THE UNIVERSITY OF SYDNEY

The C Preprocessor

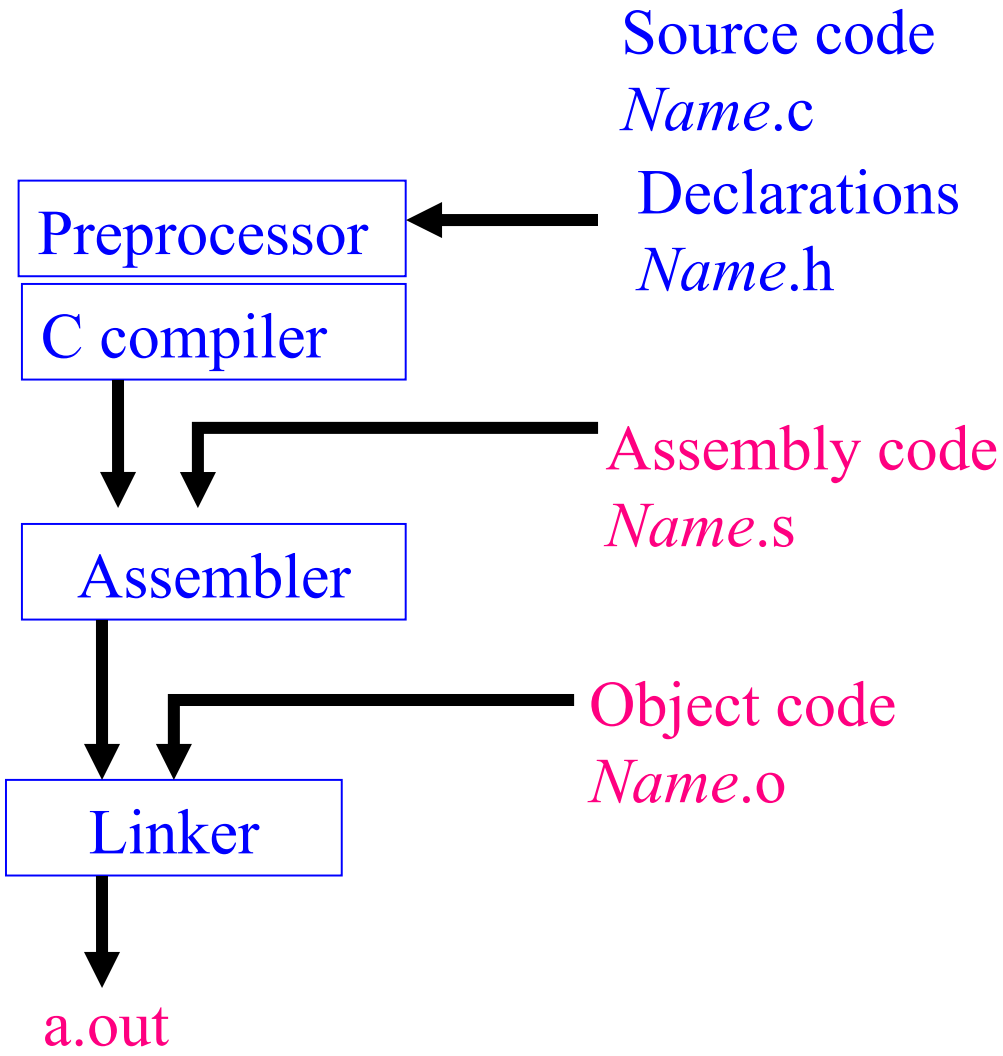
```
#define exprintf(expr,                \
        /* optional msg... */) \
    __exprintf(#expr,   \
        (int) (expr),   \
        "%s: %d\n" msg)

#define __exprintf(str_expr, \
        expr, fmt, args...)   \
    printf(fmt, \
        str_expr, expr, ##args)
```

