macro-expanded `#include' directive. If the macro expands to a string literal, the `#include' directive is processed as if the string had been specified directly. Otherwise, the macro must expand to a token stream beginning with a `<' token and including a `>' token. In this case, the tokens between the `<' and the first `>' are combined to form the filename to be included. Any whitespace between tokens is reduced to a single space; then any space after the initial `<' is retained, but a trailing space before the closing `>' is ignored. In either case, if any excess tokens remain, an error occurs and the directive is not processed.
- Treatment of a `#pragma' directive that after macro-expansion results in a standard pragma. The pragma is processed as if it were a normal standard pragma.

The following documents internal limits of GNU cpp.

- Nesting levels of `#include' files. We impose an arbitrary limit of 200 levels, to avoid runaway recursion. The standard requires at least 15 levels.
- Nesting levels of conditional inclusion. The C standard mandates this be at least 63. The GNU C preprocessor is limited only by available memory.
- Levels of parenthesised expressions within a full expression. The C standard requires this to be at least 63. In preprocessor conditional expressions it is limited only by available memory.
- Significant initial characters in an identifier or macro name. The preprocessor treats all characters as significant. The C standard requires only that the first 63 be significant.
- Number of macros simultaneously defined in a single translation unit. The standard requires at least 4095 be possible; GNU cpp is limited only by available memory.
- Number of parameters in a macro definition and arguments in a macro call. We allow USHRT_MAX, which is normally 65,535, and above the minimum of 127 required by the standard.
- Number of characters on a logical source line. The C standard requires a minimum of 4096 be permitted. GNU cpp places no limits on this, but you may get incorrect column numbers reported in diagnostics for lines longer than 65,535 characters.

Undefined Behavior and Deprecated Features

This section details GNU C preprocessor behavior that is subject to change or deprecated. You are *strongly advised* to write your software so it does not rely on anything described here; future versions of the preprocessor may subtly change such behavior or even remove the feature altogether.

Preservation of the form of whitespace between tokens is unlikely to change from current behavior (section C Preprocessor Output), but you are advised not to rely on it.

The following are undocumented and subject to change:-

- Precedence of ## operators with respect to each other Whether a sequence of ## operators is evaluated left-to-right, right-to-left or indeed in a consistent direction at all is not specified. An example of where this might matter is pasting the arguments `1', `e' and `-2'. This would be fine for left-to-right pasting, but right-to-left pasting would produce an invalid token `e-2'. It is possible to guarantee precedence by suitable use of nested macros.
- Precedence of # operator with respect to the ## operator Which of these two operators is evaluated first is not specified.

The following features are in flux and should not be used in portable code:

- Optional argument when invoking rest argument macros As an extension, GCC permits you to omit the variable arguments entirely when you use a variable argument macro. This works whether or not you give the variable argument a name. For example, the two macro invocations in the example below expand to the same thing:
- ```
  #define debug(format, ...) printf (format, __VA_ARGS__)
  ```
- ```
  debug("string");       /* Not permitted by C standard.  */
  ```
- ```
  debug("string",);      /* OK.  */
  ```

  This extension will be preserved, but the special behavior of `##' in this context has changed in the past and may change again in the future.

- ## swallowing preceding text in rest argument macros Formerly, in a macro expansion, if `##' appeared before a variable arguments parameter, and the set of tokens specified for that argument in the macro invocation was empty, previous versions of the GNU C preprocessor would back up and remove the preceding sequence of non-whitespace characters (**not** the preceding token). This extension is in direct conflict with the 1999 C standard and has been drastically pared back. In the current version of the preprocessor, if `##' appears between a comma and a variable arguments parameter, and the variable argument is omitted entirely, the comma will be removed from the expansion. If the variable argument is empty, or the token before `##' is not a comma, then `##' behaves as a normal token paste. Portable code should avoid this extension at all costs.

The following features are deprecated and will likely be removed at some point in the future:-

