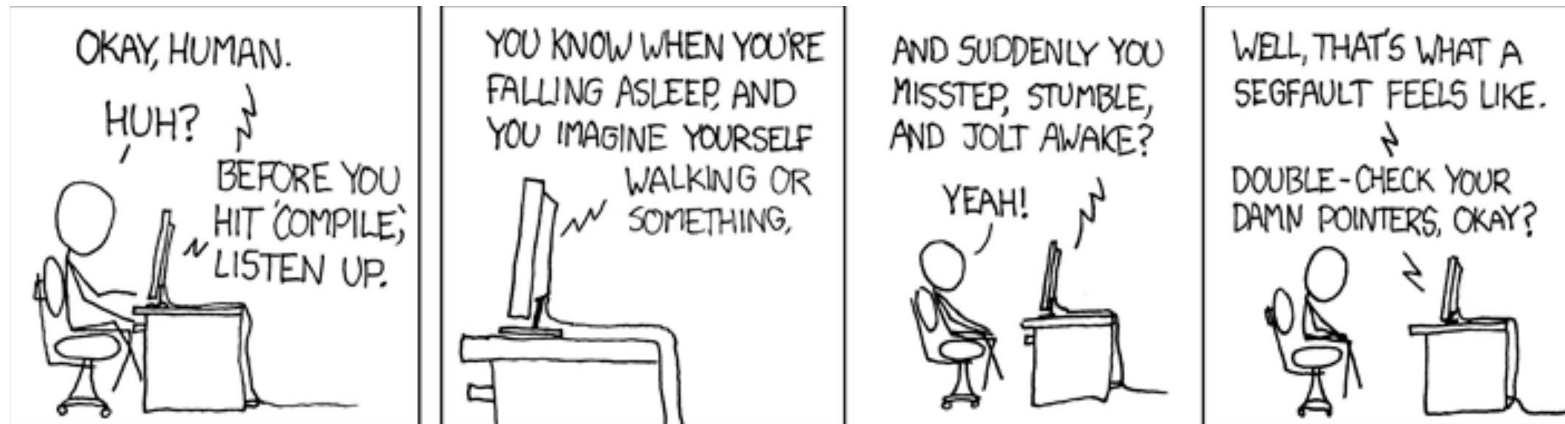


CS107, Lecture 8

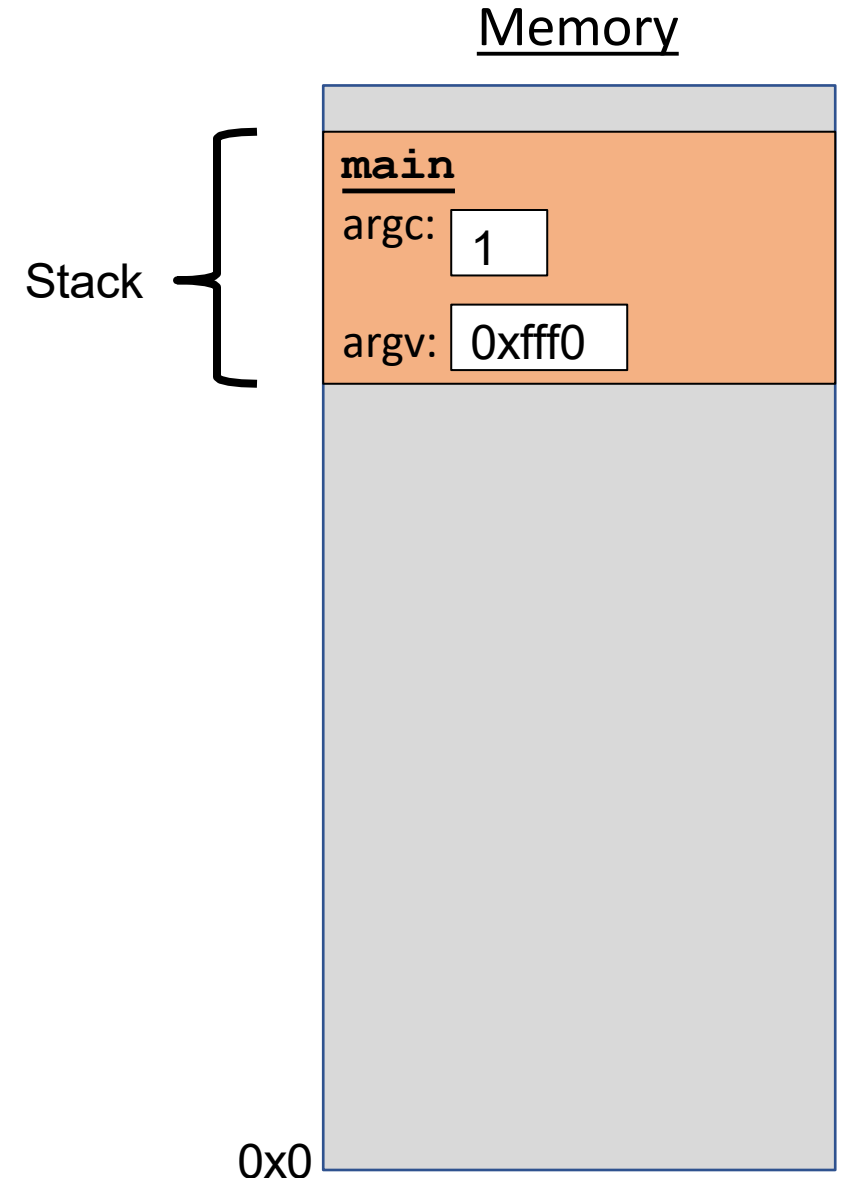
C Generics – Void *



The Stack

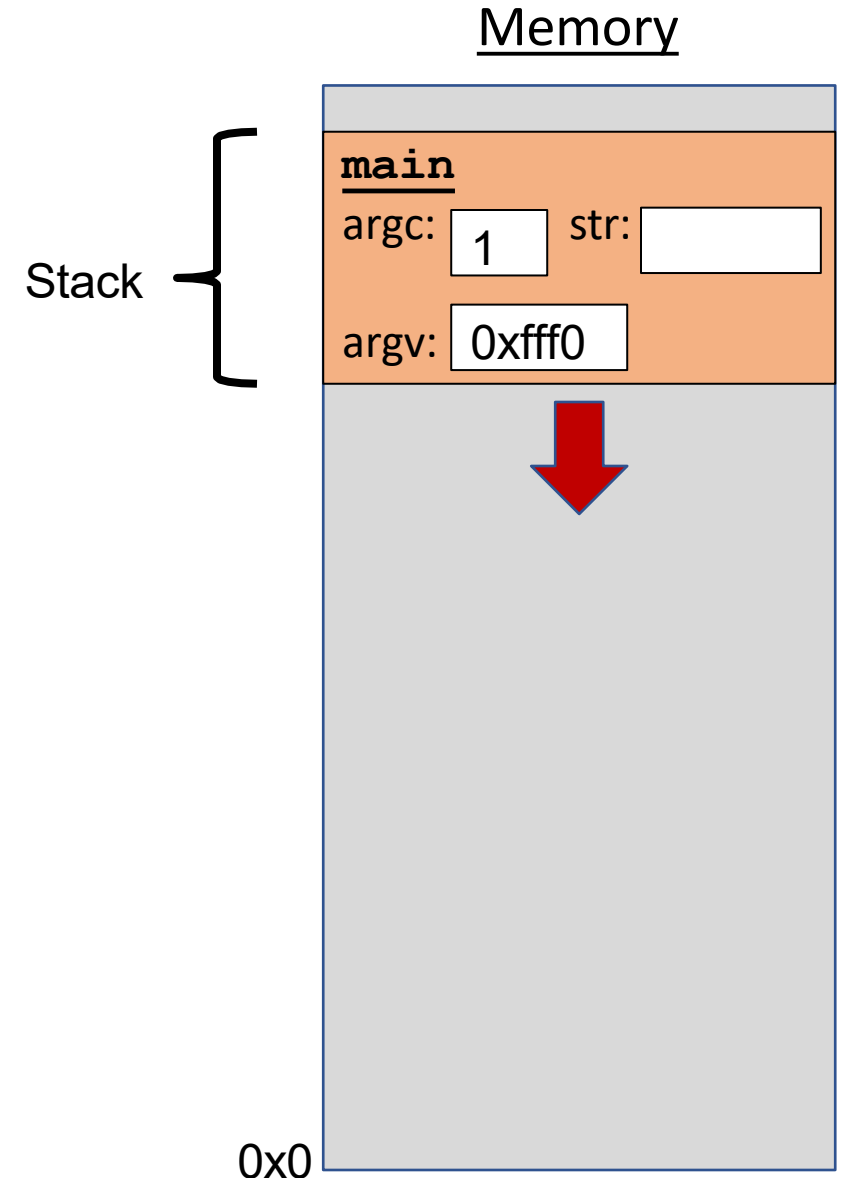
```
char *create_string(char ch, int num) {  
    char new_str[num + 1];  
    for (int i = 0; i < num; i++) {  
        new_str[i] = ch;  
    }  
    new_str[num] = '\0';  
    return new_str;  
}
```

```
int main(int argc, char *argv[]) {  
    char *str = create_string('a', 4);  
    printf("%s", str); // want "aaaa"  
    return 0;  
}
```



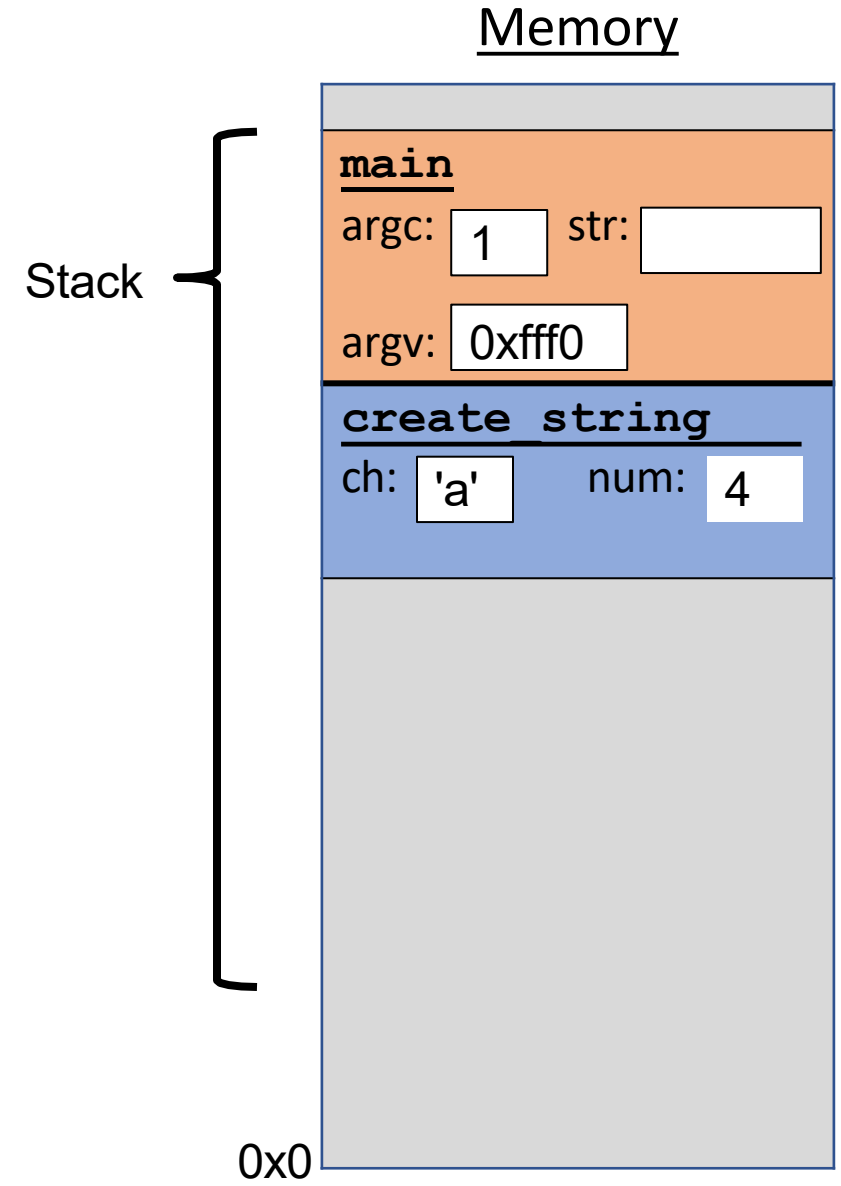
The Stack

```
char *create_string(char ch, int num) {  
    char new_str[num + 1];  
    for (int i = 0; i < num; i++) {  
        new_str[i] = ch;  
    }  
    new_str[num] = '\0';  
    return new_str;  
}  
  
int main(int argc, char *argv[]) {  
    char *str = create_string('a', 4);  
    printf("%s", str); // want "aaaa"  
    return 0;  
}
```



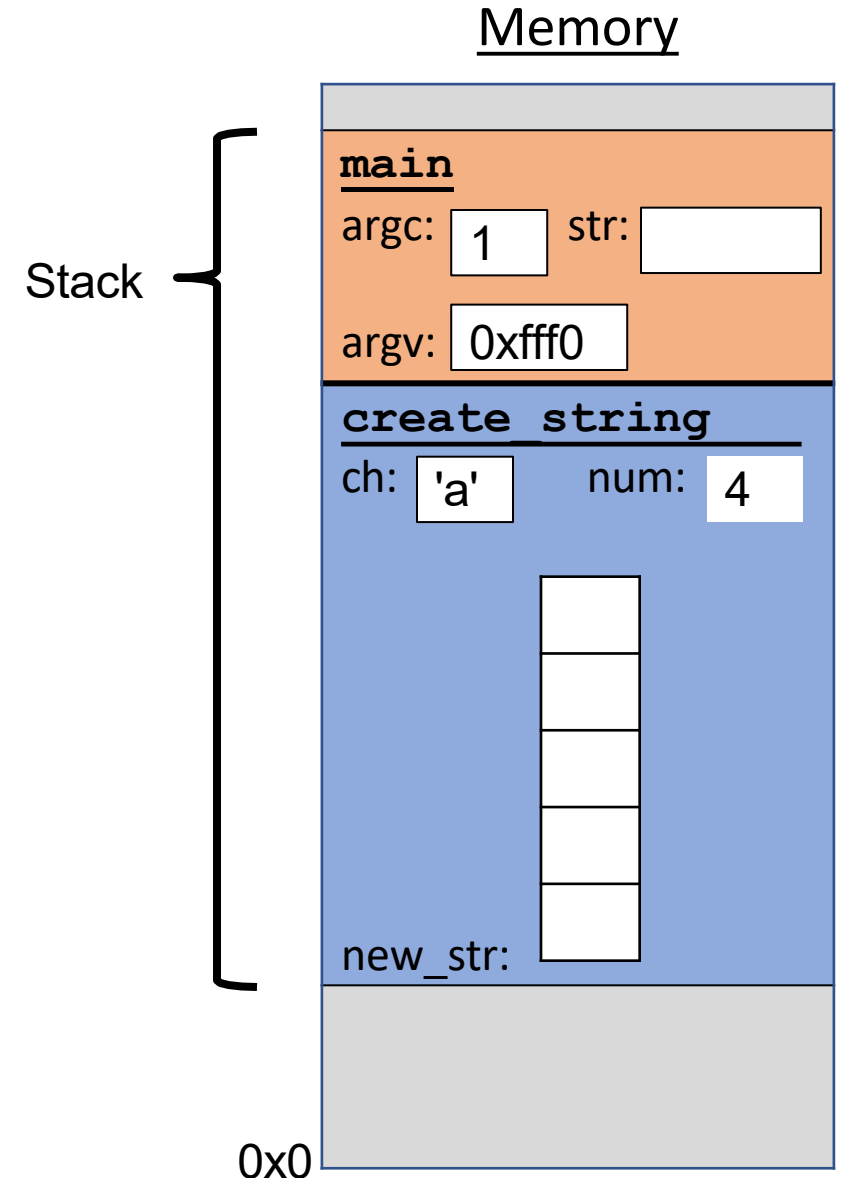
The Stack

```
char *create_string(char ch, int num) {  
    char new_str[num + 1];  
    for (int i = 0; i < num; i++) {  
        new_str[i] = ch;  
    }  
    new_str[num] = '\0';  
    return new_str;  
}  
  
int main(int argc, char *argv[]) {  
    char *str = create_string('a', 4);  
    printf("%s", str); // want "aaaa"  
    return 0;  
}
```



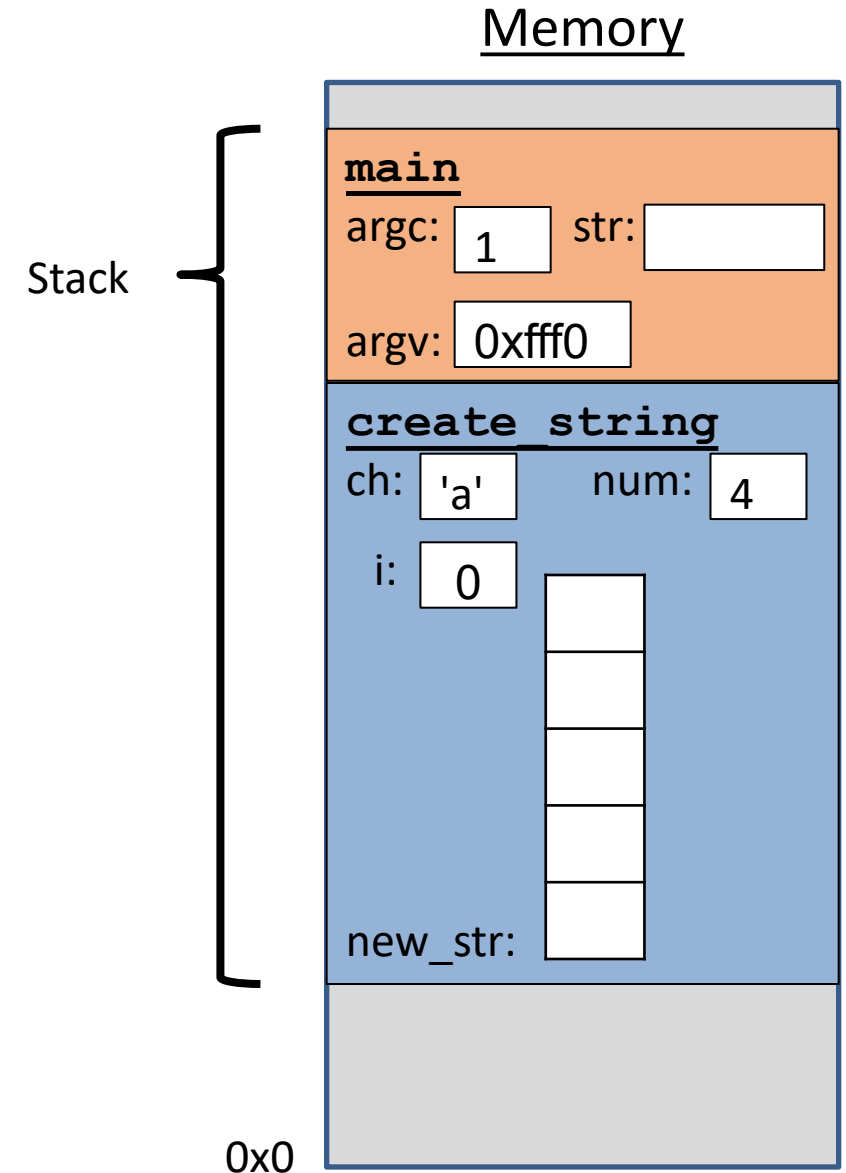
The Stack

```
char *create_string(char ch, int num) {  
    char new_str[num + 1];  
    for (int i = 0; i < num; i++) {  
        new_str[i] = ch;  
    }  
    new_str[num] = '\0';  
    return new_str;  
}  
  
int main(int argc, char *argv[]) {  
    char *str = create_string('a', 4);  
    printf("%s", str); // want "aaaa"  
    return 0;  
}
```



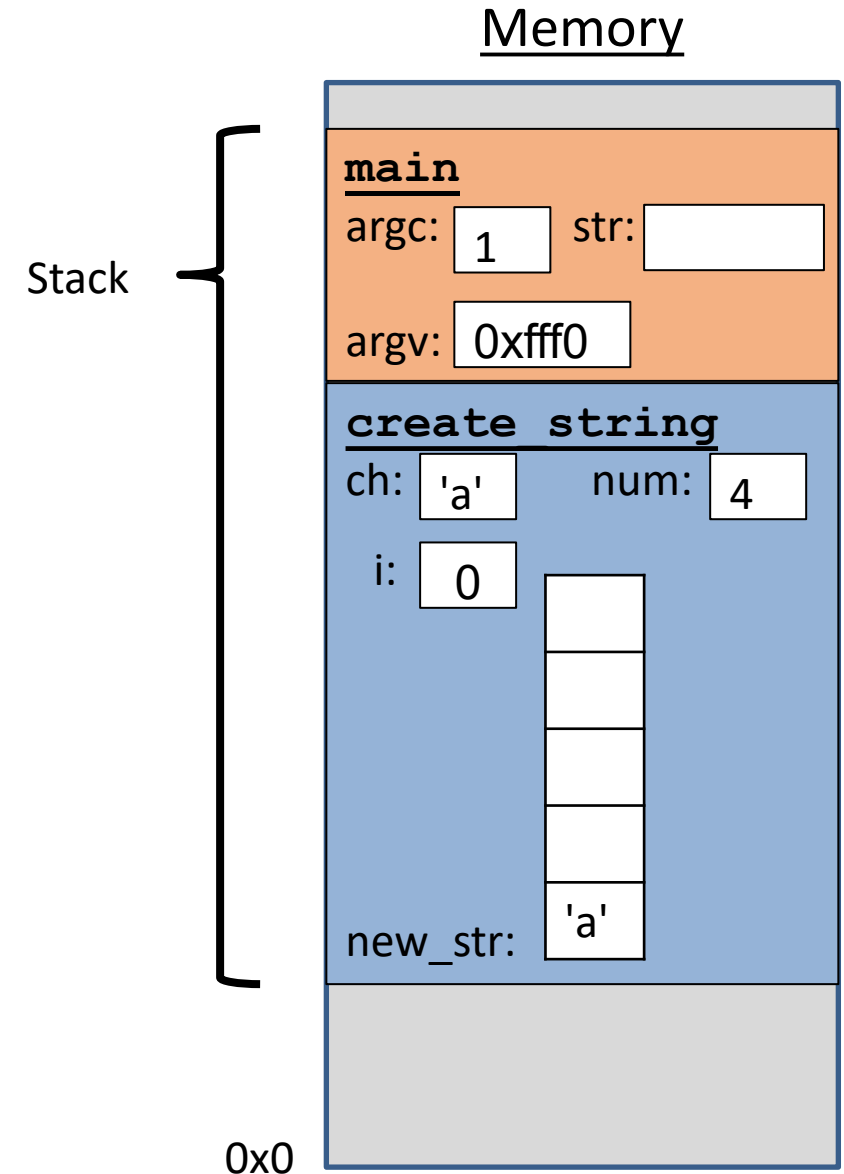
The Stack

```
char *create_string(char ch, int num) {  
    char new_str[num + 1];  
    for (int i = 0; i < num; i++) {  
        new_str[i] = ch;  
    }  
    new_str[num] = '\0';  
    return new_str;  
}  
  
int main(int argc, char *argv[]) {  
    char *str = create_string('a', 4);  
    printf("%s", str); // want "aaaa"  
    return 0;  
}
```