THE UNIVERSITY OF SYDNEY

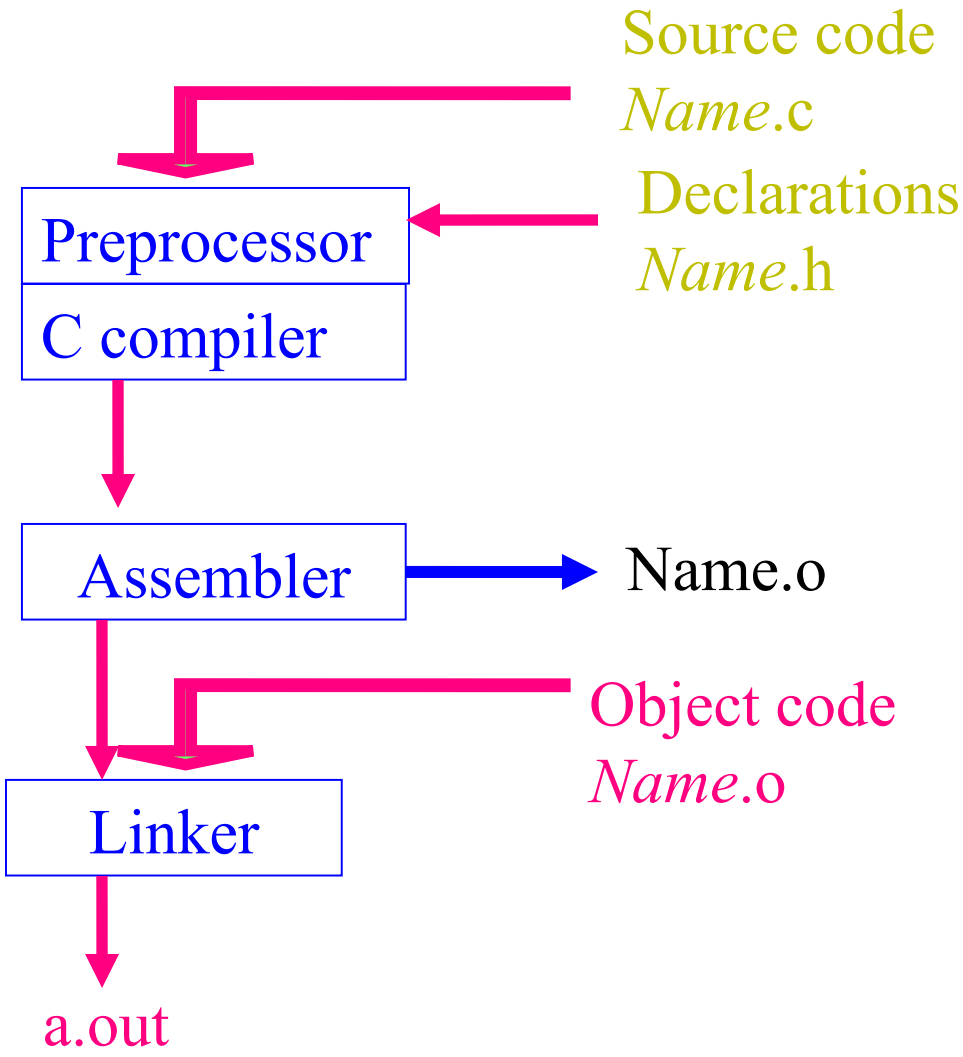
Object Code Files

- the C compiler can produce an *object code* version of a .c file that is machine language but not linked with other parts of your program
- these *object code* files end in .o

Compiling multi-file programs



Compiling multi-file programs



THE UNIVERSITY OF SYDNEY

Compiling and Linking

- the C compiler can be instructed to produce the .o file from the .c using the -c flag, eg:

```
gcc -c util.c
```

- several .c or .o files can be combined to produce an executable program:

```
gcc myprog.c util.o -o myprog
```

- the *object code* files are linked together to form the final executable program

- after changing a .c file we only need to recompile the affected file into its object code (.o) form and then relink all the .o files to produce the executable

THE UNIVERSITY OF SYDNEY

About dynamically shared libraries

- Loaded at runtime, not compile time
- Static linking is possible, but what are the advantages and disadvantages?
- Security issues? Yes
  - Easy to intercept and hijack to launching of that application, or even the file itself.
  - On trusting trust. August 1984 Volume 27 ACM Communications.

THE UNIVERSITY OF SYDNEY

Preprocessor commands

- preprocessor commands are lines starting with #

eg #include

- The C preprocessor interprets these lines

THE UNIVERSITY OF SYDNEY

Including Text from other files

- The #include statement is used to include text from another file into your program file at that point
- C programs typically consist of many source code files that each contain a small number of functions
- functions work on common data structures and so need declarations of the data structure to be included in each file

THE UNIVERSITY OF SYDNEY

Including Text from other files

- rather than copy the declarations into every file (error prone!) we can use *include files*
- Useful for:
  - externs
  - typedefs
  - struct definitions
- can even nest the included files

Two Files: before

myprog.c

```
/* My Program */
#include "decs.h"

int main(int argc,
         char *argv[])
{
…
}
```

decs.h

```
/* Declarations */

extern  int count;

struct employee
{
…
}
```

After preprocessor:

```
extern  int count;

struct employee
{
…
}

int main(int argc,
                char *argv[])
{
…
}
```

Included text

THE UNIVERSITY OF SYDNEY

- By convention, the names of included files end in ".h"
- So called “header” files because they tend to be included near the head of the program file

- Why shouldn’t you include actual code?