- Attempting to paste two tokens which together do not form a valid preprocessing token The preprocessor currently warns about this and outputs the two tokens adjacently, which is probably the behavior the programmer intends. It may not work in future, though. Most of the time, when you get this warning, you will find that `##' is being used superstitiously, to guard against whitespace appearing between two tokens. It is almost always safe to delete the `##'.
- #pragma once This pragma was once used to tell the preprocessor that it need not include a file more than once. It is now obsolete and should not be used at all.
  ```

- #pragma poison This pragma has been superseded by `#pragma GCC poison'. See section Poisoning Macros.
- Multi-line string literals in directives The GNU C preprocessor currently allows newlines in string literals within a directive. This is forbidden by the C standard and will eventually be removed. (Multi-line string literals in open text are still supported.)
- Preprocessing things which are not C The C preprocessor is intended to be used only with C, C++, and Objective C source code. In the past, it has been abused as a general text processor. It will choke on input which is not lexically valid C; for example, apostrophes will be interpreted as the beginning of character constants, and cause errors. Also, you cannot rely on it preserving characteristics of the input which are not significant to C-family languages. For instance, if a Makefile is preprocessed, all the hard tabs will be lost, and the Makefile will not work. Having said that, you can often get away with using cpp on things which are not C. Other Algol-ish programming languages are often safe (Pascal, Ada, ...) and so is assembly, with caution. `-traditional' mode is much more permissive, and can safely be used with e.g. Fortran. Many of the problems go away if you write C or C++ style comments instead of native language comments, and if you avoid elaborate macros. Wherever possible, you should use a preprocessor geared to the language you are writing in. Modern versions of the GNU assembler have macro facilities. Most high level programming languages have their own conditional compilation and inclusion mechanism. If all else fails, try a true general text processor, such as See section `Top' in *GNU `m4'*.

Invoking the C Preprocessor

Most often when you use the C preprocessor you will not have to invoke it explicitly: the C compiler will do so automatically. However, the preprocessor is sometimes useful on its own.

The C preprocessor expects two file names as arguments, *infile* and *outfile*. The preprocessor reads *infile* together with any other files it specifies with `#include'. All the output generated by the combined input files is written in *outfile*.

Either *infile* or *outfile* may be `-', which as *infile* means to read from standard input and as *outfile* means to write to standard output. Also, if either file is omitted, it means the same as if `-' had been specified for that file.

Here is a table of command options accepted by the C preprocessor. These options can also be given when compiling a C program; they are passed along automatically to the preprocessor when it is invoked by the compiler.

`-P'
: Inhibit generation of `#'-lines with line-number information in the output from the preprocessor. This might be useful when running the preprocessor on something that is not C code and will be sent to a program which might be confused by the `#'-lines. See section C Preprocessor Output.

`-C'
: Do not discard comments. All comments are passed through to the output file, except for comments in processed directives, which are deleted along with the directive. Comments appearing in the expansion list of a macro will be preserved, and appear in place wherever the macro is invoked. You

should be prepared for side effects when using `-C'; it causes the preprocessor to treat comments as tokens in their own right. For example, macro redefinitions that were trivial when comments were replaced by a single space might become significant when comments are retained. Also, comments appearing at the start of what would be a directive line have the effect of turning that line into an ordinary source line, since the first token on the line is no longer a `#'.

`-traditional'

Try to imitate the behavior of old-fashioned C, as opposed to ISO C.

- Traditional macro expansion pays no attention to single-quote or double-quote characters; macro argument symbols are replaced by the argument values even when they appear within apparent string or character constants.
- Traditionally, it is permissible for a macro expansion to end in the middle of a string or character constant. The constant continues into the text surrounding the macro call.
- However, traditionally the end of the line terminates a string or character constant, with no error.
- In traditional C, a comment is equivalent to no text at all. (In ISO C, a comment counts as whitespace.)
- Traditional C does not have the concept of a "preprocessing number". It considers `1.0e+4' to be three tokens: `1.0e', `+', and `4'.
- A macro is not suppressed within its own definition, in traditional C. Thus, any macro that is used recursively inevitably causes an error.
- The character `#' has no special meaning within a macro definition in traditional C.
- In traditional C, the text at the end of a macro expansion can run together with the text after the macro call, to produce a single token. (This is impossible in ISO C.)
- None of the GNU extensions to the preprocessor are available in `-traditional' mode.

Use the `-traditional' option when preprocessing Fortran code, so that single-quotes and double-quotes within Fortran comment lines (which are generally not recognized as such by the preprocessor) do not cause diagnostics about unterminated character or string constants. However, this option does not prevent diagnostics about unterminated comments when a C-style comment appears to start, but not end, within Fortran-style commentary. So, the following Fortran comment lines are accepted with `-traditional':

```
C This isn't an unterminated character constant
C Neither is "20000000000, an octal constant
C in some dialects of Fortran
```

However, this type of comment line will likely produce a diagnostic, or at least unexpected output from the preprocessor, due to the unterminated comment:

```
C Some Fortran compilers accept /* as starting
C an inline comment.
```

Note that `g77` automatically supplies the `-traditional' option when it invokes the preprocessor. However, a future version of `g77` might use a different, more-Fortran-aware preprocessor in place of `cpp`.

`-trigraphs'

Process ISO standard trigraph sequences. These are three-character sequences, all starting with `??', that are defined by ISO C to stand for single characters. For example, `??/' stands for `\', so `'??/n'' is a character constant for a newline. By default, GCC ignores trigraphs, but in

standard-conforming modes it converts them. See the `-std' option. The nine trigraph sequences are