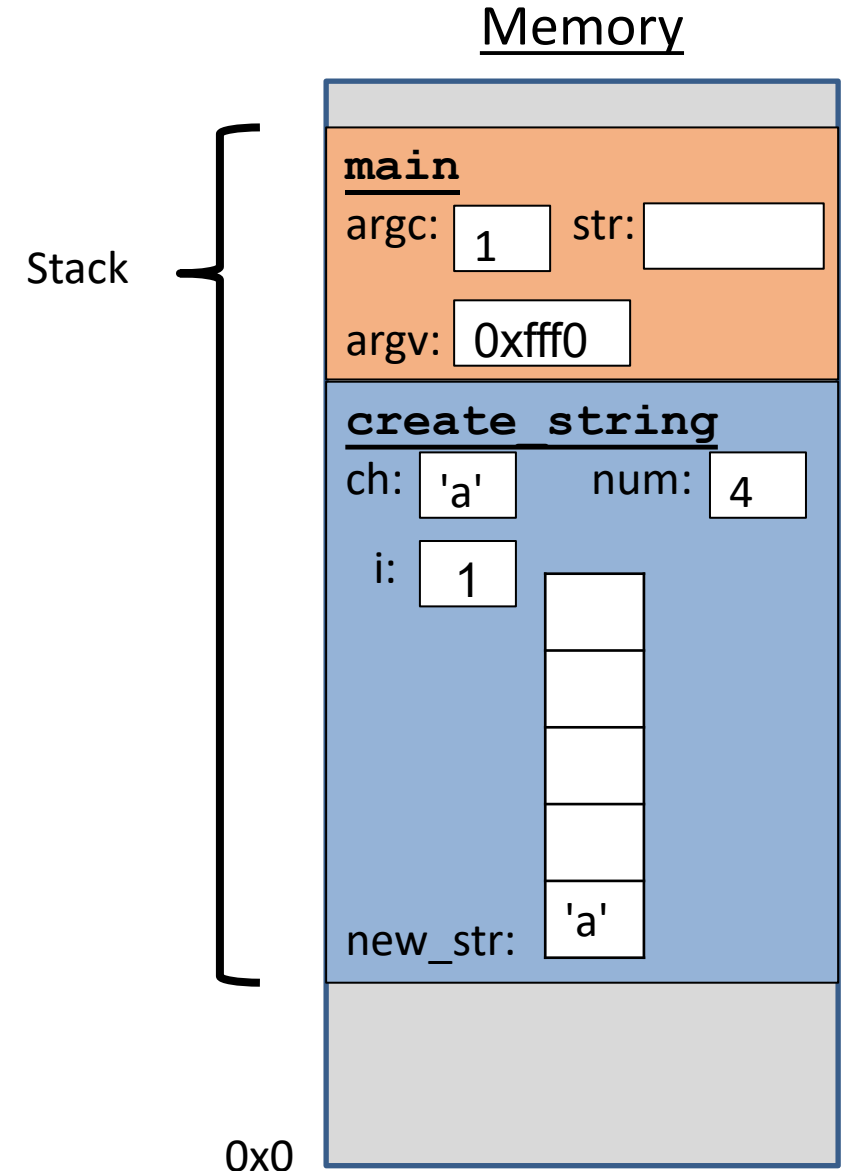
The Stack

```
char *create_string(char ch, int num) {  
    char new_str[num + 1];  
    for (int i = 0; i < num; i++) {  
        new_str[i] = ch;  
    }  
    new_str[num] = '\0';  
    return new_str;  
}  
  
int main(int argc, char *argv[]) {  
    char *str = create_string('a', 4);  
    printf("%s", str); // want "aaaa"  
    return 0;  
}
```



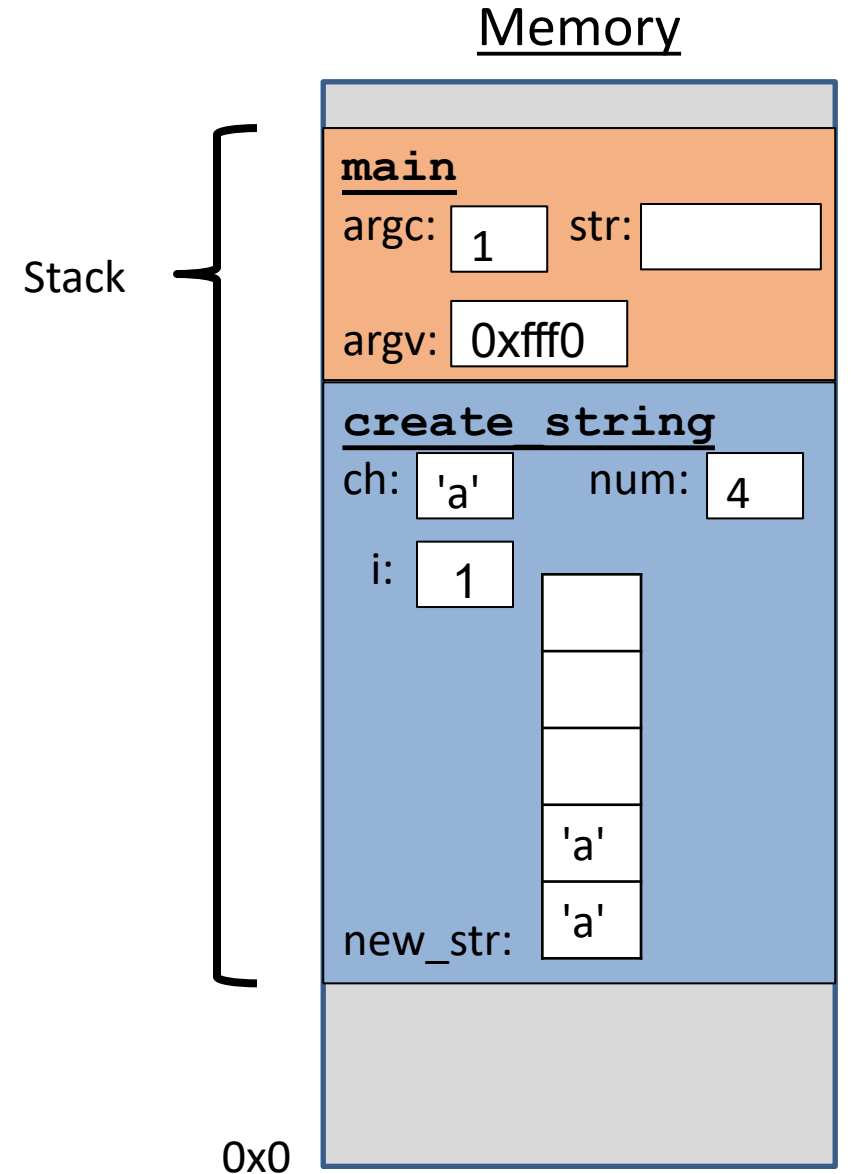
The Stack

```
char *create_string(char ch, int num) {  
    char new_str[num + 1];  
    for (int i = 0; i < num; i++) {  
        new_str[i] = ch;  
    }  
    new_str[num] = '\0';  
    return new_str;  
}  
  
int main(int argc, char *argv[]) {  
    char *str = create_string('a', 4);  
    printf("%s", str); // want "aaaa"  
    return 0;  
}
```



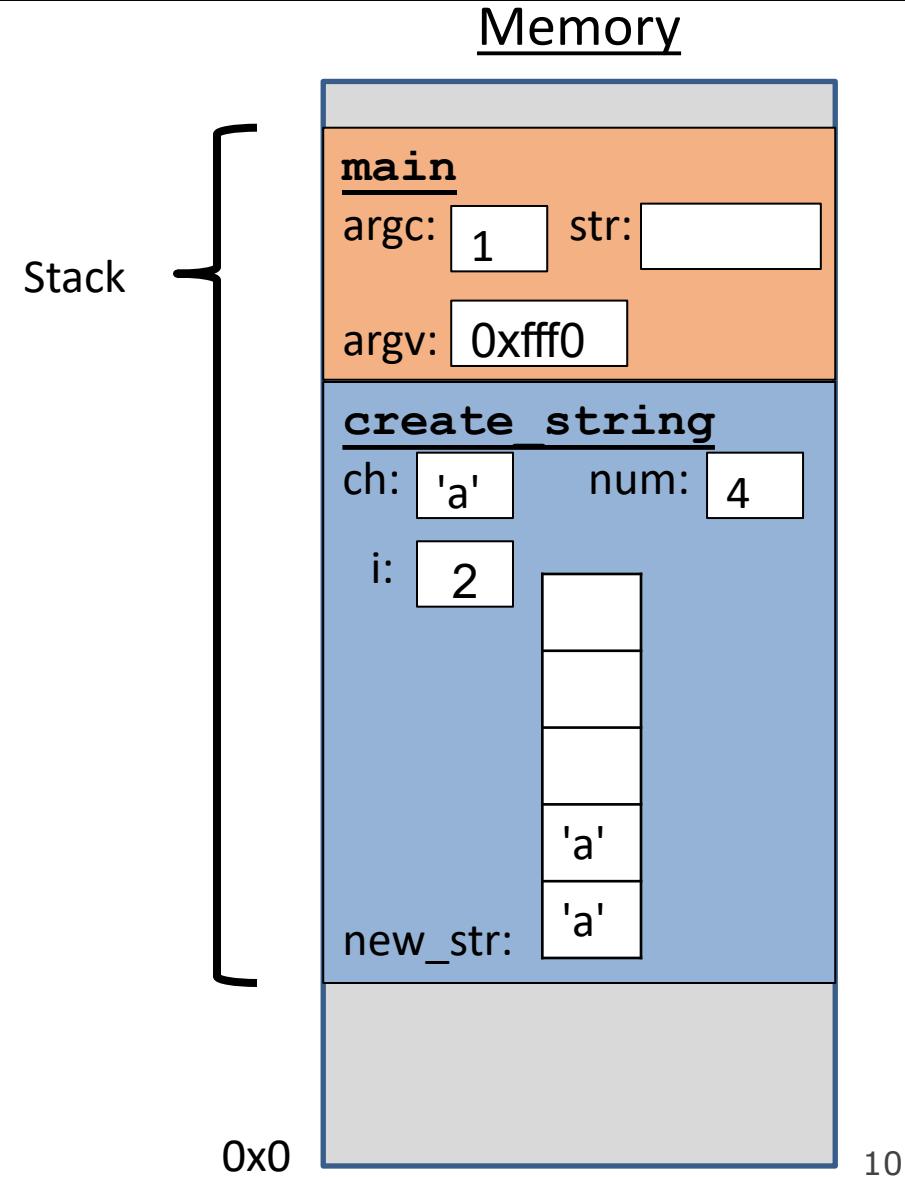
The Stack

```
char *create_string(char ch, int num) {  
    char new_str[num + 1];  
    for (int i = 0; i < num; i++) {  
        new_str[i] = ch;  
    }  
    new_str[num] = '\0';  
    return new_str;  
}  
  
int main(int argc, char *argv[]) {  
    char *str = create_string('a', 4);  
    printf("%s", str); // want "aaaa"  
    return 0;  
}
```



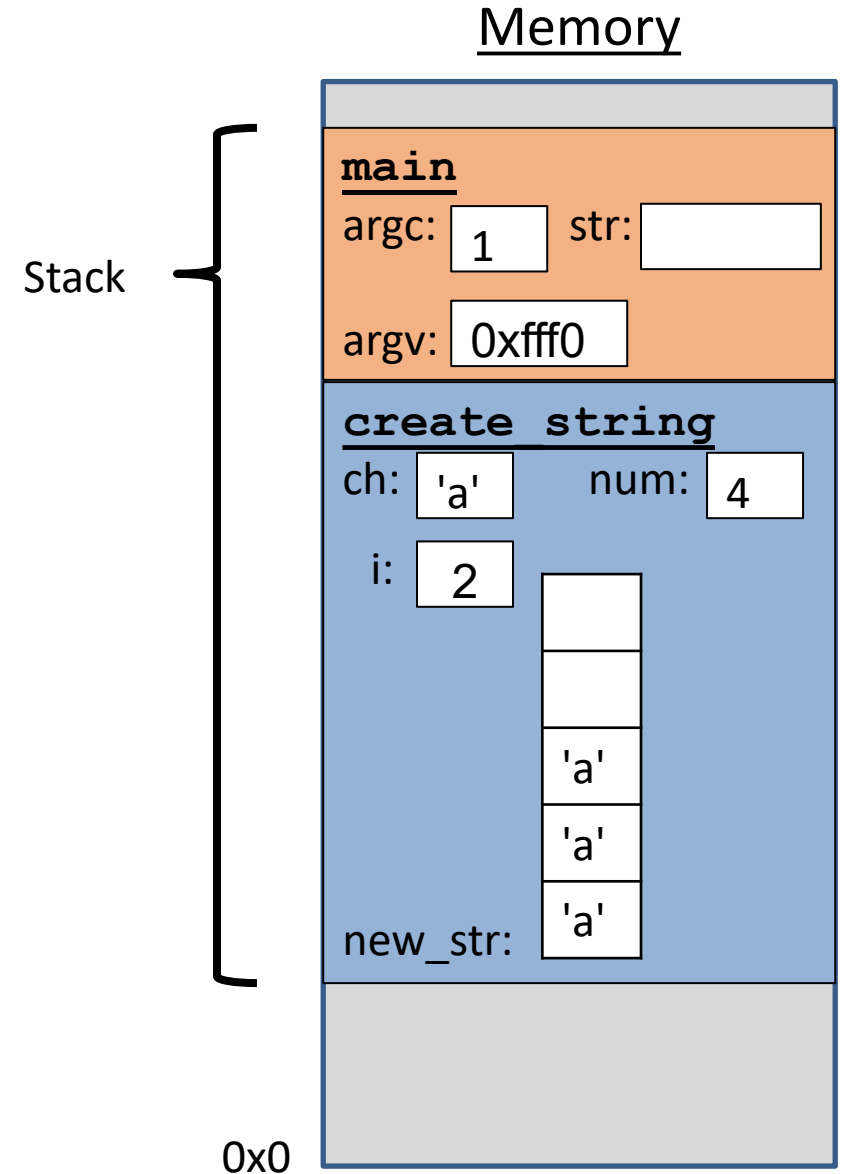
The Stack

```
char *create_string(char ch, int num) {  
    char new_str[num + 1];  
    for (int i = 0; i < num; i++) {  
        new_str[i] = ch;  
    }  
    new_str[num] = '\\0';  
    return new_str;  
}  
  
int main(int argc, char *argv[]) {  
    char *str = create_string('a', 4);  
    printf("%s", str); // want "aaaa"  
    return 0;  
}
```



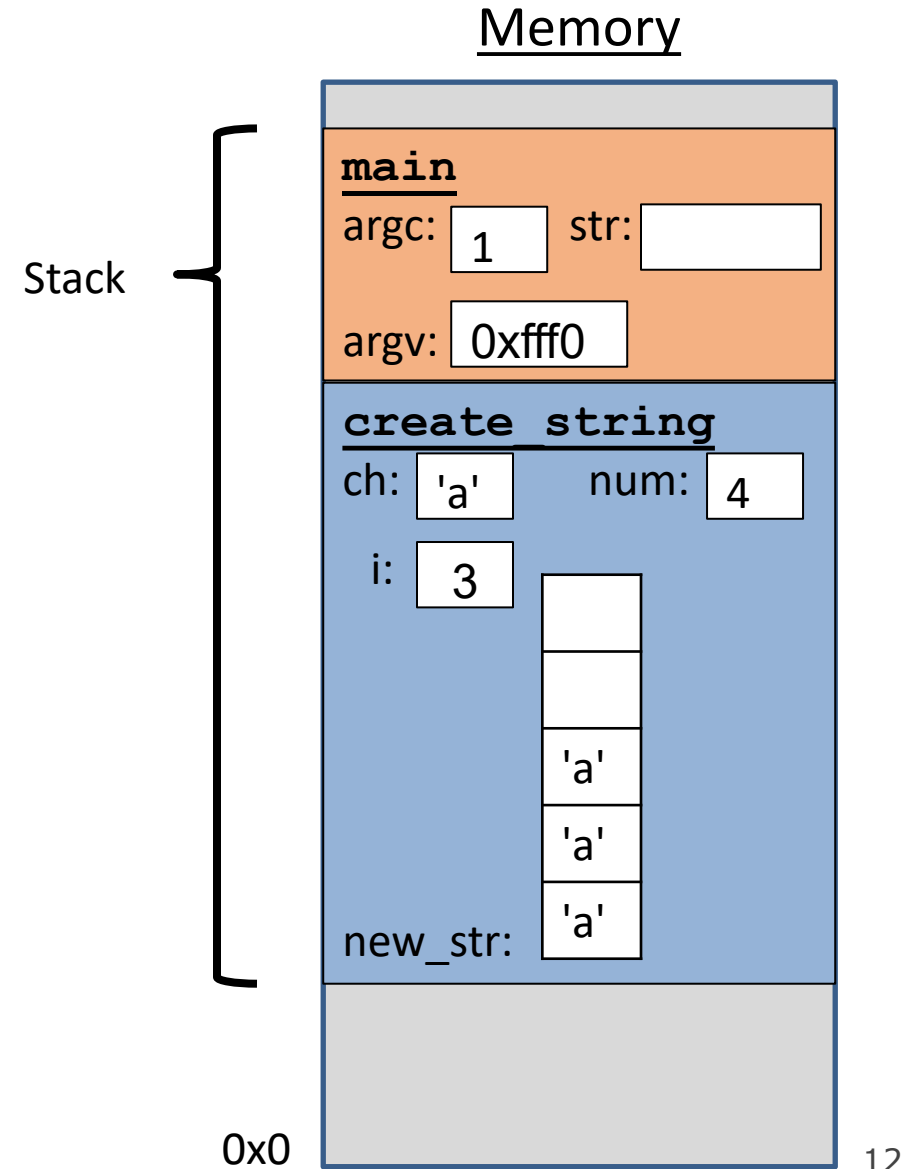
The Stack

```
char *create_string(char ch, int num) {  
    char new_str[num + 1];  
    for (int i = 0; i < num; i++) {  
        new_str[i] = ch;  
    }  
    new_str[num] = '\0';  
    return new_str;  
}  
  
int main(int argc, char *argv[]) {  
    char *str = create_string('a', 4);  
    printf("%s", str); // want "aaaa"  
    return 0;  
}
```



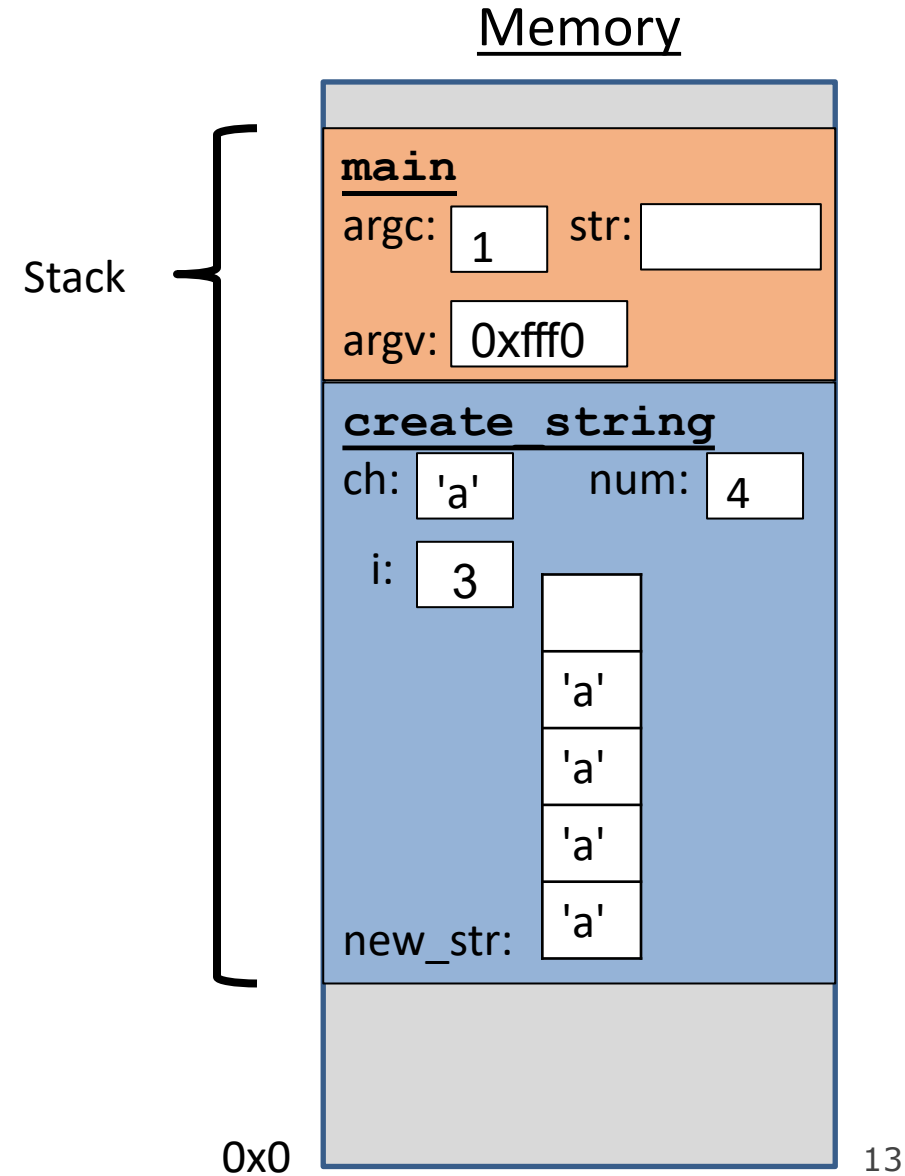
The Stack

```
char *create_string(char ch, int num) {  
    char new_str[num + 1];  
    for (int i = 0; i < num; i++) {  
        new_str[i] = ch;  
    }  
    new_str[num] = '\0';  
    return new_str;  
}  
  
int main(int argc, char *argv[]) {  
    char *str = create_string('a', 4);  
    printf("%s", str); // want "aaaa"  
    return 0;  
}
```