- Why should you include relevant header files rather than simply have them in the code?

THE UNIVERSITY OF SYDNEY

Defined Symbols

- An identifier symbol can be given a value by the preprocessor

```
#define LINES 100
```

- The preprocessor will replace the identifier LINES with the string 100 ***whenever*** it finds it in the program

Before preprocessor

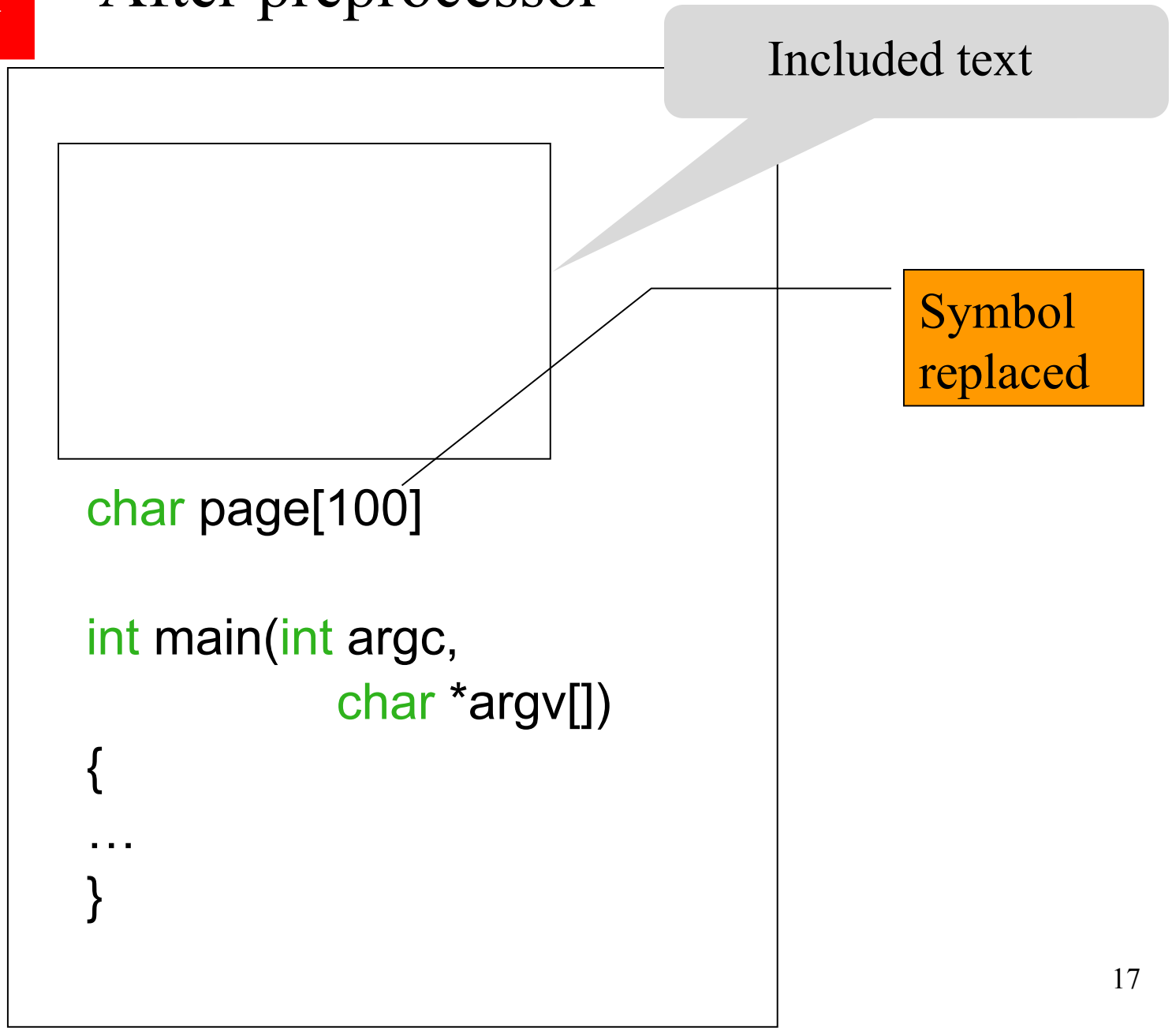
myprog.c

```
/* My Program */
#include "decs.h"
char page[LINES]
int main(int argc,
         char *argv[])
{
…
}
```