```
`??('
-> `['
`??)'
-> `]'
`??<'
-> `{'
`??>'
-> `}'
`??='
-> `#'
`??/'
-> `\'
`??''
-> `^'
`??!'
-> `|'
`??-'
-> `~'
```

Trigraph support is not popular, so many compilers do not implement it properly. Portable code should not rely on trigraphs being either converted or ignored.

`-pedantic'
: Issue warnings required by the ISO C standard in certain cases such as when text other than a comment follows `#else' or `#endif'.

`-pedantic-errors'
: Like `-pedantic', except that errors are produced rather than warnings.

`-Wcomment'
`-Wcomments'
: (Both forms have the same effect). Warn whenever a comment-start sequence `/*' appears in a `/*' comment, or whenever a backslash-newline appears in a `//' comment.

`-Wtrigraphs'
: Warn if any trigraphs are encountered. This option used to take effect only if `-trigraphs' was also specified, but now works independently. Warnings are not given for trigraphs within comments, as we feel this is obnoxious.

`-Wwhite-space'
: Warn about possible white space confusion, e.g. white space between a backslash and a newline.

`-Wall'
: Requests `-Wcomment', `-Wtrigraphs', and `-Wwhite-space' (but not `-Wtraditional' or `-Wundef').

`-Wtraditional'
: Warn about certain constructs that behave differently in traditional and ISO C.

`-Wundef'
: Warn if an undefined identifier is evaluated in an `#if' directive.

`-I *directory*'

Add the directory *directory* to the head of the list of directories to be searched for header files (see section The `#include' Directive). This can be used to override a system header file, substituting your own version, since these directories are searched before the system header file directories. If you use more than one `-I' option, the directories are scanned in left-to-right order; the standard system directories come after.

`-I-'
Any directories specified with `-I' options before the `-I-' option are searched only for the case of `#include "*file*"'; they are not searched for `#include <*file*>'. If additional directories are specified with `-I' options after the `-I-', these directories are searched for all `#include' directives. In addition, the `-I-' option inhibits the use of the current directory as the first search directory for `#include "*file*"'. Therefore, the current directory is searched only if it is requested explicitly with `-I.'. Specifying both `-I-' and `-I.' allows you to control precisely which directories are searched before the current one and which are searched after.

`-nostdinc'
Do not search the standard system directories for header files. Only the directories you have specified with `-I' options (and the current directory, if appropriate) are searched.

`-nostdinc++'
Do not search for header files in the C++-specific standard directories, but do still search the other standard directories. (This option is used when building the C++ library.)

`-remap'
When searching for a header file in a directory, remap file names if a file named `header.gcc' exists in that directory. This can be used to work around limitations of file systems with file name restrictions. The `header.gcc' file should contain a series of lines with two tokens on each line: the first token is the name to map, and the second token is the actual name to use.

`-D *name*'
Predefine *name* as a macro, with definition `1'.

`-D *name*=*definition*'
Predefine *name* as a macro, with definition *definition*. There are no restrictions on the contents of *definition*, but if you are invoking the preprocessor from a shell or shell-like program you may need to use the shell's quoting syntax to protect characters such as spaces that have a meaning in the shell syntax. If you use more than one `-D' for the same *name*, the rightmost definition takes effect.

`-U *name*'
Do not predefine *name*. If both `-U' and `-D' are specified for one name, whichever one appears later on the command line wins.

`-undef'
Do not predefine any nonstandard macros.

`-gcc'
Define the macros __*GNUC*__, __*GNUC_MINOR*__ and __*GNUC_PATCHLEVEL*__. These are defined automatically when you use `gcc -E'; you can turn them off in that case with `-no-gcc'.

`-A *predicate*=*answer*'
Make an assertion with the predicate *predicate* and answer *answer*. This form is preferred to the older form `-A *predicate*(*answer*)', which is still supported, because it does not use shell special characters. See section Assertions.

`-A -*predicate*=*answer*'

Disable an assertion with the predicate *predicate* and answer *answer*. Specifying no predicate, by `-A-' or `-A -', disables all predefined assertions and all assertions preceding it on the command line; and also undefines all predefined macros and all macros preceding it on the command line.

`-dM'
Instead of outputting the result of preprocessing, output a list of `#define' directives for all the macros defined during the execution of the preprocessor, including predefined macros. This gives you a way of finding out what is predefined in your version of the preprocessor; assuming you have no file `foo.h', the command