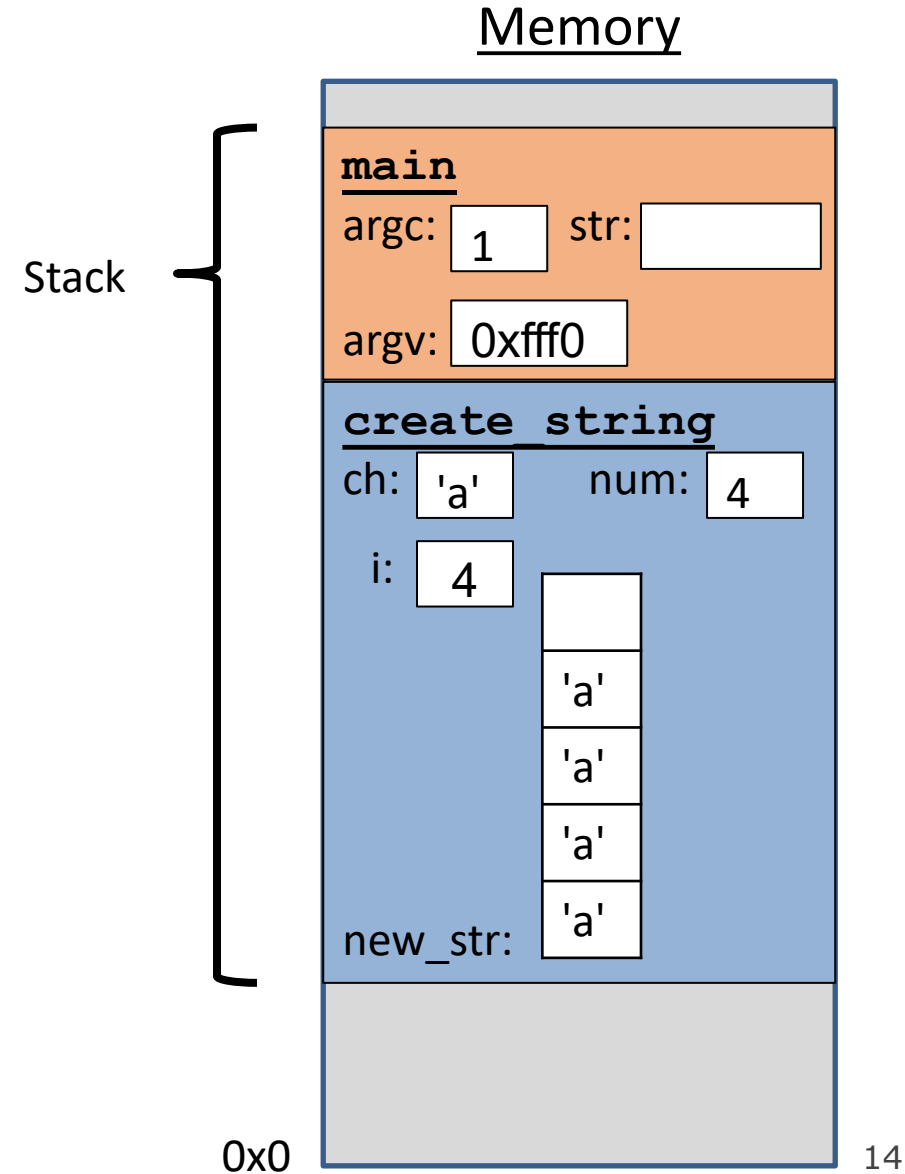
The Stack

```
char *create_string(char ch, int num) {  
    char new_str[num + 1];  
    for (int i = 0; i < num; i++) {  
        new_str[i] = ch;  
    }  
    new_str[num] = '\0';  
    return new_str;  
}  
  
int main(int argc, char *argv[]) {  
    char *str = create_string('a', 4);  
    printf("%s", str); // want "aaaa"  
    return 0;  
}
```



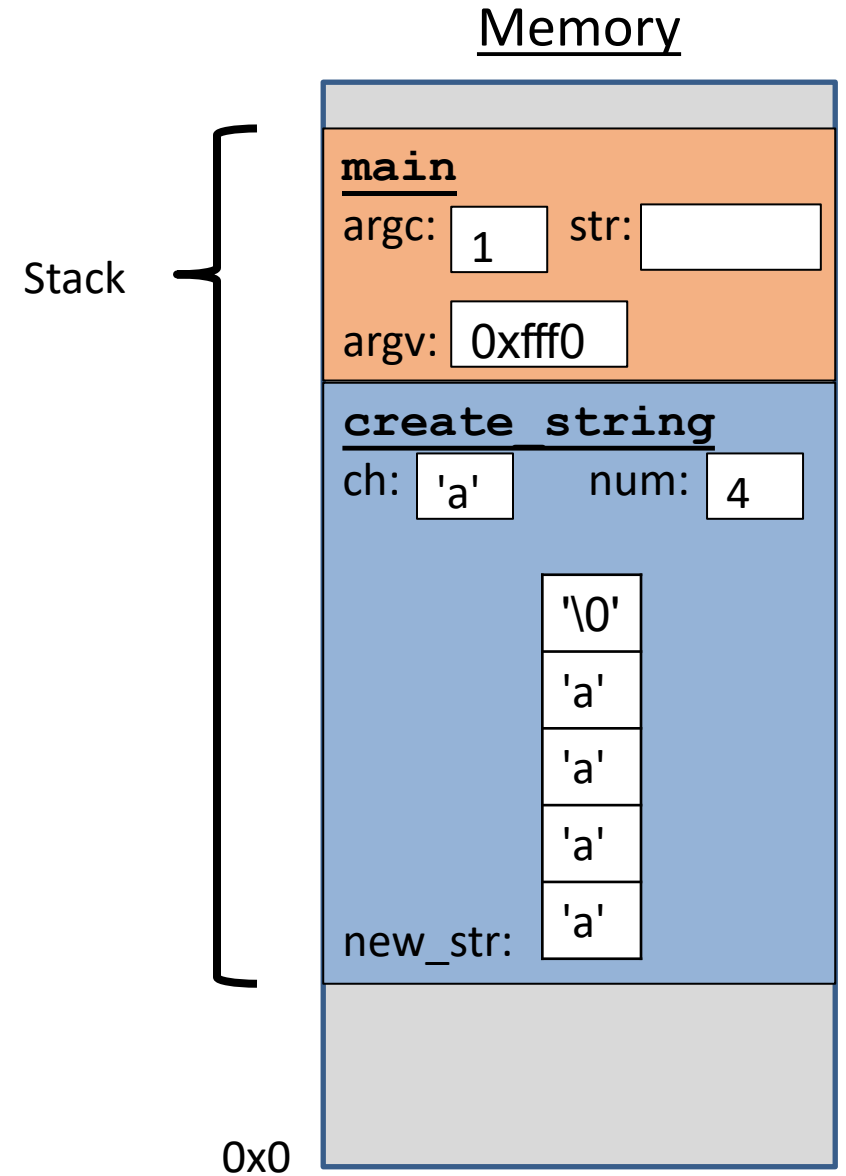
The Stack

```
char *create_string(char ch, int num) {  
    char new_str[num + 1];  
    for (int i = 0; i < num; i++) {  
        new_str[i] = ch;  
    }  
    new_str[num] = '\\0';  
    return new_str;  
}  
  
int main(int argc, char *argv[]) {  
    char *str = create_string('a', 4);  
    printf("%s", str); // want "aaaa"  
    return 0;  
}
```



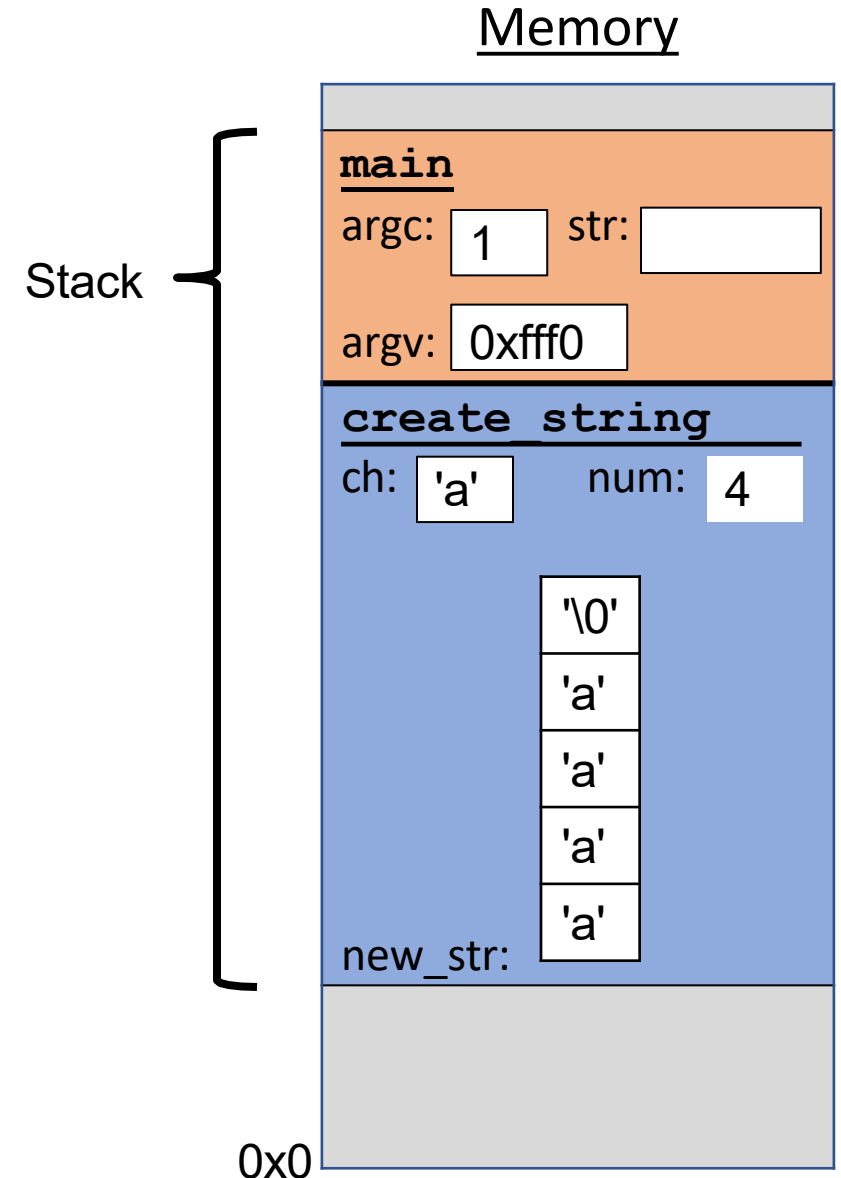
The Stack

```
char *create_string(char ch, int num) {  
    char new_str[num + 1];  
    for (int i = 0; i < num; i++) {  
        new_str[i] = ch;  
    }  
    new_str[num] = '\0';  
    return new_str;  
}  
  
int main(int argc, char *argv[]) {  
    char *str = create_string('a', 4);  
    printf("%s", str); // want "aaaa"  
    return 0;  
}
```



The Stack

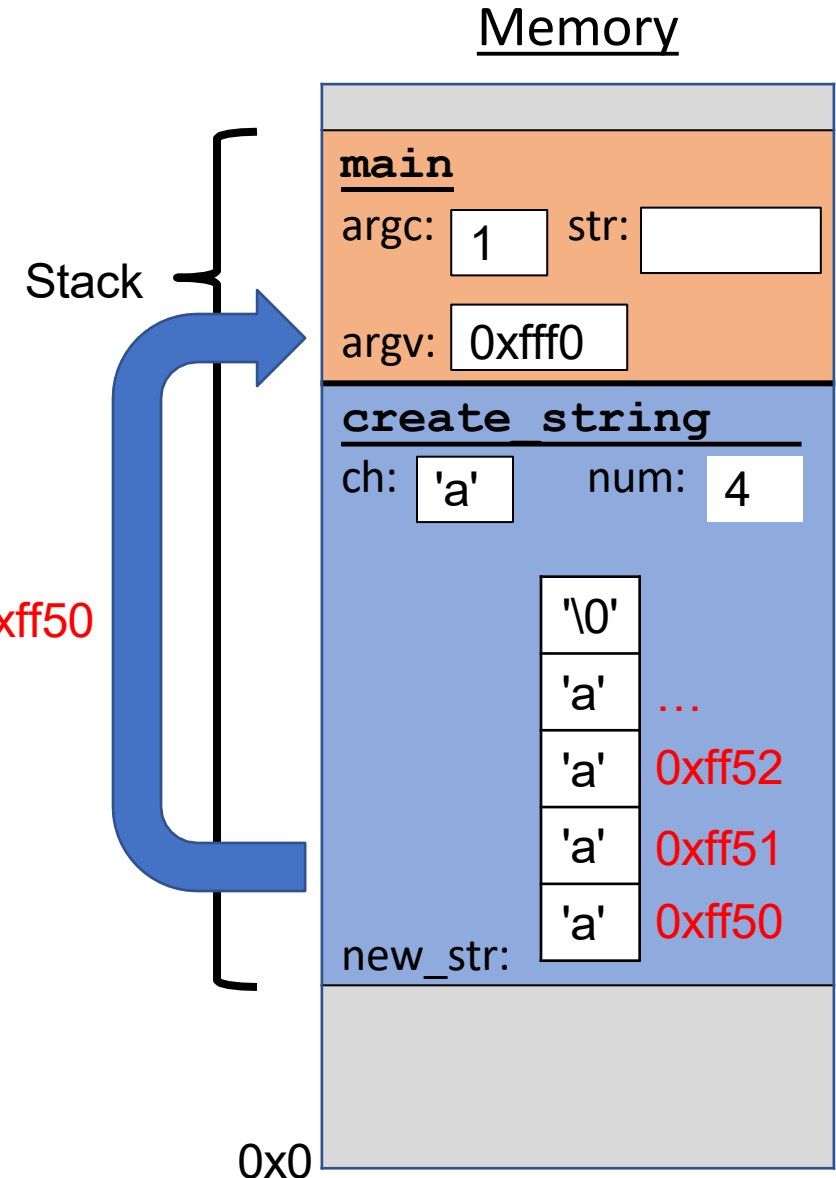
```
char *create_string(char ch, int num) {  
    char new_str[num + 1];  
    for (int i = 0; i < num; i++) {  
        new_str[i] = ch;  
    }  
    new_str[num] = '\0';  
    return new_str;  
}  
  
int main(int argc, char *argv[]) {  
    char *str = create_string('a', 4);  
    printf("%s", str); // want "aaaa"  
    return 0;  
}
```



The Stack

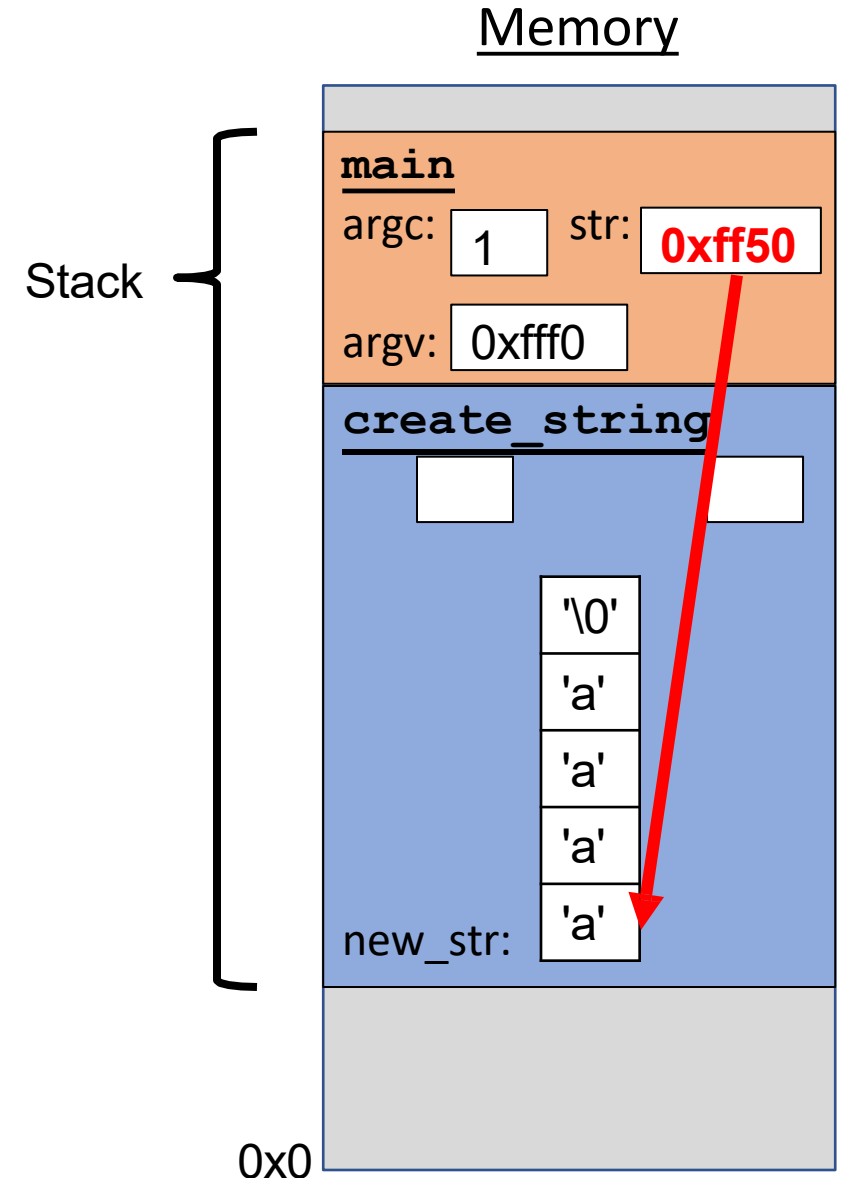
```
char *create_string(char ch, int num) {  
    char new_str[num + 1];  
    for (int i = 0; i < num; i++) {  
        new_str[i] = ch;  
    }  
    new_str[num] = '\0';  
    return new_str;  
}  
  
int main(int argc, char *argv[]) {  
    char *str = create_string('a', 4);  
    printf("%s", str); // want "aaaa"  
    return 0;  
}
```

Returns e.g. 0xff50



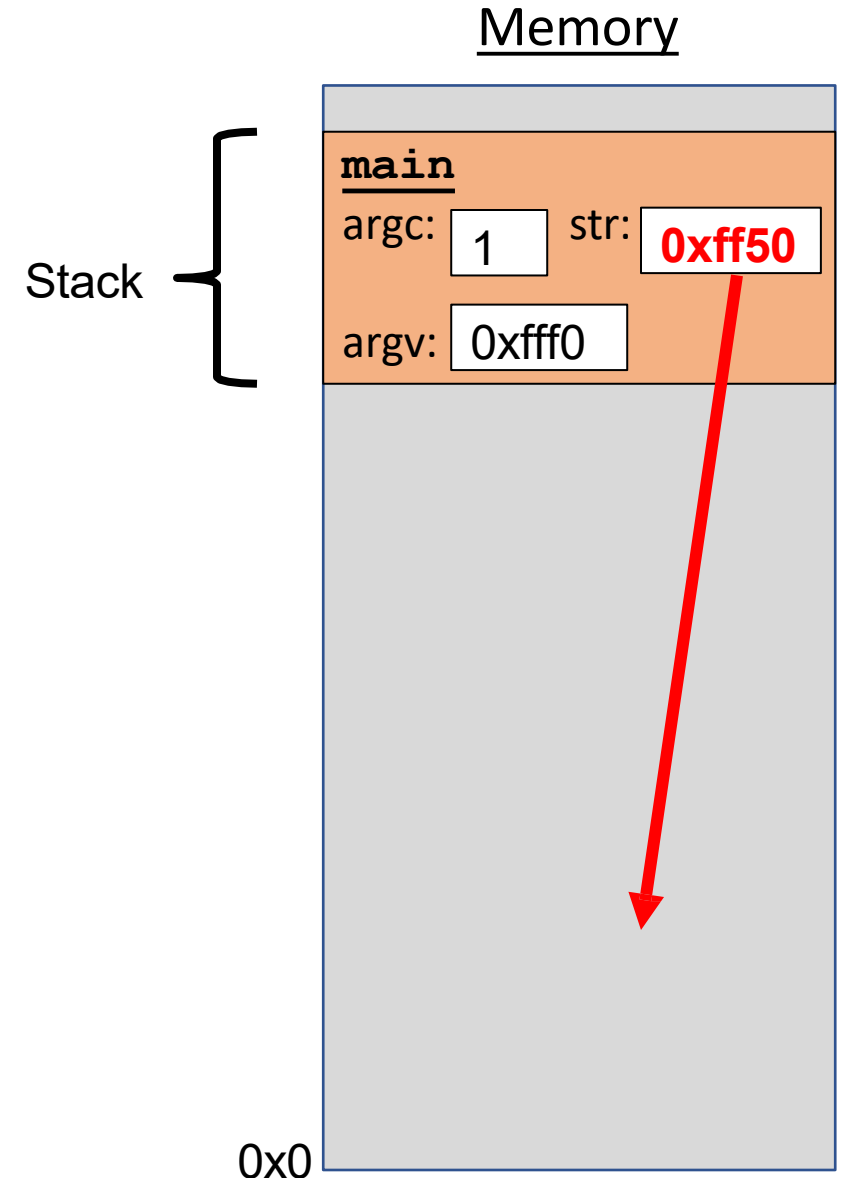
The Stack

```
char *create_string(char ch, int num) {  
    char new_str[num + 1];  
    for (int i = 0; i < num; i++) {  
        new_str[i] = ch;  
    }  
    new_str[num] = '\0';  
    return new_str;  
}  
  
int main(int argc, char *argv[]) {  
    char *str = create_string('a', 4);  
    printf("%s", str); // want "aaaa"  
    return 0;  
}
```