decs.h

```
/* Declarations */


#define LINES 100
```

After preprocessor



THE UNIVERSITY OF SYDNEY

Any replacement string

- The replacement string can be any string of characters:

```
#define LINES 5*10*20
```

Before preprocessor

myprog.c

```
/* My Program */
#include "decs.h"
char page[LINES]
int main(int argc,
         char *argv[])
{
…
}
```

decs.h

```
/* Declarations */


#define LINES 5*10*20
```

After preprocessor

Symbol replaced

```
char page[5*10*20]

int main(int argc,
                char *argv[])
{
…
}
```

THE UNIVERSITY OF SYDNEY

Warning!

- The replacement string can be any string of characters and replaces the symbol exactly

```
#define LINES 10+10
```

Before preprocessor

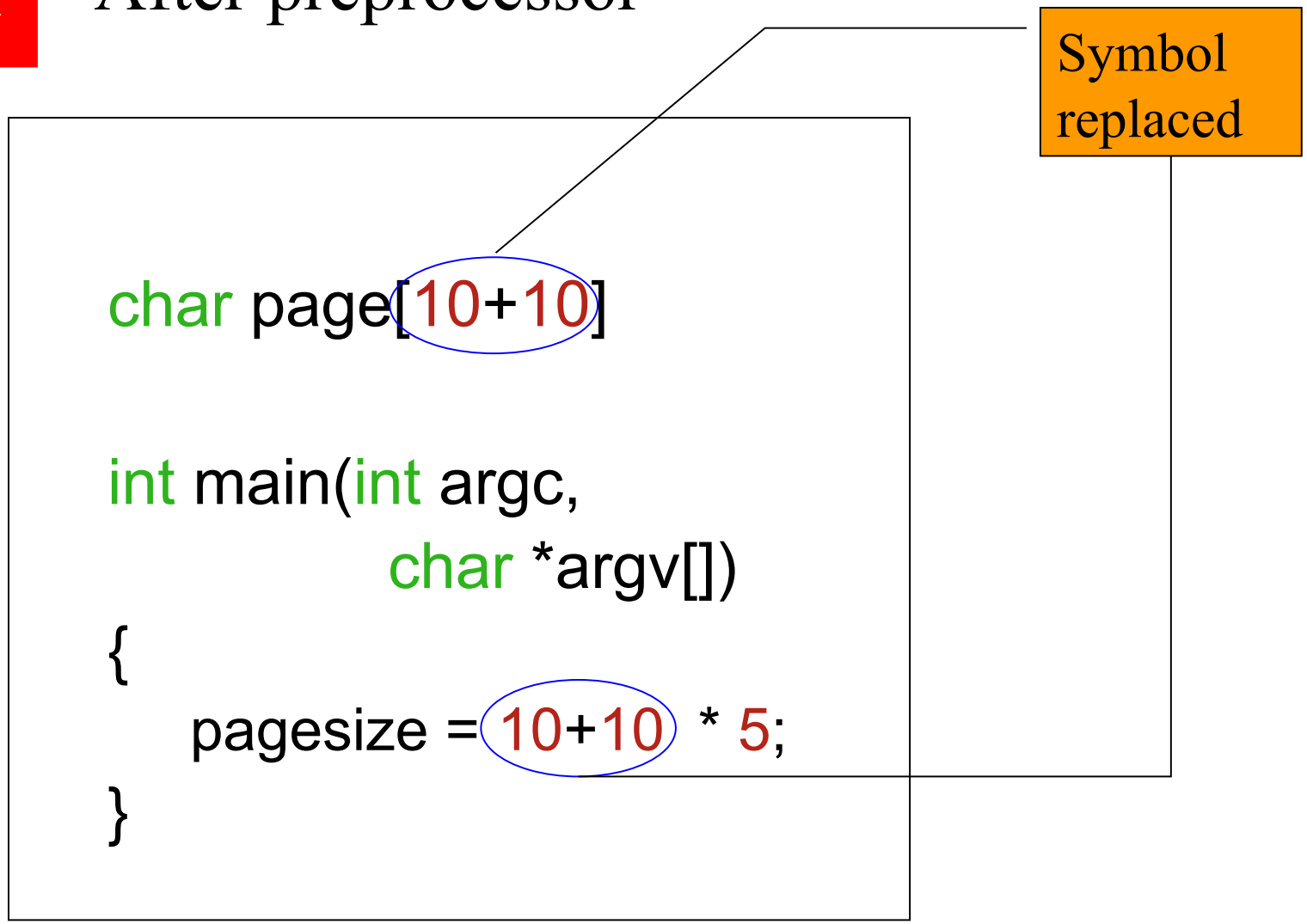
myprog.c

```
/* My Program */
#include "decs.h"
char page[LINES]
int main(int argc,
         char *argv[])
{
   pagesize = LINES * 5;
}
```