```
touch foo.h; cpp -dM foo.h
```

will show the values of any predefined macros.

`-dD'
Like `-dM' except in two respects: it does *not* include the predefined macros, and it outputs *both* the `#define' directives and the result of preprocessing. Both kinds of output go to the standard output file.

`-dN'
Like `-dD', but emit only the macro names, not their expansions.

`-dI'
Output `#include' directives in addition to the result of preprocessing.

`-M [-MG]'
Instead of outputting the result of preprocessing, output a rule suitable for `make` describing the dependencies of the main source file. The preprocessor outputs one `make` rule containing the object file name for that source file, a colon, and the names of all the included files. If there are many included files then the rule is split into several lines using `\'-newline. `-MG' says to treat missing header files as generated files and assume they live in the same directory as the source file. It must be specified in addition to `-M'. This feature is used in automatic updating of makefiles.

`-MM [-MG]'
Like `-M' but mention only the files included with `#include "*file*"'. System header files included with `#include <*file*>' are omitted.

`-MD *file*'
Like `-M' but the dependency information is written to *file*. This is in addition to compiling the file as specified -- `-MD' does not inhibit ordinary compilation the way `-M' does. When invoking `gcc`, do not specify the *file* argument. `gcc` will create file names made by replacing ".c" with ".d" at the end of the input file names. In Mach, you can use the utility `md` to merge multiple dependency files into a single dependency file suitable for using with the `make' command.

`-MMD *file*'
Like `-MD' except mention only user header files, not system header files.

`-H'
Print the name of each header file used, in addition to other normal activities.

`-imacros *file*'
Process *file* as input, discarding the resulting output, before processing the regular input file. Because the output generated from *file* is discarded, the only effect of `-imacros *file*' is to make the macros defined in *file* available for use in the main input.

`-include *file*'
Process *file* as input, and include all the resulting output, before processing the regular input file.

`-idirafter *dir*'

Add the directory *dir* to the second include path. The directories on the second include path are searched when a header file is not found in any of the directories in the main include path (the one that `-I' adds to).

`-iprefix *prefix*'
Specify *prefix* as the prefix for subsequent `-iwithprefix' options. If the prefix represents a directory, you should include the final `/'.

`-iwithprefix *dir*'
Add a directory to the second include path. The directory's name is made by concatenating *prefix* and *dir*, where *prefix* was specified previously with `-iprefix'.

`-isystem *dir*'
Add a directory to the beginning of the second include path, marking it as a system directory, so that it gets the same special treatment as is applied to the standard system directories. See section System Headers.

`-x c'
`-x c++'
`-x objective-c'
`-x assembler-with-cpp'
Specify the source language: C, C++, Objective-C, or assembly. This has nothing to do with standards conformance or extensions; it merely selects which base syntax to expect. If you give none of these options, cpp will deduce the language from the extension of the source file: `.c', `.cc', `.m', or `.S'. Some other common extensions for C++ and assembly are also recognized. If cpp does not recognize the extension, it will treat the file as C; this is the most generic mode. **Note:** Previous versions of cpp accepted a `-lang' option which selected both the language and the standards conformance level. This option has been removed, because it conflicts with the `-l' option.

`-std=*standard*'
`-ansi'
Specify the standard to which the code should conform. Currently cpp only knows about the standards for C; other language standards will be added in the future. *standard* may be one of:
`iso9899:1990`
`c89`
The ISO C standard from 1990. `c89' is the customary shorthand for this version of the standard. The `-ansi' option is equivalent to `-std=c89'.
`iso9899:199409`
The 1990 C standard, as amended in 1994.
`iso9899:1999`
`c99`
`iso9899:199x`
`c9x`
The revised ISO C standard, published in December 1999. Before publication, this was known as C9X.
`gnu89`
The 1990 C standard plus GNU extensions. This is the default.
`gnu99`
`gnu9x`
The 1999 C standard plus GNU extensions.

`-ftabstop=NUMBER'

Set the distance between tab stops. This helps the preprocessor report correct column numbers in warnings or errors, even if tabs appear on the line. Values less than 1 or greater than 100 are ignored. The default is 8.

`-$'

Forbid the use of `$' in identifiers. The C standard allows implementations to define extra characters that can appear in identifiers. By default the GNU C preprocessor permits `$', a common extension.

Concept Index

#

a

c

d

e

f

g

h

i

l

m

n

o

p

r

s

t

u

Index of Directives, Macros and Options

#

-

- -C
- -D
- -dD
- -dI
- -dM
- -dN
- -ftabstop
- -gcc
- -H
- -I
- -idirafter
- -imacros
- -include
- -iprefix
- -isystem, -isystem
- -iwithprefix
- -M
- -MD
- -MM
- -MMD
- -nostdinc
- -nostdinc++
- -P
- -pedantic
- -pedantic-errors
- -remap
- -std
- -traditional
- -trigraphs
- -U
- -undef
- -Wall
- -Wcomment
- -Wtraditional
- -Wtrigraphs
- -Wundef
- -Wwhite-space
- -x assembler-with-cpp
- -x c
- -x objective-c

—

b

d

m

n

p

- pyr

s

- sequent
- sun

u

- unix

v

- vax