The Stack

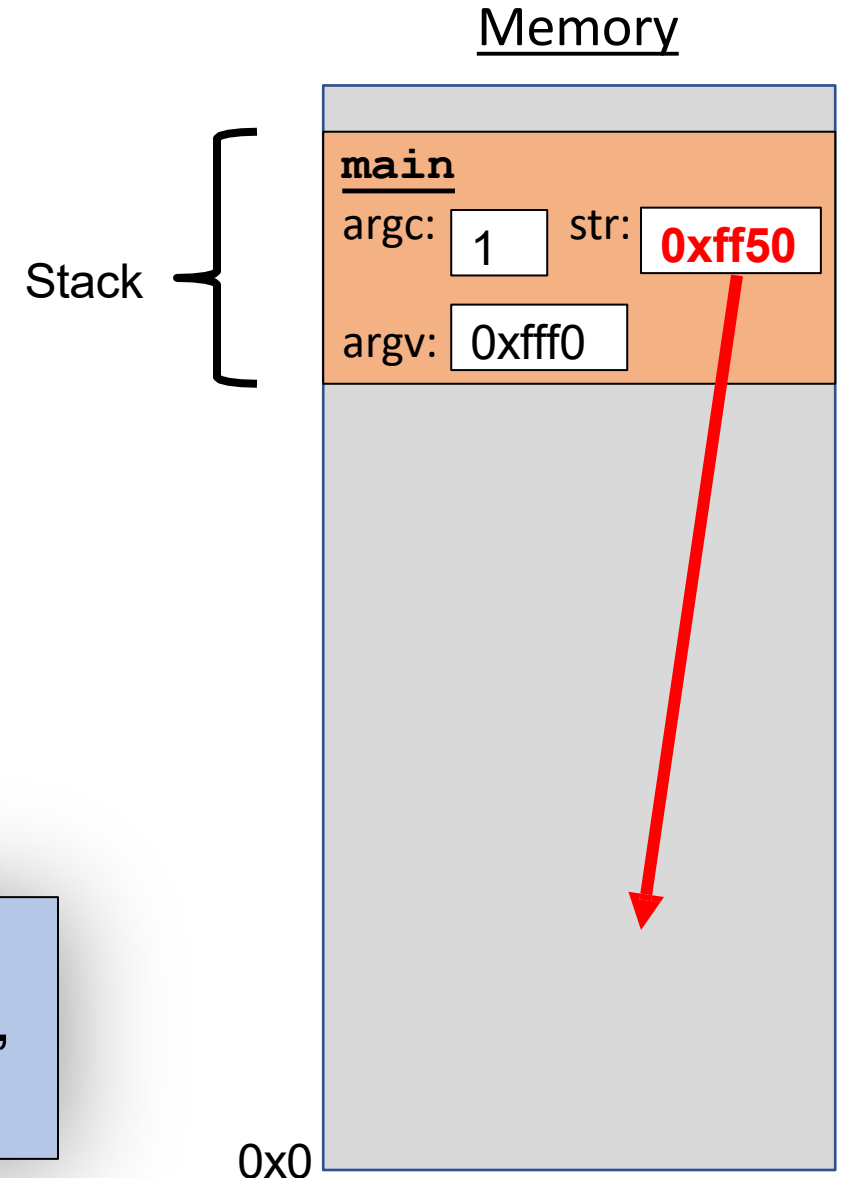
```
char *create_string(char ch, int num) {  
    char new_str[num + 1];  
    for (int i = 0; i < num; i++) {  
        new_str[i] = ch;  
    }  
    new_str[num] = '\0';  
    return new_str;  
}  
  
int main(int argc, char *argv[]) {  
    char *str = create_string('a', 4);  
    printf("%s", str); // want "aaaa"  
    return 0;  
}
```



The Stack

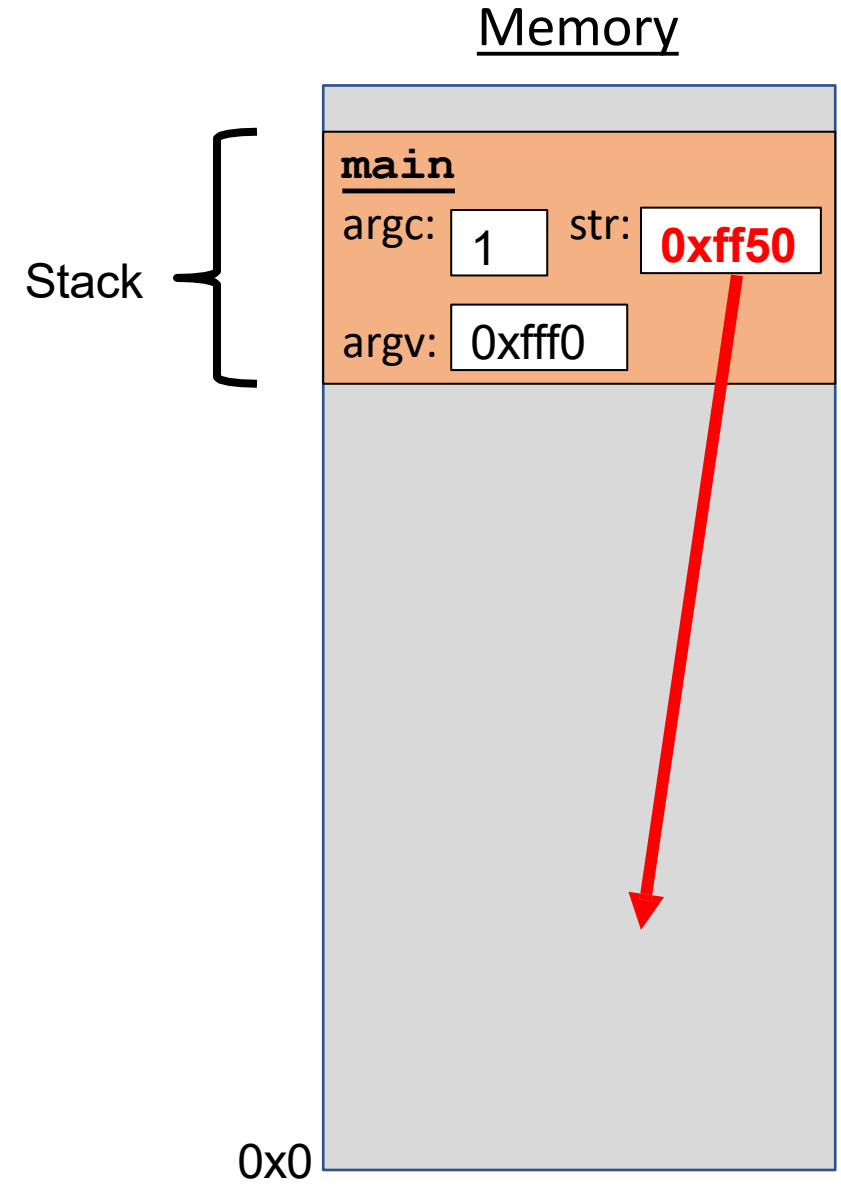
```
char *create_string(char ch, int num) {  
    char new_str[num + 1];  
    for (int i = 0; i < num; i++) {  
        new_str[i] = ch;  
    }  
    new_str[num] = '\0';  
    return new_str;  
}  
  
int main(int argc, char *argv[]) {  
    char *str = create_string('a', 4);  
    printf("%s", str); // want "aaaa"  
    return 0;  
}
```

Problem: local variables go away when a function finishes. These characters will thus no longer exist, and the address will be for unknown memory!



The Stack

```
char *create_string(char ch, int num) {  
    char new_str[num + 1];  
    for (int i = 0; i < num; i++) {  
        new_str[i] = ch;  
    }  
    new_str[num] = '\0';  
    return new_str;  
}  
  
int main(int argc, char *argv[]) {  
    char *str = create_string('a', 4);  
    printf("%s", str); // want "aaaa"  
    return 0;  
}
```



Stacked Against Us

This is a problem! We need a way to have memory that doesn't get cleaned up when a function exits.

Lecture Plan

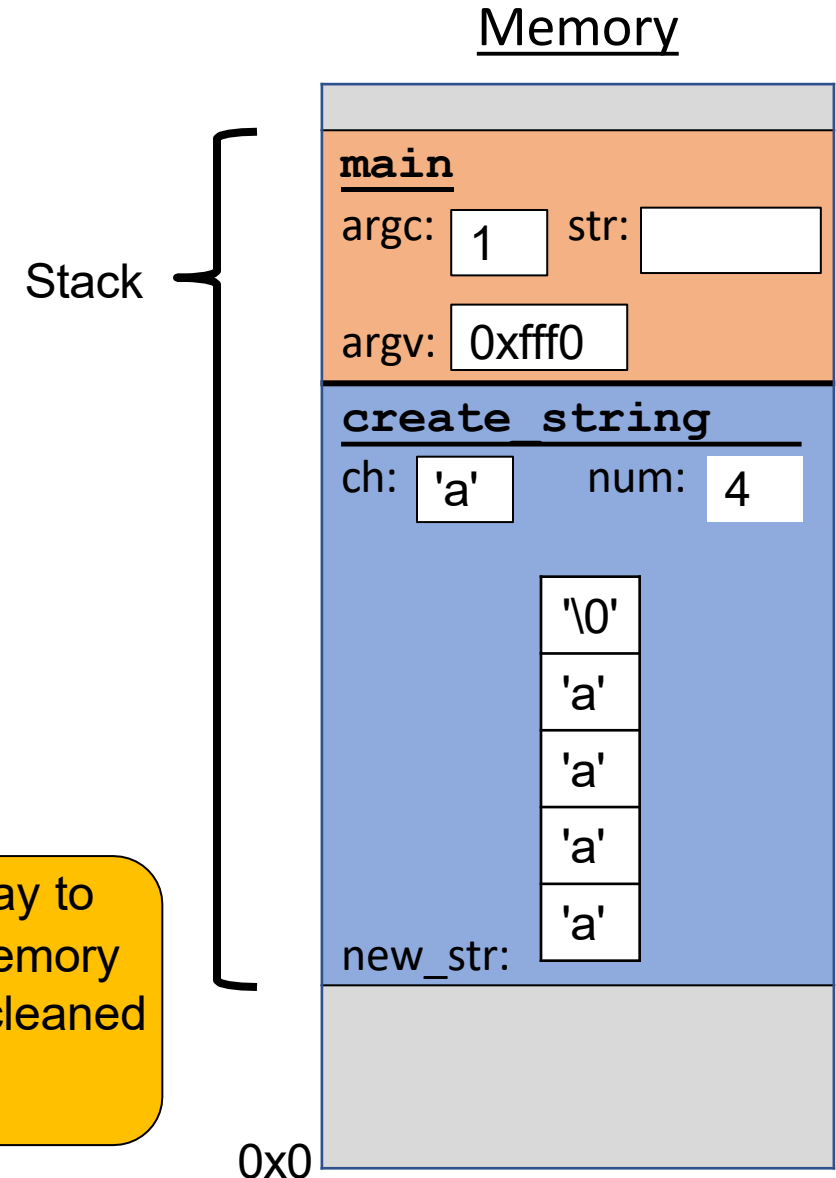
- The Stack
- **The Heap and Dynamic Memory**
- realloc
- Use After Free

The Heap

```
char *create_string(char ch, int num) {  
    char new_str[num + 1];  
    for (int i = 0; i < num; i++) {  
        new_str[i] = ch;  
    }  
    new_str[num] = '\0';  
    return new_str;  
}
```

```
int main(int argc, char *argv[]) {  
    char *str = create_string('a', 4);  
    printf("%s", str); // want "aaaa"  
    return 0;  
}
```

Us: Hey C, is there a way to make this variable in memory that isn't automatically cleaned up?

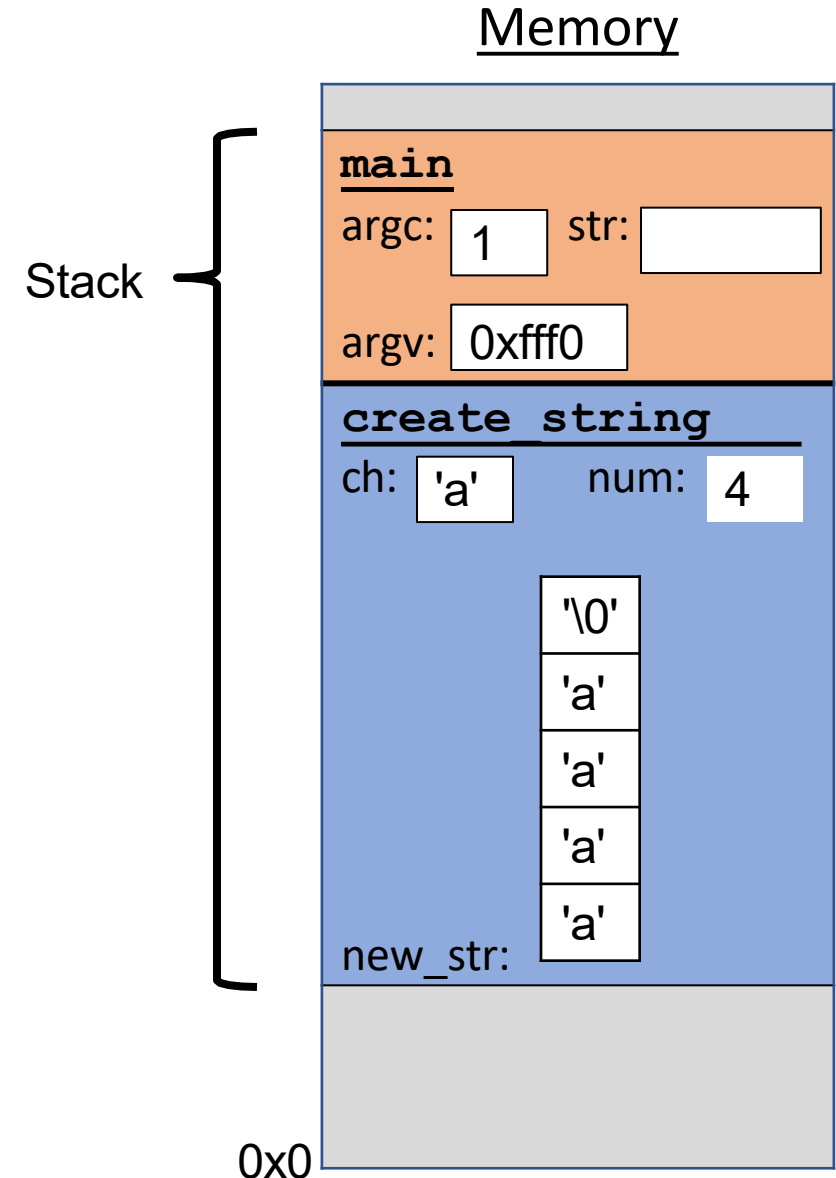


The Heap

```
char *create_string(char ch, int num) {  
    char new_str[num + 1];  
    for (int i = 0; i < num; i++) {  
        new_str[i] = ch;  
    }  
    new_str[num] = '\0';  
    return new_str;  
}
```

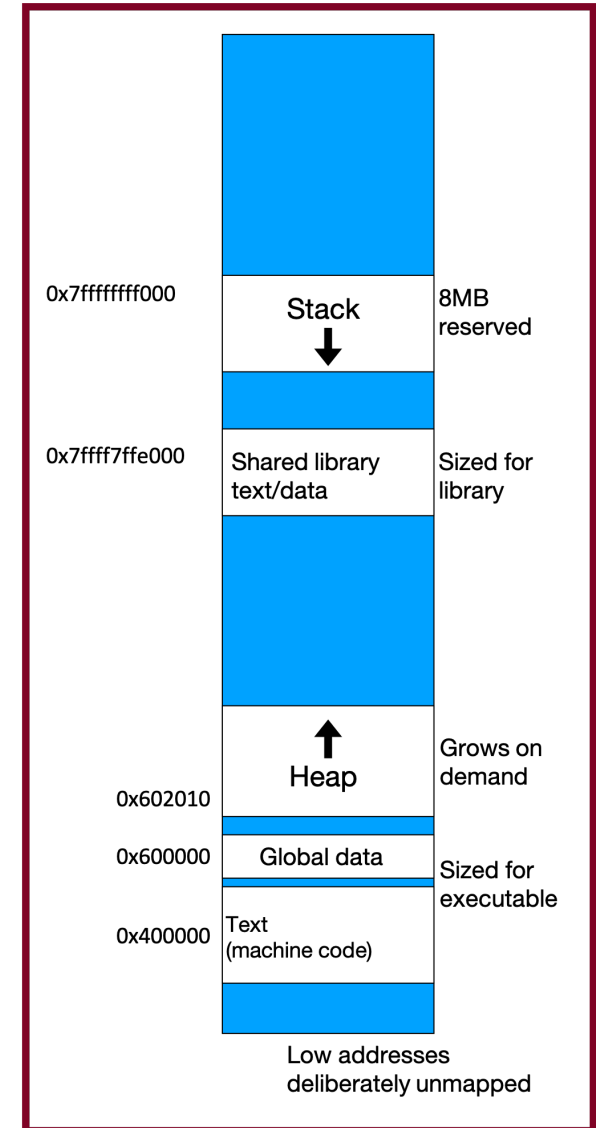
```
int main(int argc, char *argv[]) {  
    char *str = create_string('a', 4);  
    printf("%s", str); // want "aaaa"  
    return 0;  
}
```

C: Sure, but since I don't know when to clean it up anymore, it's your responsibility...



The Heap

- The **heap** is a part of memory that you can manage yourself.
- The **heap** is a part of memory below the stack that you can manage yourself. Unlike the stack, the memory only goes away when you delete it yourself.
- Unlike the stack, the heap grows **upwards** as more memory is allocated.
- The heap is **dynamic memory** – memory that can be allocated, resized, and freed during **program runtime**.



malloc

```
void *malloc(size_t size);
```

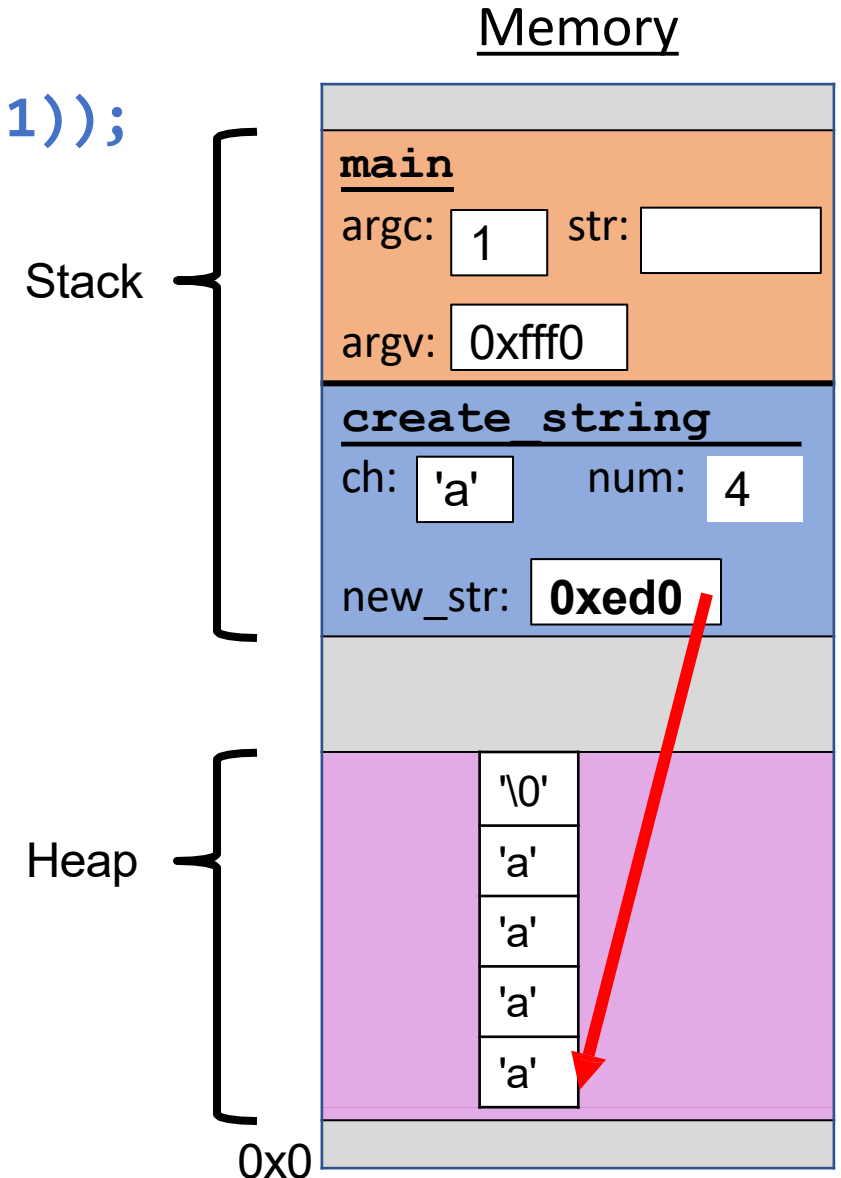
To allocate memory on the heap, use the **malloc** function (“memory allocate”) and specify the number of bytes you’d like.

- This function returns a pointer to *the **starting address** of the new memory*. It doesn’t know or care whether it will be used as an array, a single block of memory, etc.
- **void ***means a pointer to generic memory. You can set another pointer equal to it without any casting.
- The memory is *not* cleared out before being allocated to you!
- If **malloc** returns **NULL**, then there wasn’t enough memory for this request.