decs.h

```
/* Declarations */


#define LINES 10+10
```

After preprocessor



THE UNIVERSITY OF SYDNEY

Important Tip!

- Always bracket expressions in defined symbols:

```
#define LINES (10+10)
```

Defined symbols: macros with parameters

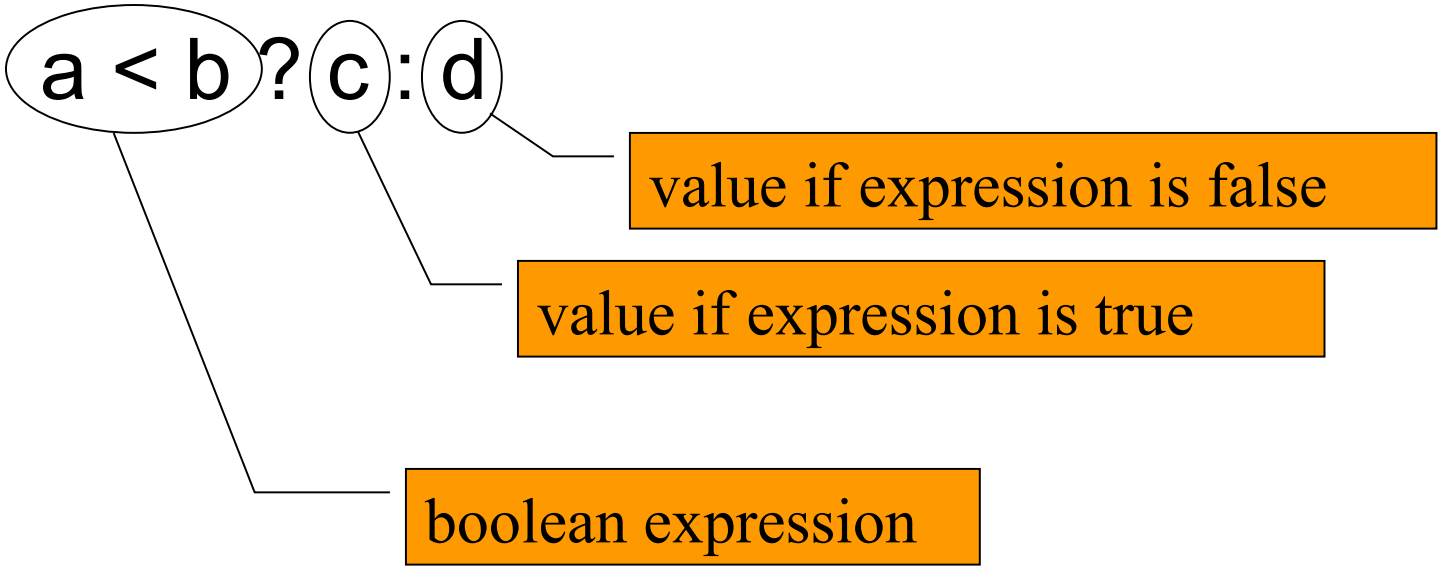
- A macro looks like a function with parameters
- A macro is processed by the preprocessor: replacing symbols in the body by parameters

```
#define min(a,b) ((a) < (b) ? (a):(b))
```

THE UNIVERSITY OF SYDNEY

Note!

Ternary operator ?:



Before preprocessor

myprog.c

```
/* My Program */
#include "decs.h"
int main(int argc,
                char *argv[])
{
    y = min(size,100)
    return 0;
}
```

decs.h

```
/* Declarations */

#define min(a,b) ((a)<(b)?(a):(b))
```

After preprocessor

```
/* My Program */
#include "decs.h"
int main(int argc,
             char *argv[])
{
    y = ((size)<(100)?(size):(100))
}
```

macro call replaced

Beware of side-effects

```
y = min(a++,b)  /* before */

y = ((a++)<(b) ? (a++):(b)  /* after */
```

Beware of side-effects

```
y = min(a++,b)  /* before */
```

```
y = ((a++)<(b) ? (a++):(b)  /* after */
```

a is incremented twice

THE UNIVERSITY OF SYDNEY

General form of macro

#define *identifier(identifier,…..) token-string*

THE UNIVERSITY OF SYDNEY

How can you tell a function from a macro?

examples:

```
if (isupper(ch)) …

if (ch = getchar())…
if (ch = getc(stdin)) …
if (ch = fgetc(stdin)) …
```

How can you tell a function from a macro?

stdio.h

```
/* The C standard explicitly says this is a
macro, so we always do the optimization for it.
*/
#define getc(_fp) _IO_getc (_fp)
```

libio.h

```
extern int _IO_getc (_IO_FILE *__fp);
```

THE UNIVERSITY OF SYDNEY

End of Segment

The C Preprocessor:

conditional inclusion

THE UNIVERSITY OF SYDNEY

Conditional inclusion

- the preprocessor allows you to select text to be included or not
- very useful for debugging: include debug printouts or not - controlled by preprocessor command

THE UNIVERSITY OF SYDNEY

Conditional inclusion

#ifdef

#if

#ifndef

#else

#elif

#undef

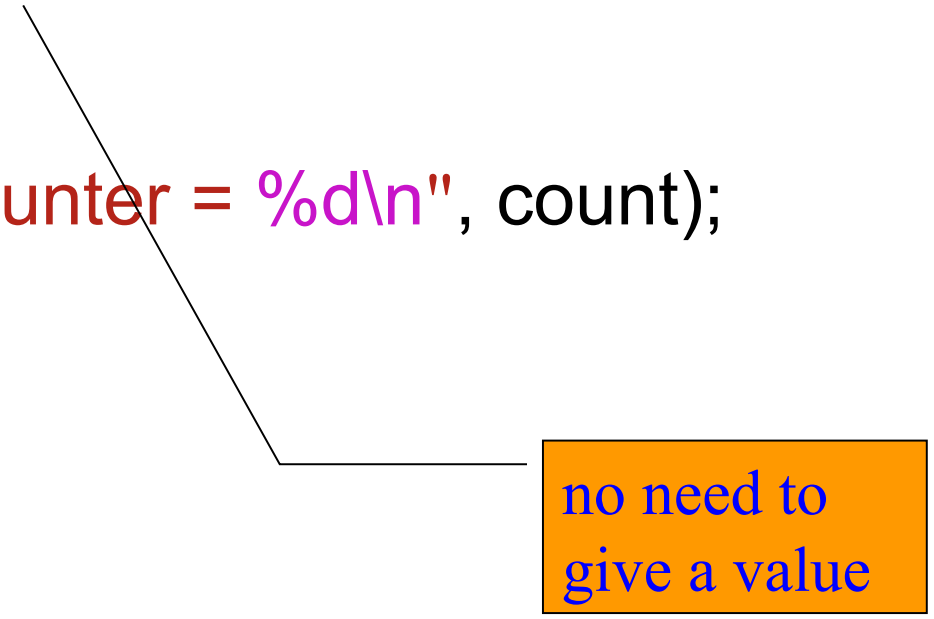
#endif

Conditional inclusion

#ifdef tests if a preprocessor symbol is defined

eg

```
#define DEBUG
#ifdef DEBUG
    printf("loop counter = %d\n", count);
#endif
```



THE UNIVERSITY OF SYDNEY

Conditional inclusion

#ifndef tests if a preprocessor symbol is *NOT* defined

eg

```
#ifndef FASTLINK
    …. /* code for slow links */
#endif
```

Conditional inclusion

#if allows more complex expressions to be used

eg

```
#if WINDOWWIDTH > 600
    …. /* code for wide windows */
#endif
```

Conditional inclusion

Both #ifdef and #if can have an #else

eg

```
#ifdef DEBUG
    …. /* Debugging version */
#else
    … /* production version */
#end
```