The Heap

```
char *create_string(char ch, int num) {  
    char *new_str = malloc(sizeof(char) * (num + 1));  
    for (int i = 0; i < num; i++) {  
        new_str[i] = ch;  
    }  
    new_str[num] = '\0';  
    return new_str;  
}
```

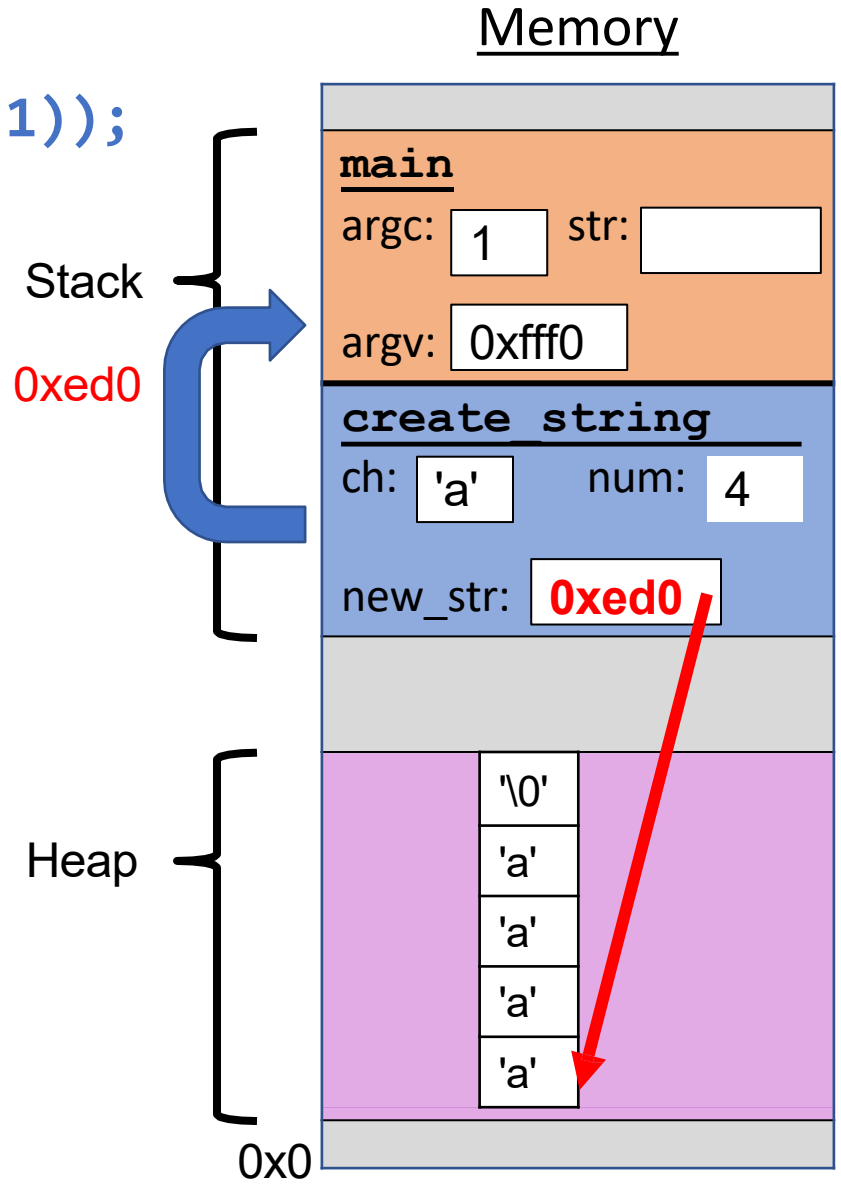
```
int main(int argc, char *argv[]) {  
    char *str = create_string('a', 4);  
    printf("%s", str); // want "aaaa"  
    return 0;  
}
```



The Heap

```
char *create_string(char ch, int num) {  
    char *new_str = malloc(sizeof(char) * (num + 1));  
    for (int i = 0; i < num; i++) {  
        new_str[i] = ch;  
    }  
    new_str[num] = '\\0';  
    return new_str;  
}  
  
int main(int argc, char *argv[]) {  
    char *str = create_string('a', 4);  
    printf("%s", str); // want "aaaa"  
    return 0;  
}
```

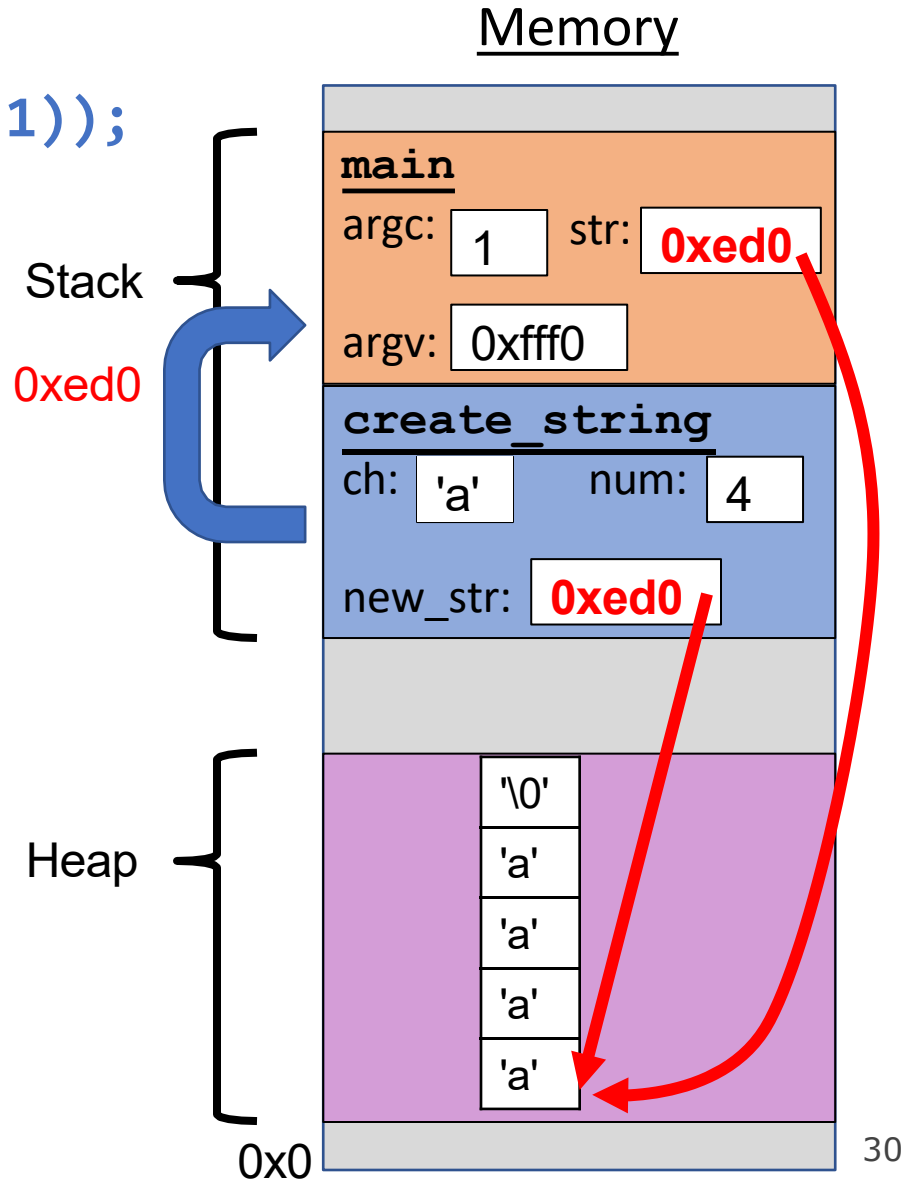
Returns e.g. 0xed0



The Heap

```
char *create_string(char ch, int num) {  
    char *new_str = malloc(sizeof(char) * (num + 1));  
    for (int i = 0; i < num; i++) {  
        new_str[i] = ch;  
    }  
    new_str[num] = '\0';  
    return new_str;  
}  
  
int main(int argc, char *argv[]) {  
    char *str = create_string('a', 4);  
    printf("%s", str); // want "aaaa"  
    return 0;  
}
```

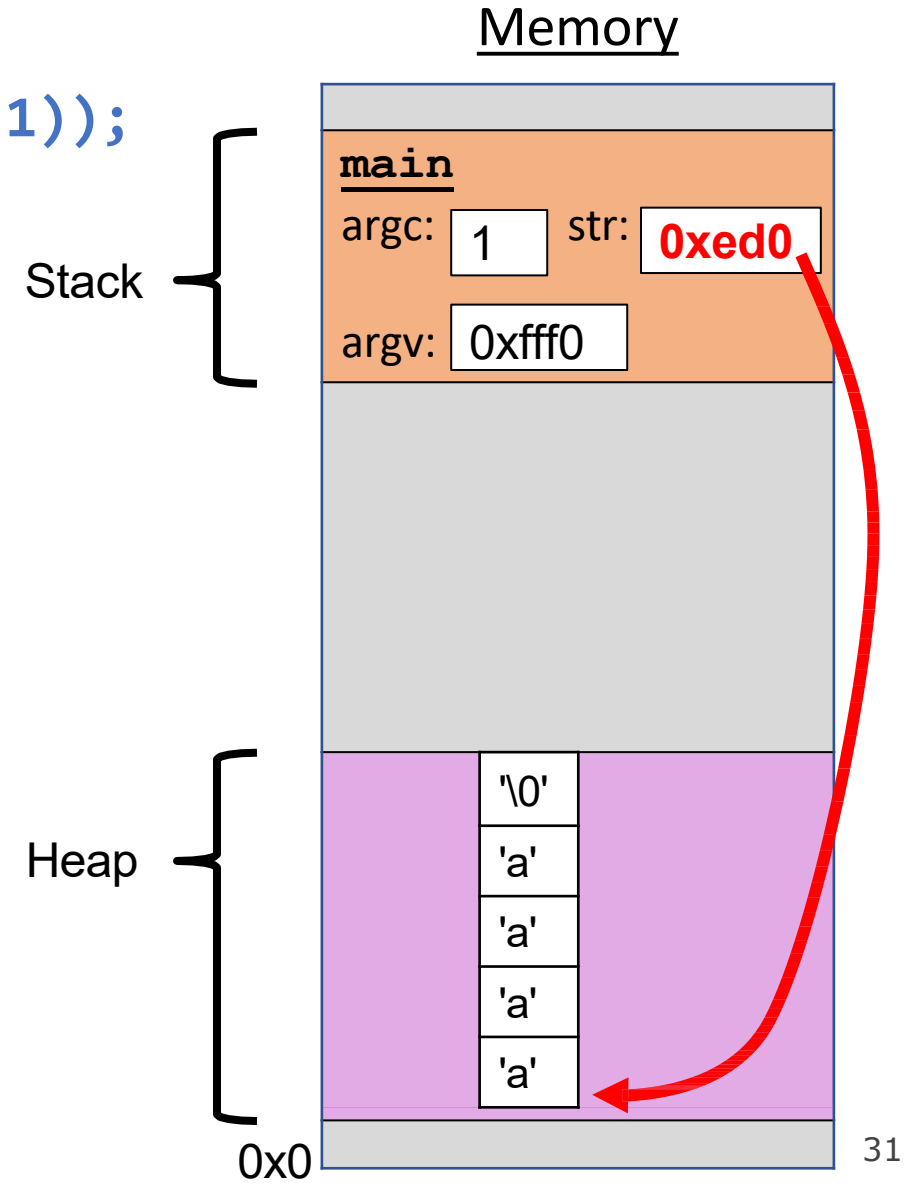
Returns e.g. 0xed0



The Heap

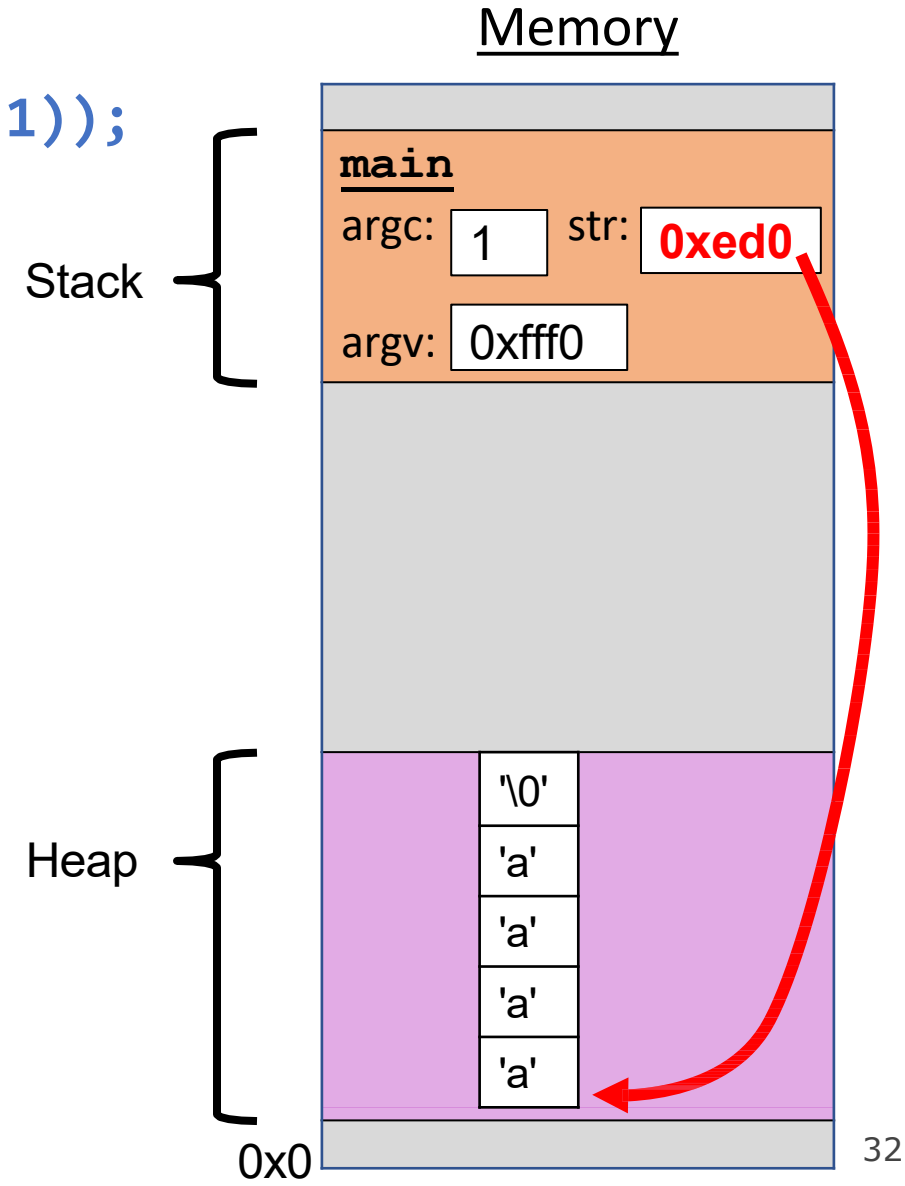
```
char *create_string(char ch, int num) {  
    char *new_str = malloc(sizeof(char) * (num + 1));  
    for (int i = 0; i < num; i++) {  
        new_str[i] = ch;  
    }  
    new_str[num] = '\\0';  
    return new_str;  
}
```

```
int main(int argc, char *argv[]) {  
    char *str = create_string('a', 4);  
    printf("%s", str); // want "aaaa"  
    return 0;  
}
```



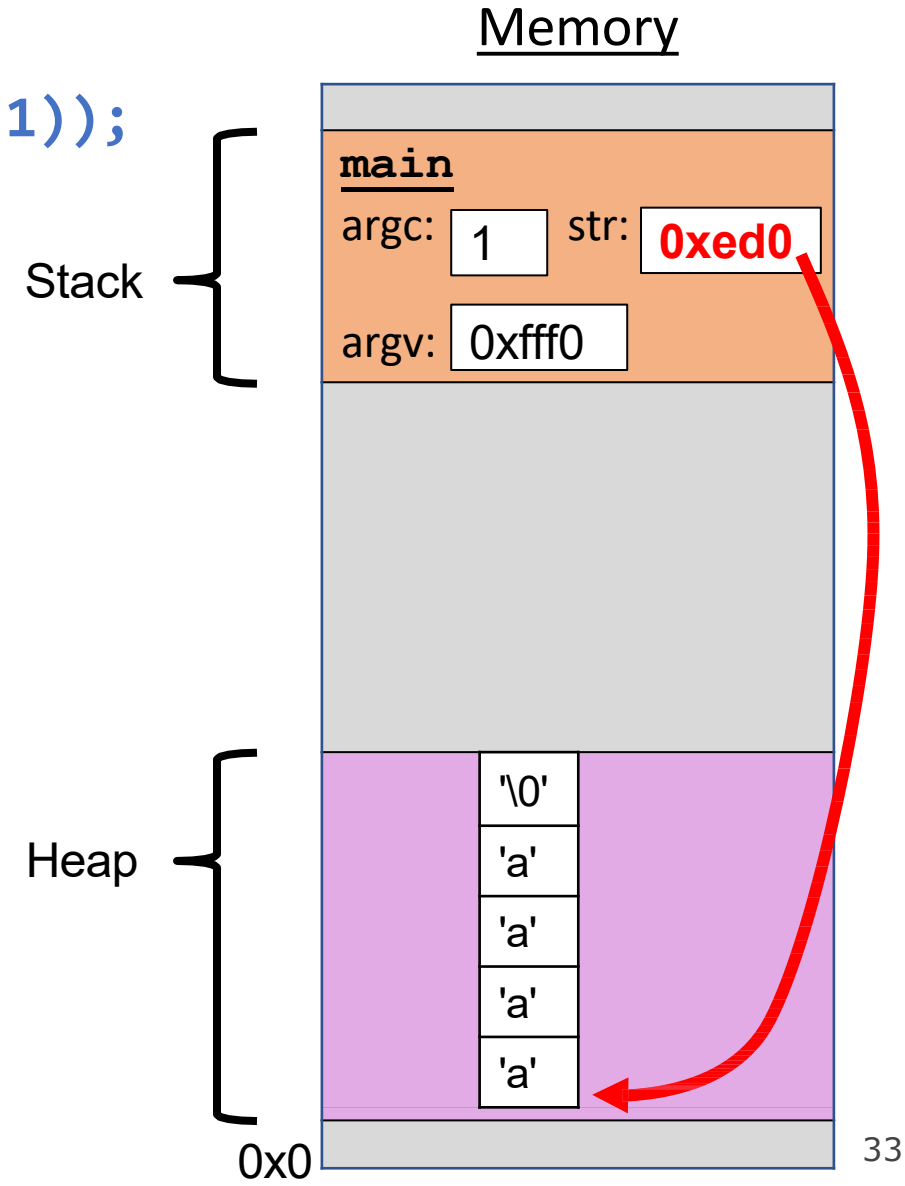
The Heap

```
char *create_string(char ch, int num) {  
    char *new_str = malloc(sizeof(char) * (num + 1));  
    for (int i = 0; i < num; i++) {  
        new_str[i] = ch;  
    }  
    new_str[num] = '\0';  
    return new_str;  
}  
  
int main(int argc, char *argv[]) {  
    char *str = create_string('a', 4);  
    printf("%s", str); // want "aaaa"  
    return 0;  
}
```



The Heap

```
char *create_string(char ch, int num) {  
    char *new_str = malloc(sizeof(char) * (num + 1));  
    for (int i = 0; i < num; i++) {  
        new_str[i] = ch;  
    }  
    new_str[num] = '\0';  
    return new_str;  
}  
  
int main(int argc, char *argv[]) {  
    char *str = create_string('a', 4);  
    printf("%s", str); // want "aaaa"  
    return 0;  
}
```



Exercise: malloc multiples

Let's write a function that returns an array of the first **len** multiples of **mult**

```
1 int *array_of_multiples(int mult, int len)
2 {
3     /* TODO: arr declaration here */
4     for (int i = 0; i < len; i++){
5         arr[i] = mult * (i + 1);
6     }
7     return arr;
8 }
```

Line 2: How should we declare arr?

- A. `int arr[len];`
- B. `int arr[] = malloc(sizeof(int));`
- C. `int *arr = malloc(sizeof(int) * len);`
- D. `int *arr = malloc(sizeof(int) * (len + 1));`
- E. Something else



Exercise: malloc multiples

Let's write a function that returns an array of the first **len** multiples of **mult**

```
1 int *array_of_multiples(int mult, int len)
2 {
3     /* TODO: arr declaration here */
4     for (int i = 0; i < len; i++){
5         arr[i] = mult * (i + 1);
6     }
7     return arr;
8 }
```

- Use a pointer to store the address returned by malloc.
- Malloc's argument is the **number of bytes** to allocate.

⚠ This code is missing an assertion.

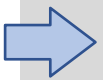
Line 2: How should we declare arr?

- A. `int arr[len];`
- B. `int arr[] = malloc(sizeof(int));`
- ☒ C. `int *arr = malloc(sizeof(int) * len);`
- D. `int *arr = malloc(sizeof(int) * (len + 1));`
- E. Something else



Always assert with the heap

Let's write a function that returns an array of the first **len** multiples of **mult**.



```
1 int *array_of_multiples(int mult, int len){
2     int *arr = malloc(sizeof(int) * len);
3     assert(arr != NULL);
4     for (int i = 0; i < len; i++){
5         arr[i] = mult * (i + 1);
6     }
7     return arr;
8 }
```



- If an allocation error occurs (e.g. out of heap memory!), malloc will return NULL. This is an important case to check **for robustness**.
- **assert** will crash the program if the provided condition is false. A memory allocation error is significant, and we should terminate the program.

Other heap allocations: calloc

```
void *calloc(size_t nmemb, size_t size);
```

calloc is like **malloc** that **zeros out** the memory for you—thanks, **calloc**!

- You might notice its interface is also a little different—it takes two parameters, which are multiplied to calculate the number of bytes (`nmemb * size`).

```
// allocate and zero 20 ints
```

```
int *scores = calloc(20, sizeof(int));
```

```
// alternate (but slower)
```

```
int *scores = malloc(20 * sizeof(int));
```

```
for (int i = 0; i < 20; i++) scores[i] = 0;
```

- **calloc** is more expensive than **malloc** because it zeros out memory. Use only when necessary!

Other heap allocations: strdup

```
char *strdup(const char *s);
```

strdup is a convenience function that returns a **null-terminated**, heap-allocated string with the provided text, instead of you having to **malloc** and copy in the string yourself.

```
char *str = strdup("Hello, world!"); // on heap  
str[0] = 'h';
```

strdup means string duplicate

How can we implement **strdup** using functions we've already seen?

```
1 char *mystrdup(const char *str) {  
2     char *heapstr = _____(A)_____;  
3     _____(B)_____;  
4     _____(C)_____;  
5     return heapstr;  
6 }
```

[Note] Use library functions:

<stdlib.h>: malloc

<assert.h>: assert

<string.h>: strcpy, strlen



strdup means string duplicate

How can we implement **strdup** using functions we've already seen?

```
1 char *mystrdup(const char *str) {  
2     char *heapstr = malloc(strlen(str) + 1);  
3     _____(B)_____;  
4     _____(C)_____;  
5     return heapstr;  
6 }
```

[Note] Use library functions:

<stdlib.h>: malloc

<assert.h>: assert

<string.h>: strcpy, strlen



strdup means string duplicate

How can we implement **strdup** using functions we've already seen?

```
1 char *mystrdup(const char *str) {  
2     char *heapstr = malloc(strlen(str) + 1);  
3     assert(heapstr != NULL);  
4     _____(C)_____;  
5     return heapstr;  
6 }
```

[Note] Use library functions:

<stdlib.h>: malloc

<assert.h>: assert

<string.h>: strcpy, strlen



strdup means string duplicate

How can we implement **strdup** using functions we've already seen?

```
1 char *mystrdup(const char *str) {  
2     char *heapstr = malloc(strlen(str) + 1);  
3     assert(heapstr != NULL);  
4     strcpy(heapstr, str);  
5     return heapstr;  
6 }
```

char arrays differ from other arrays in that valid strings must be null-terminated (i.e., have an extra ending char).

(Note: library strdup doesn't have an assert—it leaves the assert to the caller)

Cleaning Up with free

```
void free(void *ptr);
```

- If we allocated memory on the heap and no longer need it, it is our responsibility to **delete** it.
- To do this, use the **free** command and pass in the *starting address on the heap for the memory you no longer need*.
- Example:

```
char *bytes = malloc(4);  
... free(bytes);
```

Free

```
void free(void *ptr);
```

When you free an allocation, you are freeing up what it *points* to. You are not freeing the pointer itself. You can still use the pointer to point to something else.

```
char *str = strdup("hello");
```

```
...
```

```
free(str);
```

```
str = strdup("hi");
```



free details

Even if you have multiple pointers to the same block of memory, each memory block should only be freed once.

```
char *bytes = malloc(4);  
char *ptr = bytes;
```

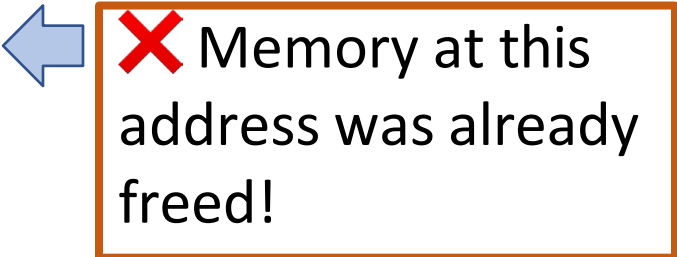
...

```
free(bytes);
```



...

```
free(ptr);
```



You must free the address you received in the previous allocation call; you cannot free just part of a previous allocation.

```
char *bytes = malloc(4);  
char *ptr = malloc(10);
```

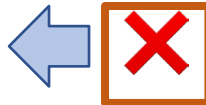
...

```
free(bytes);
```



...

```
free(ptr + 1);
```



Cleaning Up

You may need to free memory allocated by other functions if that function expects the caller to handle memory cleanup.

```
char *str = strdup("Hello!");
```

```
...
```

```
free(str);    // our responsibility to free!
```

Free

```
void free(void *ptr);
```

A **memory leak** is when you do not free memory you previously allocated.

```
char *str = strdup("hello");
```

```
...
```

```
str = strdup("hi"); // memory leak! Lost previous str
```

Recommendation: Don't worry about putting in frees until **after** you're finished with functionality.

Memory leaks will rarely crash your CS107 programs.



Memory Leaks

- A memory **leak** is when you **allocate** memory on the heap, **but** do **not free** it.
- **Your** program should be **responsible for cleaning up** any memory it allocates but no longer needs.
- If you never free any memory and allocate an extremely large amount, you may run out of memory in the heap!

However, memory leaks rarely (if ever) cause crashes.

- We recommend not to worry about freeing memory until your program is written. Then, go back and free memory as appropriate.
- **Valgrind** is a very helpful tool for finding memory leaks!

Lecture Plan

- The Stack
- The Heap and Dynamic Memory
- **realloc**
- Use After Free

realloc

```
void *realloc(void *ptr, size_t size);
```

- The **realloc** function takes an existing allocation pointer and enlarges to a new requested size. It returns the new pointer.
- If *there is enough space* after the existing memory block on the heap for the new size, **realloc** simply adds that space to the allocation.
- If there is *not enough space*, **realloc** *moves the memory to a larger location*, frees the old memory for you, and *returns a pointer to the new location*.

realloc

```
char *str = strdup("Hello");  
assert(str != NULL);
```

...

```
// want to make str longer to hold "Hello world!"
```

```
char *addition = " world!";  
str = realloc(str, strlen(str) + strlen(addition) + 1);  
assert(str != NULL);
```

```
strcat(str, addition);  
printf("%s", str);  
free(str);
```

realloc

- realloc **only** accepts pointers that were previously returned by malloc/etc.
- Make sure to **not** pass pointers to the middle of heap-allocated memory.
- Make sure to **not** pass pointers to stack memory.

Cleaning Up with free and realloc

You only need to free the new memory coming out of realloc—the previous (smaller) one was already reclaimed by realloc.

```
char *str = strdup("Hello");  
assert(str != NULL);  
...  
// want to make str longer to hold "Hello world!"  
char *addition = " world!";  
str = realloc(str, strlen(str) + strlen(addition) + 1);  
assert(str != NULL);  
strcat(str, addition);  
printf("%s", str);  
free(str);
```



Heap allocator analogy: A hotel

Request memory by size (`malloc`)

- Receive room key to first of connecting rooms

Need more room? (`realloc`)

- Extend into connecting room if available
- If not, trade for new room, employee moves your stuff for you

Check out when done (`free`)

- You need to remember your room number though



Errors! What happens if you...

- Forget to check out?
- Bust through connecting door to neighbor? What if the room is in use? Yikes...
- Return to your room after checkout?

Heap Allocation: Error prone

Invalid usage of heap functions includes when you:

- Realloc a stack pointer
- Free the middle of an allocation
- Free a stack pointer
- Etc....

Working with the heap

Working with the heap consists of 3 core steps:

1. Allocate memory with malloc/realloc/strdup/calloc
2. Assert heap pointer is not NULL
3. Free when done

The heap is **dynamic memory**, so you may encounter many **runtime errors**, even if your code compiles!

Goodbye, Free Memory

Where/how should we free memory below so that all memory is freed properly?

```
1 char *str = strdup("Hello");
2 assert(str != NULL);
3 char *ptr = str + 1;
4 for (int i = 0; i < 5; i++) {
5     int *num = malloc(sizeof(int));
6     *num = i;
7     printf("%s %d\n", ptr, *num);
8     free(num);
9 }
10 printf("%s\n", str);
11 free(str);
```

Recommendation: Don't worry about putting in frees until **after** you're finished with functionality.

Memory leaks will rarely crash your CS107 programs.

//Leak Check: valgrind --leak-check=full --show-leak-kinds=all ...

How Would We Improve This?

Correct and leak-free — but look closely at how this code uses the heap.

```
1 char *str = strdup("Hello");
2 assert(str != NULL);
3 char *ptr = str + 1;
4 for (int i = 0; i < 5; i++) {
5     int *num = malloc(sizeof(int));
6     *num = i;
7     printf("%s %d\n", ptr, *num);
8     free(num);
9 }
10 printf("%s\n", str);
11 free(str);
```

Look closely at num:

Each pass allocates 4 bytes, uses them once, and frees them — 5 malloc/free pairs in all.

Could one allocation serve all five iterations?

Move malloc Out of the Loop

Step 1: allocate num once, before the loop, and reuse it each iteration.

```
1 char *str = strdup("Hello");
2 assert(str != NULL);
3 char *ptr = str + 1;
4 int *num = malloc(sizeof(int));
5 assert(num != NULL);
6 for (int i = 0; i < 5; i++) {
7     *num = i;
8     printf("%s %d\n", ptr, *num);
9 }
10 printf("%s\n", str);
11 free(str);
```

One allocation now serves every iteration — the loop just overwrites *num each time.

But num is never freed now. Where should free(num) go?

Not in the loop — pass 1 would free it, and pass 2 would write to freed memory!

...Then Move free Out, Too

Step 2: free num once, after the loop — one malloc, one matching free.

```
1 char *str = strdup("Hello");
2 assert(str != NULL);
3 char *ptr = str + 1;
4 int *num = malloc(sizeof(int));
5 assert(num != NULL);
6 for (int i = 0; i < 5; i++) {
7     *num = i;
8     printf("%s %d\n", ptr, *num);
9 }
10 free(num);
11 printf("%s\n", str);
12 free(str);
```

Allocate once, reuse, free once.

Each malloc now has exactly one free at the same scope — and valgrind reports 0 bytes lost.

Sharp eyes: this int never needed the heap at all — int num; on the stack works too!

strcat_extend

Write a function that takes in a heap-allocated **str1**, enlarges it, and concatenates **str2** onto it.

```
1 char *strcat_extend(char *heap_str, const char *concat_str) {  
2     (_____(1)_____);  
3     heap_str = realloc(____(2A)____, ____ (2B)____);  
4     (_____(3)_____);  
5     strcat(____(4A)____, ____ (4B)____);  
6     return heap_str;  
7 }
```

Example usage:

```
char *str = strdup("Hello ");  
str = strcat_extend(str, "world!");  
printf("%s\n", str);  
free(str);
```

strcat_extend

Write a function that takes in a heap-allocated **str1**, enlarges it, and concatenates **str2** onto it.

```
1 char *strcat_extend(char *heap_str, const char *concat_str) {  
2     int new_length = strlen(heap_str) + strlen(concat_str) + 1;  
3     heap_str = realloc(heap_str, new_length);  
4     assert(heap_str != NULL);  
5     strcat(heap_str, concat_str);  
6     return heap_str;  
7 }
```

Example usage:

```
char *str = strdup("Hello ");  
str = strcat_extend(str, "world!");  
printf("%s\n", str);  
free(str);
```

Heap allocation interface: A summary

```
void *malloc(size_t size);  
void *calloc(size_t nmemb, size_t size);  
void *realloc(void *ptr, size_t size);  
char *strdup(const char *s);  
void free(void *ptr);
```

Compare and contrast the heap memory functions we've learned about.



Heap allocation interface: A summary

```
void *malloc(size_t size);  
void *calloc(size_t nmemb, size_t size);  
void *realloc(void *ptr, size_t size);  
char *strdup(const char *s);  
void free(void *ptr);
```

Heap **memory allocation** guarantee:

- NULL on failure, so check with assert
- Memory is contiguous; it is not recycled unless you call free
- realloc preserves existing data
- calloc zero-initializes bytes, malloc and realloc do not

Undefined behavior occurs:

- If you overflow (i.e., you access beyond bytes allocated)
- If you use after free, or if free is called twice on a location.
- If you realloc/free non-heap address

Engineering principles: stack vs heap

Stack (“local variables”)

- **Fast**
Fast to allocate/deallocate
- **Convenient.**
Automatic allocation/ deallocation;
declare/initialize in one step
- **Reasonable type safety**
Thanks to the compiler
- ⚠ **Not especially plentiful**
Total stack size fixed, default 8MB
- ⚠ **Somewhat inflexible**
Cannot add/resize at runtime, scope
dictated by control flow in/out of functions

Heap (dynamic memory)

Engineering principles: stack vs heap

Stack (“local variables”)

- **Fast**
Fast to allocate/deallocate; okay to oversize
- **Convenient.**
Automatic allocation/ deallocation;
declare/initialize in one step
- **Reasonable type safety**
Thanks to the compiler
- ⚠ **Not especially plentiful**
Total stack size fixed, default 8MB
- ⚠ **Somewhat inflexible**
Cannot add/resize at runtime, scope
dictated by control flow in/out of functions

Heap (dynamic memory)

- **Plentiful.**
Can provide more memory on demand!
- **Very flexible.**
Runtime decisions about how much/when to
allocate, can resize easily with realloc
- **Scope under programmer control**
Can precisely determine lifetime
- ⚠ **Lots of opportunity for error**
Low type safety, forget to allocate/free
before done, allocate wrong size, etc.,
Memory leaks (much less critical)

Stack and Heap

- Generally, unless a situation requires dynamic allocation, stack allocation is preferred. Often both techniques are used together in a program.
- Heap allocation is a necessity when:
 - you have a very large allocation that could blow out the stack
 - you need to control the memory lifetime, or memory must persist outside of a function call
 - you need to resize memory after its initial allocation

Lecture Plan

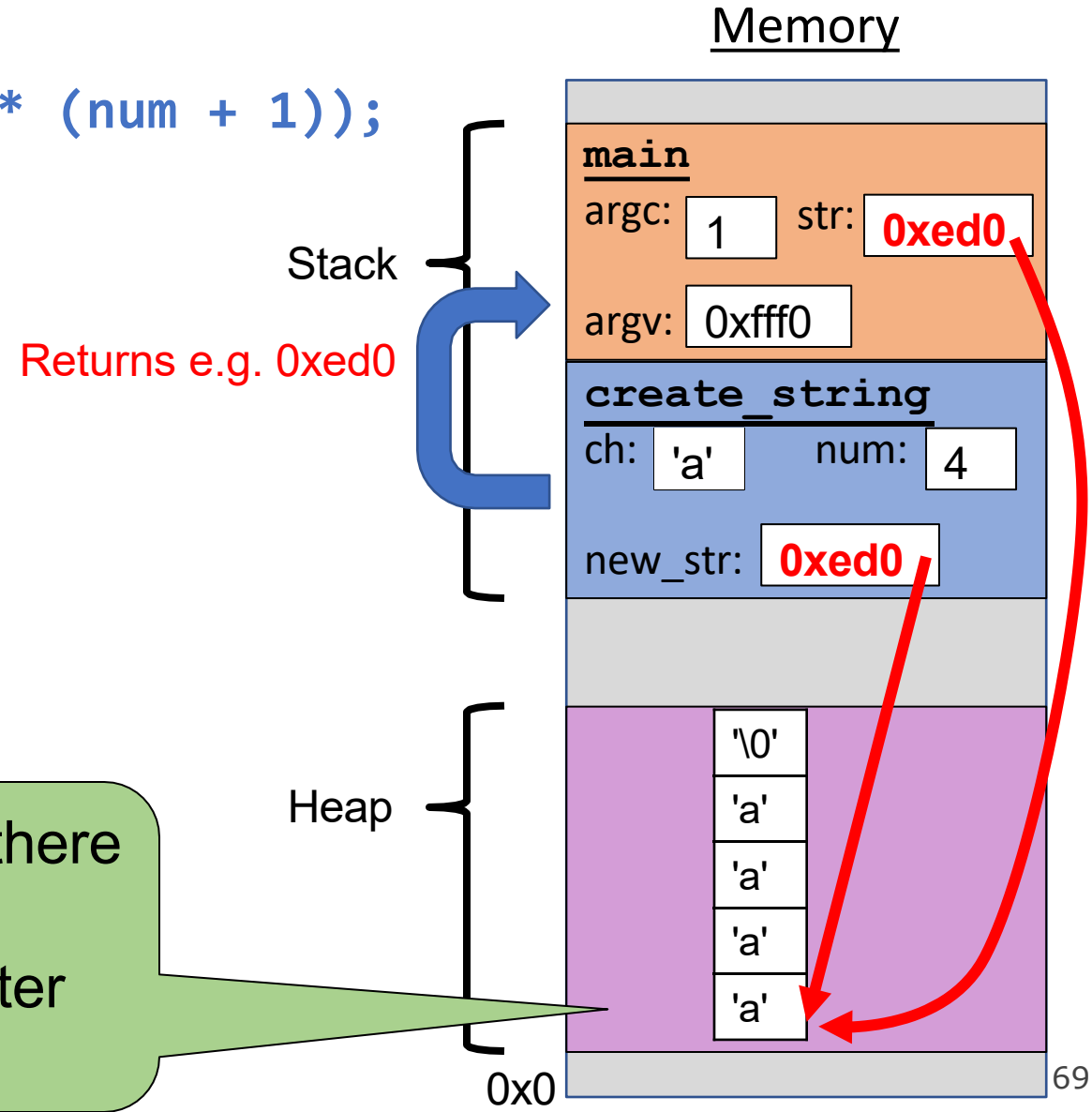
- The Stack
- The Heap and Dynamic Memory
- realloc
- **Use After Free**

Recall: The Heap

```
char *create_string(char ch, int num) {  
    char *new_str = malloc(sizeof(char) * (num + 1));  
    for (int i = 0; i < num; i++) {  
        new_str[i] = ch;  
    }  
    new_str[num] = '\\0';  
    return new_str;  
}
```

```
int main(int argc, char *argv[]) {  
    char *str = create_string('a', 4);  
    printf("%s", str); // want "aaaa"  
    return 0;  
}
```

C: Nice memory you got there
... would be a shame if
someone tried to use it after
you freed it.



Use after Free

What happens if we attempt to reference something on the heap after it has been freed?

```
char *bytes = malloc(4);  
char *ptr = bytes;  
...  
free(bytes);  
...  
strncpy(ptr, argv[1], 3);
```

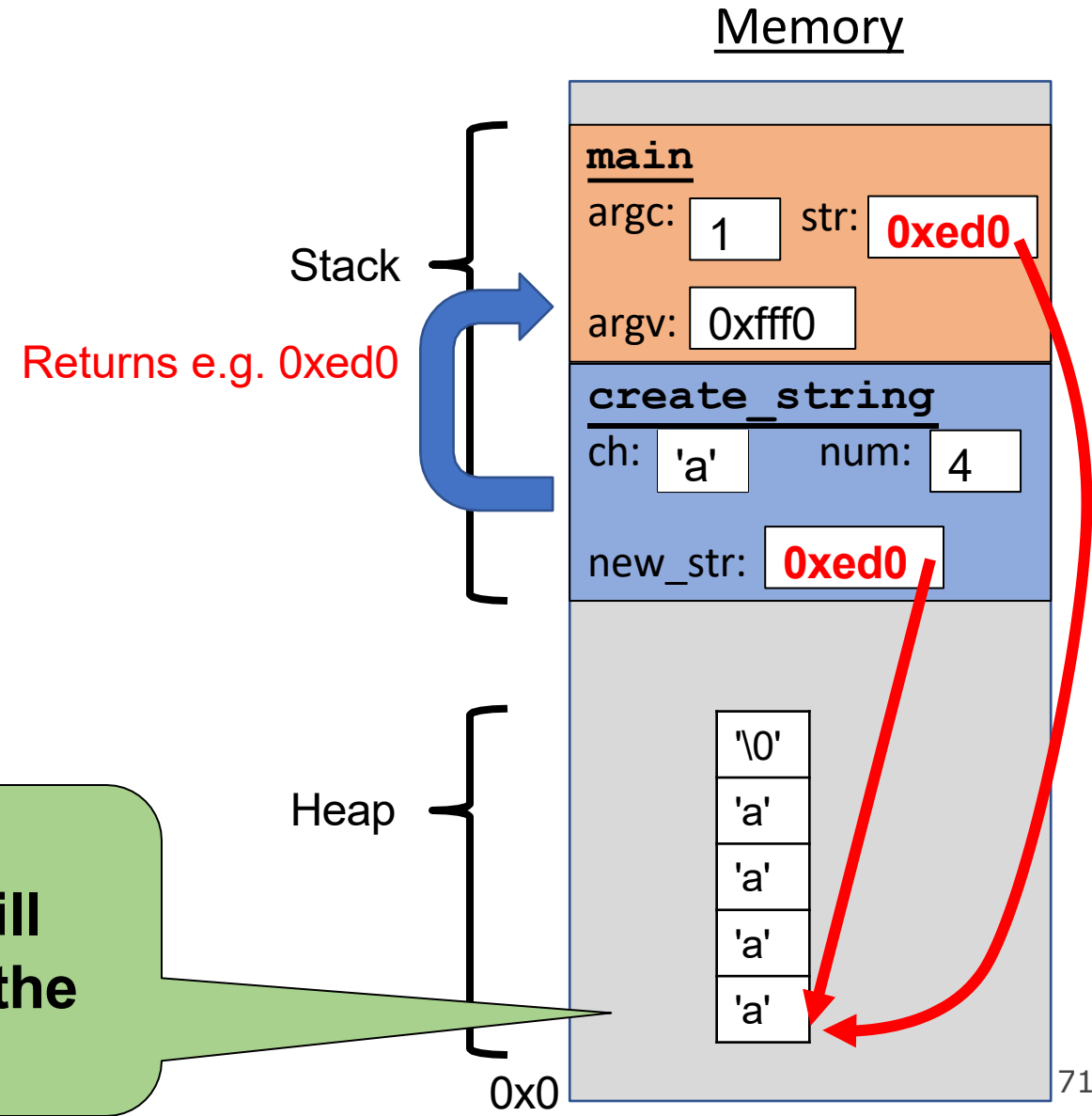
← We freed `bytes` but did not set `ptr` to NULL

✗ Memory at this address was already freed, but now we are using it!

What Could Happen?

- Each program has its own memory space (heap + stack)
- Vulnerabilities are within the context of a program.

The memory has been freed, but the pointer still reaches its location on the heap.



“undefined behavior occurs”

Using a pointer to the heap after the pointer's memory has been freed results in undefined behavior because the memory could have its original contents or could have been overwritten.

This undefined behavior also creates a vulnerability and an opportunity to hack.

**How can we use our
knowledge of memory and
data representation to
write code that works with
any data type?**

Learning Goals

- Learn how to write C code that works with any data type.
- Learn about using void * and avoiding potential pitfalls.
- Before we dive in, remember void * is just like any other type
 - This means it is just a way of saying:
 - This variable stores the address of an unspecified type of data
 - Information still lives in memory
 - Whatever information is there still has whatever structure it started with

Generics

- We always strive to write code that is as general-purpose as possible.
- Generic code reduces code duplication and means you can make improvements and fix bugs in one place rather than many.
- Generics are used throughout C for common functions like:
 - Sorting
 - Searching
 - Free-ing arbitrary memory
 - And more.
- How can we write generic code in C?



Imagine

**Let's imagine we need to
implement some swapping for our
team or a client**

Generic Lecture Plan

- **Generic Swap**
- Generics Pitfalls
- Generic Swap Ends
- Generic Stack
- Array Rotation

Swap

You're asked to write a function that swaps two numbers.

```
void swap_int(int *a, int *b) {  
    int temp = *a;  
    *a = *b;  
    *b = temp;  
}  
  
int main(int argc, char *argv[]) {  
    int x = 2;  
    int y = 5;  
    swap_int(&x, &y);  
    // want x = 5, y = 2  
    printf("x = %d, y = %d\n", x, y);  
    return 0;  
}
```

Swap

You're asked to write a function that swaps two numbers.

```
void swap_int(int *a, int *b) {  
    int temp = *a;  
    *a = *b;  
    *b = temp;  
}  
  
int main(int argc, char *argv[]) {  
    int x = 2;  
    int y = 5;  
    swap_int(&x, &y);  
    // want x = 5, y = 2  
    printf("x = %d, y = %d\n", x, y);  
    return 0;  
}
```

main()



		Stack	
		Address	Value
		x	0xff14
		y	0xff10
			...

Swap

You're asked to write a function that swaps two numbers.

```
void swap_int(int *a, int *b) {  
    int temp = *a;  
    *a = *b;  
    *b = temp;  
}  
  
int main(int argc, char *argv[]) {  
    int x = 2;  
    int y = 5;  
    swap_int(&x, &y);  
    // want x = 5, y = 2  
    printf("x = %d, y = %d\n", x, y);  
    return 0;  
}
```

main()

swap_int()

		Stack
		Address Value
x	0xff14	...
		2
y	0xff10	5
		...
b	0xf18	0xff10
a	0xf10	0xff14
		...

Swap

You're asked to write a function that swaps two numbers.

```
void swap_int(int *a, int *b) {  
    int temp = *a;  
    *a = *b;  
    *b = temp;  
}  
  
int main(int argc, char *argv[]) {  
    int x = 2;  
    int y = 5;  
    swap_int(&x, &y);  
    // want x = 5, y = 2  
    printf("x = %d, y = %d\n", x, y);  
    return 0;  
}
```

main()

swap_int()

		Stack
		Address Value
x	0xff14	...
		2
		5
y	0xff10	...
		0xff10
		0xff14
temp	0xf0c	...
		2
		...