Conditional inclusion

#if can have an #elif

eg

```
#if WIDTH > 600
    …. /* wide version */
#elif WIDTH > 400
    … /* medium version */
#else
        … /* narrow version */
#endif
```

Before preprocessor

```
#include "declarations.h"
int main(int argc, char *argv[])
{
#ifdef DEBUG
    printf("MyProg (debug version)\n");
#else
    printf("MyProg (production version)\n");
#endif
    return 0;
}
```

THE UNIVERSITY OF SYDNEY

declarations.h

```
#define DEBUG
```

THE UNIVERSITY OF SYDNEY

After preprocessor

```
int main(int argc, char *argv[])
{

    printf("MyProg (debug version)\n");

}
```

Controlling the preprocessor from the gcc command

gcc -DWIDTH=600 prog.c

has the same effect as

#define WIDTH 600

at the beginning of the program

#define or #undef **within** the program overrides the command line setting

gcc -Didentifier

is equivalent to

#define identifier

multiple -D arguments can be used:

gcc -DWIDTH=600 -DTEST prog.c

Useful for debugging

gcc -DEBUG prog.c

prog.c:

```
#ifdef EBUG
#define DEBUG(m) printf("debug: %s\n", (m))
#else
#define DEBUG(m)  /* null statement */
#endif

…
   DEBUG("called proc fn");
…
```

Alternatives

```
#ifdef DEBUG
     printf(…)
#endif
```

compared with

```
enum {DEBUG = 0}
…
if (DEBUG)
    printf(…)
```

Pre-defined Symbols

- the preprocessor defines several symbols automatically
- the most useful of these are:

__LINE__ contains the current line number at any point

__FILE__ contains the name of the current program file

THE UNIVERSITY OF SYDNEY

When would you need them?

- __LINE__?
- __FILE__?

Debug example revisited

gcc -DEBUG prog.c

continuation indicator

prog.c:

```
#ifdef EBUG
#define DEBUG(m) \
  printf("debug: %s at line %d in file %s\n", \
                              (m), __LINE__, __FILE__)
#else
#define DEBUG(m) /* null statement */
#endif
…
    DEBUG("called doit function");
…
```

THE UNIVERSITY OF SYDNEY

#include revisited

- the normal form of #include has a file name in double quotes - this specifies a relative or absolute path for the file
- if the file name is enclosed in <> the file is searched for in /usr/include
- the preprocessor can be instructed to look in other directories using the -I directory flag
- this allows you to have your own directory for include files

Example

#include <defs.h>

- the preprocessor will look in
  /usr/include for the defs.h file

- if we use the command
  gcc -I/home/john/include myprog.c

the preprocessor will look in
/home/john/include for the file

Caution!

```
#define IF          if(
#define THEN        )
#define BEGIN       {
#define END         }
```

```
IF a == 1 THEN
BEGIN
        dothis()
        dothat();
END
```

A new language!
Unreadable for the next programmer.

THE UNIVERSITY OF SYDNEY

Preprocessor as a tool

- gcc -E ….

- runs just the preprocessor
- can be used as a tool exploiting
  - #define call by name
  - #ifdef for conditional generation

THE UNIVERSITY OF SYDNEY

Example: hack templates

- Generate text in different forms as required:
  - #include parts
  - #define to replace parts

THE UNIVERSITY OF SYDNEY

Role of preprocessor

- lots of it around, especially for
  - machine-dependencies
  - OS versions
- pretty hard to read code with lots of conditional compilation through it
- how to debug the conditional compilations proprocessor “code”?

THE UNIVERSITY OF SYDNEY

- the preprocessor is very useful for configuring programs
  - different versions
  - debugging
- not found in Java

- image: https://packagecontrol.io/packages/C%20Improved

End of Segment

THE UNIVERSITY OF SYDNEY

Security matters for data types

What if we started with zero (in floating point representation) and then added 0.1 to it 1000 times?

What would the result be?