Swap

You're asked to write a function that swaps two numbers.

```
void swap_int(int *a, int *b) {  
    int temp = *a;  
    *a = *b;  
    *b = temp;  
}  
  
int main(int argc, char *argv[]) {  
    int x = 2;  
    int y = 5;  
    swap_int(&x, &y);  
    // want x = 5, y = 2  
    printf("x = %d, y = %d\n", x, y);  
    return 0;  
}
```

main()

swap_int()

		Stack
		Value
Address		
		...
x	0xff14	5
y	0xff10	5
		...
b	0xf18	0xff10
a	0xf10	0xff14
temp	0xf0c	2
		...

Swap

You're asked to write a function that swaps two numbers.

```
void swap_int(int *a, int *b) {  
    int temp = *a;  
    *a = *b;  
    *b = temp;  
}  
  
int main(int argc, char *argv[]) {  
    int x = 2;  
    int y = 5;  
    swap_int(&x, &y);  
    // want x = 5, y = 2  
    printf("x = %d, y = %d\n", x, y);  
    return 0;  
}
```

main()

swap_int()

		Stack
		Value
Address		
		...
x	0xff14	5
y	0xff10	2
		...
b	0xf18	0xff10
a	0xf10	0xff14
temp	0xf0c	2
		...

Swap

You're asked to write a function that swaps two numbers.

```
void swap_int(int *a, int *b) {  
    int temp = *a;  
    *a = *b;  
    *b = temp;  
}
```

```
int main(int argc, char *argv[]) {  
    int x = 2;  
    int y = 5;  
    swap_int(&x, &y);  
    // want x = 5, y = 2  
    printf("x = %d, y = %d\n", x, y);  
    return 0;  
}
```

main()

		Stack	
Address		Value	
		...	
x	0xff14	5	
y	0xff10	2	
		...	

**“Oh, when I said ‘numbers’
I meant shorts, not ints.”**



Swap

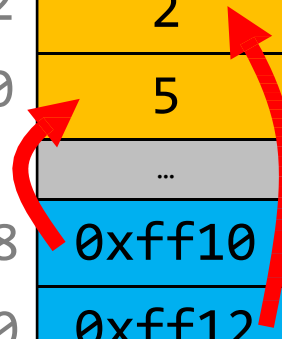
```
void swap_short(short *a, short *b) {  
    short temp = *a;  
    *a = *b;  
    *b = temp;  
}
```

```
int main(int argc, char *argv[]) {  
    short x = 2;  
    short y = 5;  
    swap_short(&x, &y);  
    // want x = 5, y = 2  
    printf("x = %d, y = %d\n", x, y);  
    return 0;  
}
```

main()

swap_short()

		Stack	
		Address	Value
x	0xff12		...
			2
			5
y	0xff10		...
			0xff10
			0xff12
b	0xf18		...
			2
			...



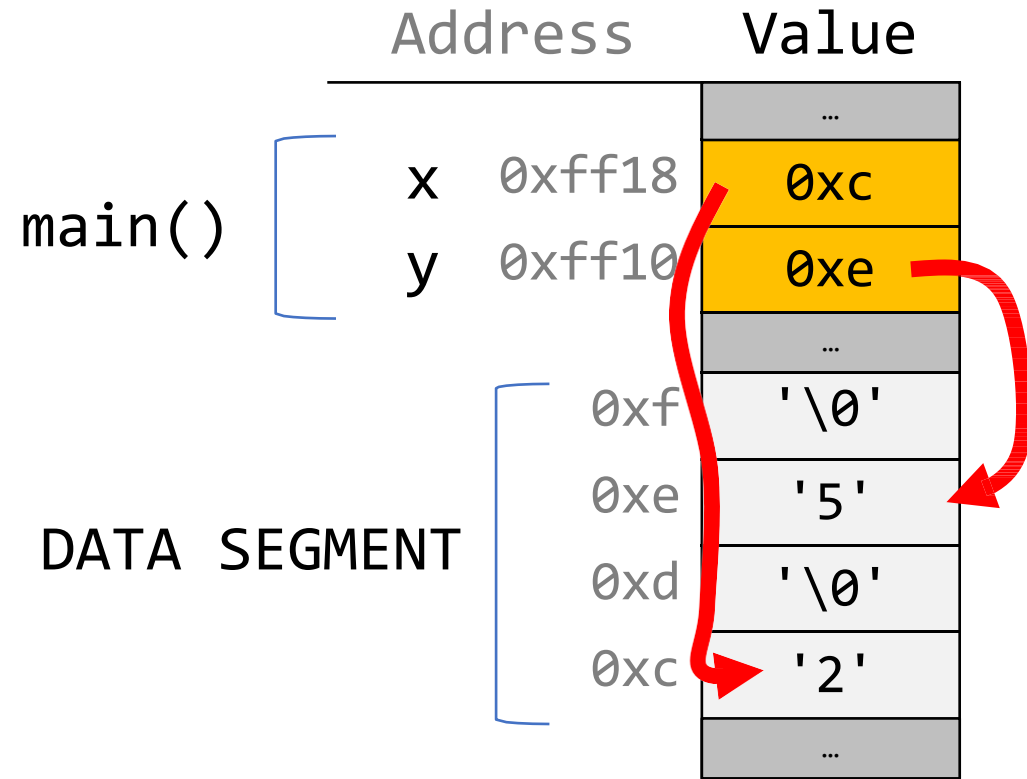
**“You know what, I goofed.
We’re going to use strings.
Could you write something
to swap those?”**



Swap

```
void swap_string(char **a, char **b) {  
    char *temp = *a;  
    *a = *b;  
    *b = temp;  
}
```

```
int main(int argc, char *argv[]) {  
    char *x = "2";  
    char *y = "5";  
    swap_string(&x, &y);  
    // want x = 5, y = 2  
    printf("x = %s, y = %s\n", x, y);  
    return 0;  
}
```



Swap

```
void swap_string(char **a, char **b) {  
    char *temp = *a;  
    *a = *b;  
    *b = temp;  
}
```

```
int main(int argc, char *argv[]) {  
    char *x = "2";  
    char *y = "5";  
    swap_string(&x, &y);  
    // want x = 5, y = 2  
    printf("x = %s, y = %s\n", x, y);  
    return 0;  
}
```

main()

swap_string()

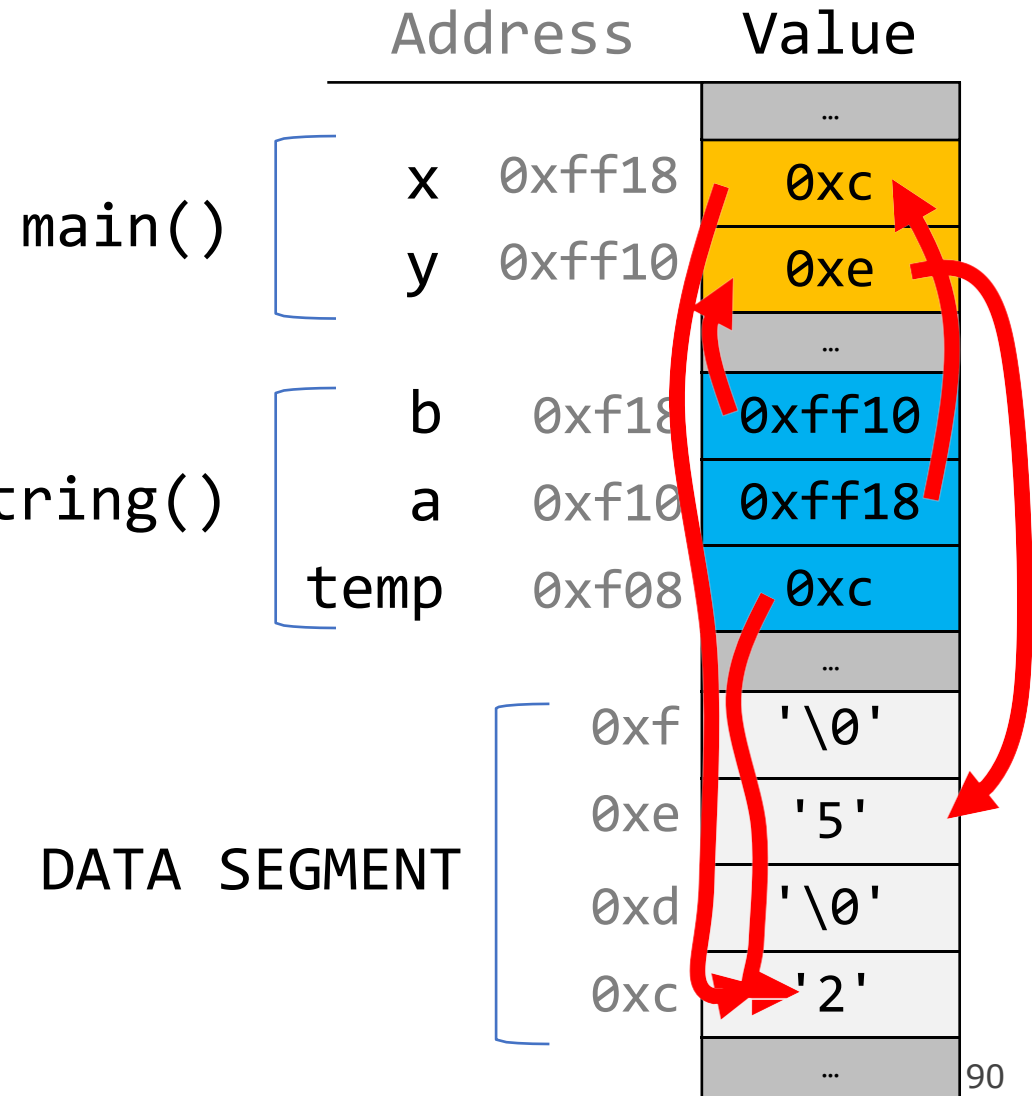
DATA SEGMENT

Address		Value
		...
x	0xff18	0xc
y	0xff10	0xe
		...
b	0xf18	0xff10
a	0xf10	0xff18
		...
	0xf	'\0'
	0xe	'5'
	0xd	'\0'
	0xc	'2'
		...

Swap

```
void swap_string(char **a, char **b) {  
    char *temp = *a;  
    *a = *b;  
    *b = temp;  
}
```

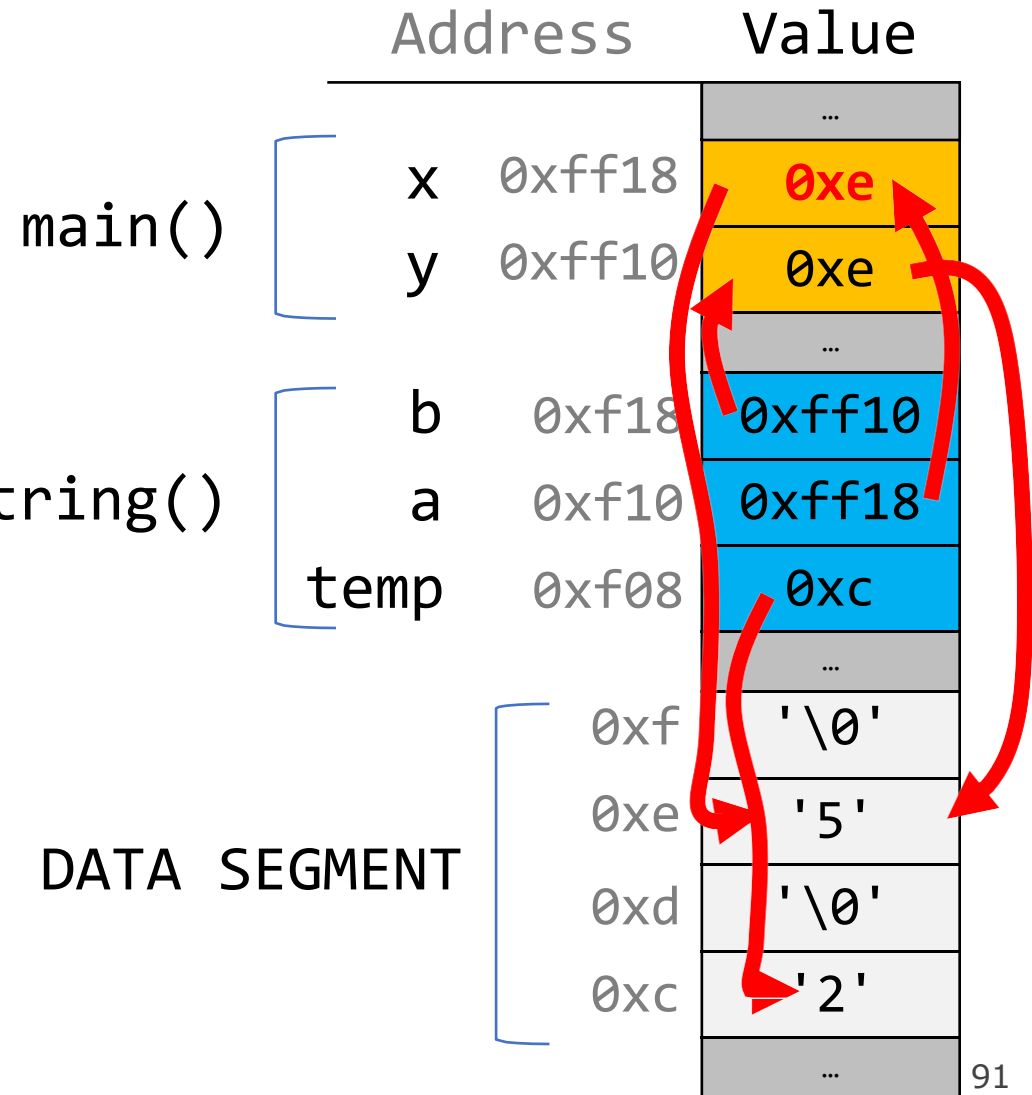
```
int main(int argc, char *argv[]) {  
    char *x = "2";  
    char *y = "5";  
    swap_string(&x, &y);  
    // want x = 5, y = 2  
    printf("x = %s, y = %s\n", x, y);  
    return 0;  
}
```



Swap

```
void swap_string(char **a, char **b) {  
    char *temp = *a;  
    *a = *b;  
    *b = temp;  
}
```

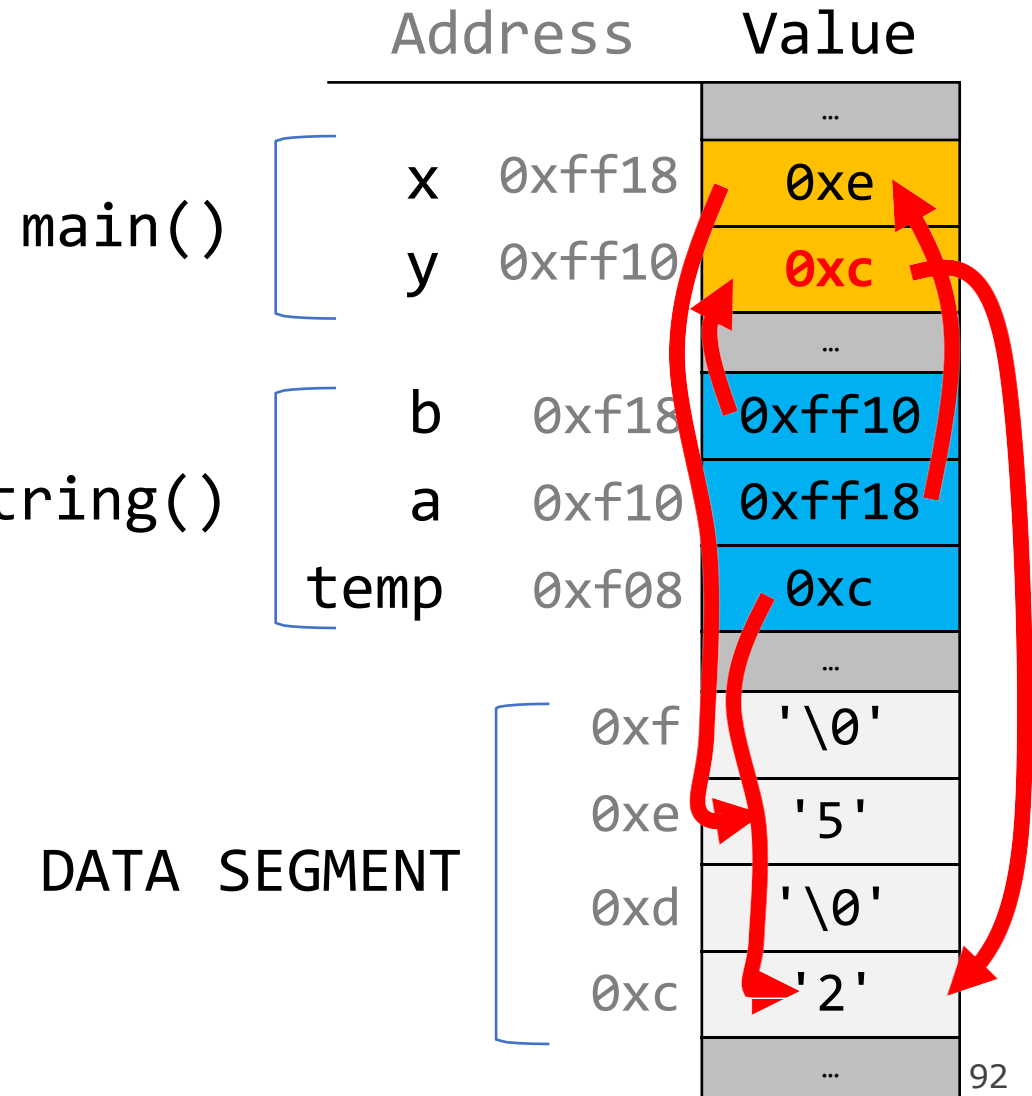
```
int main(int argc, char *argv[]) {  
    char *x = "2";  
    char *y = "5";  
    swap_string(&x, &y);  
    // want x = 5, y = 2  
    printf("x = %s, y = %s\n", x, y);  
    return 0;  
}
```



Swap

```
void swap_string(char **a, char **b) {  
    char *temp = *a;  
    *a = *b;  
    *b = temp;  
}
```

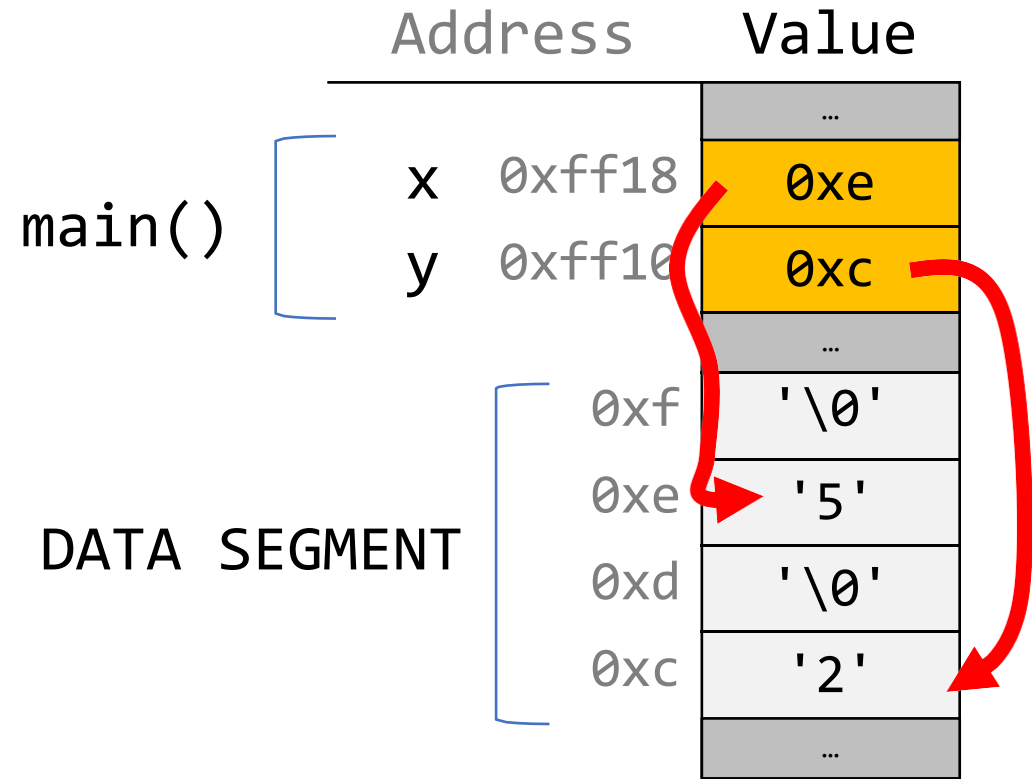
```
int main(int argc, char *argv[]) {  
    char *x = "2";  
    char *y = "5";  
    swap_string(&x, &y);  
    // want x = 5, y = 2  
    printf("x = %s, y = %s\n", x, y);  
    return 0;  
}
```



Swap

```
void swap_string(char **a, char **b) {  
    char *temp = *a;  
    *a = *b;  
    *b = temp;  
}
```

```
int main(int argc, char *argv[]) {  
    char *x = "2";  
    char *y = "5";  
    swap_string(&x, &y);  
    // want x = 5, y = 2  
    printf("x = %s, y = %s\n", x, y);  
    return 0;  
}
```



“Awesome! Thanks.”



“Awesome! Thanks. We also have 20 custom struct types. Could you write swap for those too?”



Generic Swap

What if we could write *one* function to swap two values of any single type?

```
void swap_int(int *a, int *b) { ... }  
void swap_float(float *a, float *b) { ... }  
void swap_size_t(size_t *a, size_t *b) { ... }  
void swap_double(double *a, double *b) { ... }  
void swap_string(char **a, char **b) { ... }  
void swap_mystruct(mystruct *a, mystruct *b) { ... }  
...
```

Generic Swap

```
void swap_int(int *a, int *b) {  
    int temp = *a;  
    *a = *b;  
    *b = temp;  
}
```

```
void swap_short(short *a, short *b) {  
    short temp = *a;  
    *a = *b;  
    *b = temp;  
}
```

```
void swap_string(char **a, char **b) {  
    char *temp = *a;  
    *a = *b;  
    *b = temp;  
}
```

Generic Swap

```
void swap_int(int *a, int *b) {  
    int temp = *a;  
    *a = *b;  
    *b = temp;  
}  
  
void swap_short(short *a, short *b)  
{  
    short temp = *a;  
    *a = *b;  
    *b = temp;  
}  
  
void swap_string(char **a, char **b) {  
    char *temp = *a;  
    *a = *b;  
    *b = temp;  
}
```

All 3:

- Take pointers to values to swap
- Create temporary storage to store one of the values
- Move data at **b** into where **a** points
- Move data in temporary storage into where **b** points

Generic Swap

```
void swap(pointer to data1, pointer to data2) {  
    store a copy of data1 in temporary storage  
    copy data2 to location of data1  
    copy data in temporary storage to location of data2  
}
```

Generic Swap

```
void swap(pointer to data1, pointer to data2) {  
    store a copy of data1 in temporary storage  
    copy data2 to location of data1  
    copy data in temporary storage to location of data2  
}
```

```
int temp = *data1ptr;
```

4 bytes

```
short temp = *data1ptr;
```

2 bytes

```
char *temp = *data1ptr;
```

8 bytes

Problem: each type may need a different size temp!