What does this print?

```
float   x = 0.1;

        printf("%10.1f\n", x);
        printf("%10.5f\n", x);
        printf("%20.15f\n", x);
        printf("%50.45f\n", x);
```

Catastrophic cancellation

- Two operands are close to each other
- They cancel out because of order of the operation, eg
  - bignum - bignum + smallnum
  - Quadratic equation discriminant

Catastrophic cancellation

b = 3.34
a = 1.22
c = 2.28

exact (b * b) - 4 * a * c
.0292

(b * b) rounds to 11.2
4ac to 11,1
answer 0.1

Benign cancellation

- Avoid the problems by reorganising the expression
- Example: catastrophic cancellation
  - (x * x) - (y * y) has catastrophic
  - Becomes benign (more accurate) as

    (x - y)(x + y)

About errors

- Sources
  - The raw data eg 7.3
  - Propagation error eg 7.3 + 8.15
  - Representational errors (finiteness)
- Awareness
  - Significant figures
  - Printing
  - calculations
- Measures
  - Absolute error
  - Relative error

- In 1979, NORAD defense radar misinterpreted the moon as an incoming missile…
- In 1983, a software "bug" resulted in an F-14 flying off the edge of an aircraft carrier and into the North Sea.
- In 1984, a 180-degree error caused a Soviet test missile to head towards Hamburg instead of the Arctic.
- The splashdown point of Gemini V was off by over 100 miles because the guidance system neglected movement of the Earth around the Sun.

June 4, 1996
Ariane 5 is launched


ARIANESPACE

Ariane 5 exploded in its 37th second of flight including its payload of 4 satellites.

It was later found to be due to a small data-representation problem...

- Computed horizontal velocity as floating point number
  - Converted to 16-bit integer
    - Worked OK for Ariane 4
    - Overflowed for Ariane 5
      - Used same software

February 25, 1991
The (first) Gulf War

An Iraqi Scud missile penetrates US Patriot defenses

"On February 25th, 1991, an Iraqi Scud hit the barracks in Dahran, Saudi Arabia, killing 28 soldiers from the US Army's 14th Quartermaster detatchment.

28 soldiers died and over 100 were injured as a result.

It was later found to be due to a small data-representation problem...

A government investigation revealed that the failed intercept at Dhahran had been caused by a software error in the system's clock. The Patriot missile battery at Dhahran had been in operation for 100 hours, by which time the system's internal clock had drifted by one third of a second. Due to the closure speed of the interceptor and the target, this resulted in a miss distance of 600 meters.

Security of type limits

- Integer overflow and underflow
  - Undefined Behaviour (UB)
- SSH1 exploit for CRC check: https://www.kb.cert.org/vuls/id/945216

```
uint8_t val = k + n; // overflow if k+n > 255
uint8_t val = k - n; // underflow if k-n < 0

// code that expects monotonically
increasing/decreasing sequences
   Loop counters
   Simulation time (t += tstep)
   index access for memory
   index access for files (fseek)

Memory allocation issues
… = malloc(val * sizeof(int));
```

Security of type limits

- Memory allocation issues

```
int *resize_for_next_minute(
      uint16_t time, uint16_t *buffer_size)
{
    *buffer_size += 60 * sizeof(int);
    int *buffer = malloc(*buffer_size);
    ...
}

int *resize_for_next_minute(uint16_t time) {
    buffer = malloc( (time + 60) * sizeof(int) );
    ...
}
```

THE UNIVERSITY OF SYDNEY

Security of type limits

- Memory allocation issues

```
char *ptr;
char *max;
size_t len;
if (ptr + len > max)
     return EINVAL;

// revision 1
if (ptr + len < ptr || ptr + len > max) {
   return EINVAL;
}
```

Works on out debugging version...but not the production version!

THE UNIVERSITY OF SYDNEY

Security of type limits

Production version: Compiles with optimisations `-O2`

Mathematically, for integers `Ptr, A, B`:

```
Ptr + A < Ptr + B
```

Is the same as: `A < B`

So we can optimise this expression

```
Ptr + len < ptr  → len < 0
```

Since len is a positive integer, `len < 0` is always false
Hence, the boolean expression `ptr + len < ptr` is always false!
Remove this redundant computation!

THE UNIVERSITY OF SYDNEY

Security of type limits

There is a need to to rewrite expressions to account for **unbounded** len values:

```
if (len > max - ptr)
        return EINVAL;
```

Summary

- Understand that there exists undefined behaviour for types
  - big disasters may result from small errors
  - small errors are hard to find
- Common pitfalls to be aware of
  - Expect that floats may be inaccurate: the 0.1 problem due to 0.1 being between two representable floats
  - the out of range problem - number too big or too small
- Never, ever test floats for equality

Effective C. Seacord, Robert C. Computer (Long Beach, Calif.), 2020-11, Vol.53 (11), p.79-82. Peer reviewed

End of Segment