Generic Swap

```
void swap(pointer to data1, pointer to data2) {  
    store a copy of data1 in temporary storage  
    copy data2 to location of data1  
    copy data in temporary storage to location of data2  
}
```

`*data1Ptr = *data2ptr;`

4 bytes

`*data1Ptr = *data2ptr;`

2 bytes

`*data1Ptr = *data2ptr;`

8 bytes

Problem: each type needs to copy a different amount of data!

Generic Swap

```
void swap(pointer to data1, pointer to data2) {  
    store a copy of data1 in temporary storage  
    copy data2 to location of data1  
    copy data in temporary storage to location of data2  
}
```

`*data2ptr = temp;`

4 bytes

`*data2ptr = temp;`

2 bytes

`*data2ptr = temp;`

8 bytes

Problem: each type needs to copy a different amount of data!

**C knows the size of temp,
and knows how many bytes
to copy, because of the
variable types.**

Is there a way to make a version that doesn't care about the variable types?

Generic Swap

```
void swap(pointer to data1, pointer to data2){  
    store a copy of data1 in temporary storage  
    copy data2 to location of data1  
    copy data in temporary storage to location of data2  
}
```

Generic Swap

```
void swap(void *data1ptr, void *data2ptr) {  
    store a copy of data1 in temporary storage  
    copy data2 to location of data1  
    copy data in temporary storage to location of data2  
}
```

Generic Swap

```
void swap(void *data1ptr, void *data2ptr) {  
    // store a copy of data1 in temporary storage  
    // copy data2 to location of data1  
    // copy data in temporary storage to location of data2  
}
```

Generic Swap

```
void swap(void *data1ptr, void *data2ptr) {  
    // store a copy of data1 in temporary storage  
    // copy data2 to location of data1  
    // copy data in temporary storage to location of data2  
}
```

If we don't know the data type, we don't know how many bytes it is. Let's take that as another parameter.

Generic Swap

```
void swap(void *data1ptr, void *data2ptr, size_t nbytes) {  
    // store a copy of data1 in temporary storage  
    // copy data2 to location of data1  
    // copy data in temporary storage to location of data2  
}
```

If we don't know the data type, we don't know how many bytes it is. Let's take that as another parameter.

Generic Swap

```
void swap(void *data1ptr, void *data2ptr, size_t nbytes) {  
    // store a copy of data1 in temporary storage  
    // copy data2 to location of data1  
    // copy data in temporary storage to location of data2  
}
```

Let's start by making space to store the temporary value. How can we make **nbytes** of temp space?

Generic Swap

```
void swap(void *data1ptr, void *data2ptr, size_t nbytes) {  
    void temp; ???  
    // store a copy of data1 in temporary storage  
    // copy data2 to location of data1  
    // copy data in temporary storage to location of data2  
}
```

Let's start by making space to store the temporary value. How can we make **nbytes** of temp space?

Generic Swap

```
void swap(void *data1ptr, void *data2ptr, size_t nbytes){  
    char temp[nbytes];  
    // store a copy of data1 in temporary storage  
    // copy data2 to location of data1  
    // copy data in temporary storage to location of data2  
}
```

temp is **nbytes** of memory,
since each **char** is 1 byte!

Generic Swap

```
void swap(void *data1ptr, void *data2ptr, size_t nbytes){  
    char temp[nbytes];  
    // store a copy of data1 in temporary storage  
    // copy data2 to location of data1  
    // copy data in temporary storage to location of data2  
}
```

Now, how can we copy in what **data1ptr** points to into **temp**?

Generic Swap

```
void swap(void *data1ptr, void *data2ptr, size_t nbytes){  
    char temp[nbytes];  
    // store a copy of data1 in temporary storage  
    temp = *data1ptr; ???  
    // copy data2 to location of data1  
    // copy data in temporary storage to location of data2  
}
```

Now, how can we copy in what **data1ptr** points to into **temp**?

Generic Swap

```
void swap(void *data1ptr, void *data2ptr, size_t nbytes){  
    char temp[nbytes];  
    // store a copy of data1 in temporary storage  
    temp = *data1ptr; ???  
    // copy data2 to location of data1  
    // copy data in temporary storage to location of data2  
}
```

We can't dereference a **void *** (or set an array equal to something). C doesn't know what it points to! Therefore, it doesn't know how many bytes there it should be looking at.

memcpy

memcpy is a function that copies a specified amount of bytes at one address to another address.

```
void *memcpy(void *dest, const void *src, size_t n);
```

It copies the next *n* bytes that *src* points to to the location contained in *dest*. (It also returns **dest**). It does not support regions of memory that overlap.

```
int x = 5;  
int y = 4;  
memcpy(&x, &y, sizeof(x)); // like x = y
```

memcpy must take **pointers** to the bytes to work with to know where they live and where they should be copied to.

memmove

memmove is the same as `memcpy`, but supports overlapping regions of memory. (Unlike its name implies, it still “copies”).

```
void *memmove(void *dest, const void *src, size_t n);
```

It copies the next `n` bytes that `src` points to to the location contained in `dest`. (It also returns **`dest`**).

memmove

When might memmove be useful?

1	2	3	4	5	6	7
---	---	---	---	---	---	---



4	5	6	7	5	6	7
---	---	---	---	---	---	---

4 Copies - Which To Choose?

strcpy (string copy):

- Used for copying null-terminated strings
- Copies from source to destination until it encounters a null terminator
- Does not check buffer sizes, which can lead to buffer overflows if not used carefully

strncpy (string numbered copy):

- Similar to strcpy, but requires specifying a maximum number of characters to copy
- If the source string is shorter than the specified length, it pads the destination with null characters
- Does not guarantee null-termination if the source string is longer than the specified length

memcpy (memory copy):

- Used for copying a specified number of bytes from one memory location to another
- Can be used for any type of data, not just strings
- Does not check for overlap between source and destination memory regions

memmove (memory move):

- Similar to memcpy, but designed to handle overlapping memory regions safely
- Copies data to a temporary buffer first if source and destination overlap
- Generally slower than memcpy but safer when dealing with overlapping memory

Generic Swap

```
void swap(void *data1ptr, void *data2ptr, size_t nbytes){  
    char temp[nbytes];  
    // store a copy of data1 in temporary storage  
    temp = *data1ptr; ???  
    // copy data2 to location of data1  
    // copy data in temporary storage to location of data2  
}
```

We can't dereference a **void ***. C doesn't know what it points to! Therefore, it doesn't know how many bytes there it should be looking at.

Generic Swap

```
void swap(void *data1ptr, void *data2ptr, size_t nbytes){  
    char temp[nbytes];  
    // store a copy of data1 in temporary storage  
    temp = *data1ptr; ???  
    // copy data2 to location of data1  
    // copy data in temporary storage to location of data2  
}
```

How can **memcpy** or **memmove** help us here?

```
void *memcpy(void *dest, const void *src, size_t n);
```

```
void *memmove(void *dest, const void *src, size_t n);
```

Generic Swap

```
void swap(void *data1ptr, void *data2ptr, size_t nbytes){  
    char temp[nbytes];  
    // store a copy of data1 in temporary storage  
    memcpy(temp, data1ptr, nbytes);  
    // copy data2 to location of data1  
    // copy data in temporary storage to location of data2  
}
```

Generic Swap

```
void swap(void *data1ptr, void *data2ptr, size_t nbytes){  
    char temp[nbytes];  
    // store a copy of data1 in temporary storage  
    memcpy(temp, data1ptr, nbytes);  
    // copy data2 to location of data1  
    // copy data in temporary storage to location of data2  
}
```

We can copy the bytes ourselves into temp! This is equivalent to **temp = *data1ptr** in non-generic versions, but this works for *any* type of *any* size.

Generic Swap

```
void swap(void *data1ptr, void *data2ptr, size_t nbytes){  
    char temp[nbytes];  
    // store a copy of data1 in temporary storage  
    memcpy(temp, data1ptr, nbytes);  
    // copy data2 to location of data1  
    // copy data in temporary storage to location of data2  
}
```

How can we copy data2 to the location of data1?

Generic Swap

```
void swap(void *data1ptr, void *data2ptr, size_t nbytes){  
    char temp[nbytes];  
    // store a copy of data1 in temporary storage  
    memcpy(temp, data1ptr, nbytes);  
    // copy data2 to location of data1  
    *data1ptr = *data2ptr; ???  
    // copy data in temporary storage to location of data2  
}
```

How can we copy data2 to the location of data1?

Generic Swap

```
void swap(void *data1ptr, void *data2ptr, size_t nbytes){  
    char temp[nbytes];  
    // store a copy of data1 in temporary storage  
    memcpy(temp, data1ptr, nbytes);  
    // copy data2 to location of data1  
    memcpy(data1ptr, data2ptr, nbytes);  
    // copy data in temporary storage to location of data2  
}
```

How can we copy data2 to the location of data1?
memcpy!

Generic Swap

```
void swap(void *data1ptr, void *data2ptr, size_t nbytes){  
    char temp[nbytes];  
    // store a copy of data1 in temporary storage  
    memcpy(temp, data1ptr, nbytes);  
    // copy data2 to location of data1  
    memcpy(data1ptr, data2ptr, nbytes);  
    // copy data in temporary storage to location of data2  
}
```

How can we copy temp's data to the location of data2?

Generic Swap

```
void swap(void *data1ptr, void *data2ptr, size_t nbytes){  
    char temp[nbytes];  
    // store a copy of data1 in temporary storage  
    memcpy(temp, data1ptr, nbytes);  
    // copy data2 to location of data1  
    memcpy(data1ptr, data2ptr, nbytes);  
    // copy data in temporary storage to location of data2  
    memcpy(data2ptr, temp, nbytes);  
}
```

How can we copy temp's data to the location of data2? **memcpy!**

Generic Swap

```
void swap(void *data1ptr, void *data2ptr, size_t nbytes){  
    char temp[nbytes];  
    // store a copy of data1 in temporary storage  
    memcpy(temp, data1ptr, nbytes);  
    // copy data2 to location of data1  
    memcpy(data1ptr, data2ptr, nbytes);  
    // copy data in temporary storage to location of data2  
    memcpy(data2ptr, temp, nbytes);  
}
```

```
int x = 2;  
int y = 5;  
swap(&x, &y, sizeof(x));
```

Generic Swap

```
void swap(void *data1ptr, void *data2ptr, size_t nbytes){  
    char temp[nbytes];  
    // store a copy of data1 in temporary storage  
    memcpy(temp, data1ptr, nbytes);  
    // copy data2 to location of data1  
    memcpy(data1ptr, data2ptr, nbytes);  
    // copy data in temporary storage to location of data2  
    memcpy(data2ptr, temp, nbytes);  
}
```

```
short x = 2;  
short y = 5;  
swap(&x, &y, sizeof(x));
```

Generic Swap

```
void swap(void *data1ptr, void *data2ptr, size_t nbytes){  
    char temp[nbytes];  
    // store a copy of data1 in temporary storage  
    memcpy(temp, data1ptr, nbytes);  
    // copy data2 to location of data1  
    memcpy(data1ptr, data2ptr, nbytes);  
    // copy data in temporary storage to location of data2  
    memcpy(data2ptr, temp, nbytes);  
}
```

```
char *x = "2";  
char *y = "5";  
swap(&x, &y, sizeof(x));
```

Generic Swap

```
void swap(void *data1ptr, void *data2ptr, size_t nbytes){  
    char temp[nbytes];  
    // store a copy of data1 in temporary storage  
    memcpy(temp, data1ptr, nbytes);  
    // copy data2 to location of data1  
    memcpy(data1ptr, data2ptr, nbytes);  
    // copy data in temporary storage to location of data2  
    memcpy(data2ptr, temp, nbytes);  
}
```

```
mystruct x = {...};  
mystruct y = {...};  
swap(&x, &y, sizeof(x));
```

void *, memcpy, memmove

- We can use **void *** and **memcpy** to handle memory as generic bytes.
- If we are given where the data of importance is, and how big it is, we can handle it!

```
// even more robust (handles overlapping swap pointers)
void swap(void *data1ptr, void *data2ptr, size_t nbytes) {
    char temp[nbytes];
    memcpy(temp, data1ptr, nbytes);
    memmove(data1ptr, data2ptr, nbytes);
    memcpy(data2ptr, temp, nbytes);
}
```

Generic Lecture Plan

- Generic Swap
- **Generics Pitfalls**
- Generic Swap Ends
- Generic Stack
- Array Rotation

Void * Pitfalls

- **void** *s are powerful, but dangerous - C cannot do as much checking!
- E.g. with **int**, C would never let you swap *half* of an int. With **void** *s, this can happen! (*How? Let's find out!*)

Void *Pitfalls

- Void * has more room for error because it manipulates arbitrary bytes without knowing what they represent. This can result in some strange memory Frankensteins!



Generic Lecture Plan

- Generic Swap
- Generics Pitfalls
- **Generic Swap Ends**
- Generic Stack
- Array Rotation

Swap Ends

You're asked to write a function that swaps the first and last elements in an array of numbers.

```
void swap_ends_int(int *arr, size_t nelems) {  
    int tmp = arr[0];  
    arr[0] = arr[nelems - 1];  
    arr[nelems - 1] = tmp;  
}
```

Wait – we just wrote a generic swap function. Let's use that!

```
int main(int argc, char *argv[]) {  
    int nums[] = {5, 2, 3, 4, 1};  
    size_t nelems = sizeof(nums) / sizeof(nums[0]);  
    swap_ends_int(nums, nelems);  
    // want nums[0] = 1, nums[4] = 5  
    printf("nums[0] = %d, nums[4] = %d\n", nums[0], nums[4]);  
    return 0;  
}
```

Swap Ends

You're asked to write a function that swaps the first and last elements in an array of numbers.

```
void swap_ends_int(int *arr, size_t nelems) {  
    swap(arr, arr + nelems - 1, sizeof(*arr));  
}  
  
int main(int argc, char *argv[]) {  
    int nums[] = {5, 2, 3, 4, 1};  
    size_t nelems = sizeof(nums) / sizeof(nums[0]);  
    swap_ends_int(nums, nelems);  
    // want nums[0] = 1, nums[4] = 5  
    printf("nums[0] = %d, nums[4] = %d\n", nums[0], nums[4]);  
    return 0;  
}
```

Wait – we just wrote a generic swap function. Let's use that!

Swap Ends

Let's write out what some other versions would look like (just in case).

```
void swap_ends_int(int *arr, size_t nelems) {  
    swap(arr, arr + nelems - 1, sizeof(*arr));  
}
```

```
void swap_ends_short(short *arr, size_t nelems) {  
    swap(arr, arr + nelems - 1, sizeof(*arr));  
}
```

```
void swap_ends_string(char **arr, size_t nelems) {  
    swap(arr, arr + nelems - 1, sizeof(*arr));  
}
```

```
void swap_ends_float(float *arr, size_t nelems) {  
    swap(arr, arr + nelems - 1, sizeof(*arr));  
}
```

The code seems to be the same regardless of the type!

Swap Ends

Let's write a version of swap_ends that works for any type of array.

```
void swap_ends(void *arr, size_t nelems) {  
    swap(arr, arr + nelems - 1, sizeof(*arr));  
}
```

Is this generic? Does this work?

Swap Ends

Let's write a version of swap_ends that works for any type of array.

```
void swap_ends(void *arr, size_t nelems) {  
    swap(arr, arr + nelems - 1, sizeof(*arr));  
}
```

Is this generic? Does this work?

Unfortunately not. First, we *no longer know the element size*. Second, *pointer arithmetic depends on the type of data* being pointed to. With a void *, we lose that information!

Swap Ends

Let's write a version of swap_ends that works for any type of array.

```
void swap_ends(void *arr, size_t nelems) {  
    swap(arr, arr + nelems - 1, sizeof(*arr));  
}
```

We need to know the element size, so let's add a parameter.

Swap Ends

Let's write a version of `swap_ends` that works for any type of array.

```
void swap_ends(void *arr, size_t nelems, size_t elem_bytes) {  
    swap(arr, arr + nelems - 1, elem_bytes);  
}
```

We need to know the element size, so let's add a parameter.

Pointer Arithmetic

`arr + nelems - 1`

Let's say `nelems = 4`. How many bytes into/beyond `arr` is this?

If it's an array of...

Int?

Pointer Arithmetic

`arr + nelems - 1`

Let's say `nelems = 4`. How many bytes into/beyond `arr` is this?

If it's an array of...

int: adds 3 places to `arr`, and `3 * sizeof(int) = 12 bytes`

Pointer Arithmetic

`arr + nelems - 1`

Let's say `nelems = 4`. How many bytes into/beyond `arr` is this?

If it's an array of...

Int: adds 3 places to `arr`, and $3 * \text{sizeof}(\text{int}) = 12$ bytes

Short?

Pointer Arithmetic

`arr + nelems - 1`

Let's say `nelems = 4`. How many bytes into/beyond `arr` is this?

If it's an array of...

Int: adds 3 places to `arr`, and $3 * \text{sizeof}(\text{int}) = 12$ bytes

Short: adds 3 places to `arr`, and $3 * \text{sizeof}(\text{short}) = 6$ bytes

Pointer Arithmetic

`arr + nelems - 1`

Let's say `nelems = 4`. How many bytes into/beyond `arr` is this?

If it's an array of...

Int: adds 3 places to `arr`, and `3 * sizeof(int) = 12` bytes

Short: adds 3 places to `arr`, and `3 * sizeof(short) = 6` bytes

Char *: adds 3 places to `arr`, and `3 * sizeof(char *) = 24` bytes

In each case, we need to know the element size to do the arithmetic.

Swap Ends

Let's write a version of `swap_ends` that works for any type of array.

```
void swap_ends(void *arr, size_t nelems, size_t elem_bytes) {  
    swap(arr, arr + nelems - 1, elem_bytes);  
}
```

How many bytes past `arr` should we go to get to the last element?

`(nelems - 1) * elem_bytes`

Swap Ends

Let's write a version of swap_ends that works for any type of array.

```
void swap_ends(void *arr, size_t nelems, size_t elem_bytes) {  
    swap(arr, arr + (nelems - 1) * elem_bytes, elem_bytes);  
}
```

How many bytes past arr should we go to get to the last element?

$(\text{nelems} - 1) * \text{elem_bytes}$

Swap Ends

Let's write a version of swap_ends that works for any type of array.

```
void swap_ends(void *arr, size_t nelems, size_t elem_bytes) {  
    swap(arr, arr + (nelems - 1) * elem_bytes, elem_bytes);  
}
```

But C still can't do arithmetic with a void*. We need to tell it to not worry about it, and just add bytes. **How can we do this?**

Swap Ends

Let's write a version of swap_ends that works for any type of array.

```
void swap_ends(void *arr, size_t nelems, size_t elem_bytes) {  
    swap(arr, (char *)arr + (nelems - 1) * elem_bytes, elem_bytes);  
}
```

But C still can't do arithmetic with a void*. We need to tell it to not worry about it, and just add bytes. **How can we do this?**

char * pointers already add bytes!

Swap Ends

Well, now it can swap_ends for an array of anything!

```
void swap_ends(void *arr, size_t nelems, size_t elem_bytes) {  
    swap(arr, (char *)arr + (nelems - 1) * elem_bytes, elem_bytes);  
}
```

We can do pointer arithmetic with a **void *** pointer by casting it.

Swap Ends

Lets see some examples!

```
void swap_ends(void *arr, size_t nelems, size_t elem_bytes) {  
    swap(arr, (char *)arr + (nelems - 1) * elem_bytes, elem_bytes);  
}
```

```
int nums[] = {5, 2, 3, 4, 1};  
size_t nelems = sizeof(nums) / sizeof(nums[0]);  
swap_ends(nums, nelems, sizeof(nums[0]));
```

Swap Ends

Lets see some examples!

```
void swap_ends(void *arr, size_t nelems, size_t elem_bytes) {  
    swap(arr, (char *)arr + (nelems - 1) * elem_bytes, elem_bytes);  
}
```

```
short nums[] = {5, 2, 3, 4, 1};  
size_t nelems = sizeof(nums) / sizeof(nums[0]);  
swap_ends(nums, nelems, sizeof(nums[0]));
```

Swap Ends

Lets see some examples!

```
void swap_ends(void *arr, size_t nelems, size_t elem_bytes) {  
    swap(arr, (char *)arr + (nelems - 1) * elem_bytes, elem_bytes);  
}
```

```
char *strs[] = {"Hi", "Hello", "Howdy"};  
size_t nelems = sizeof(strs) / sizeof(strs[0]);  
swap_ends(strs, nelems, sizeof(strs[0]));
```

Swap Ends

Lets see some examples!

```
void swap_ends(void *arr, size_t nelems, size_t elem_bytes) {  
    swap(arr, (char *)arr + (nelems - 1) * elem_bytes, elem_bytes);  
}
```

```
mystruct structs[] = ...;  
size_t nelems = ...;  
swap_ends(structs, nelems, sizeof(structs[0]));
```

Generic Lecture Plan

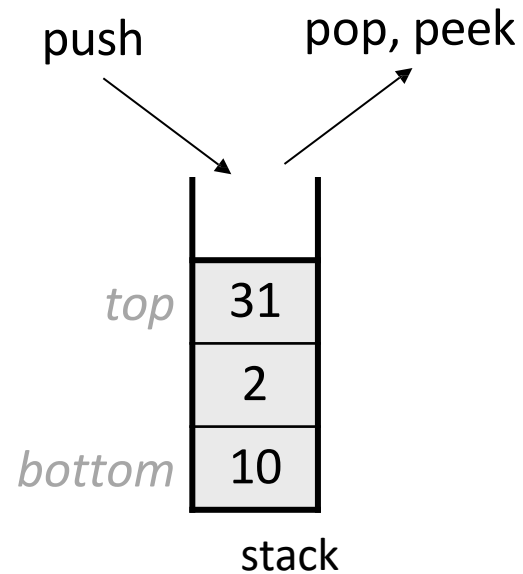
- Generic Swap
- Generics Pitfalls
- Generic Swap Ends
- **Generic Stack**
- Array Rotation

Stacks

- C generics are particularly powerful in helping us create generic data structures.
- Let's see how we might go about making a Stack in C.

Refresher: Stacks

- A **Stack** is a data structure representing a stack of things.
- Objects can be ***pushed*** on top of or ***popped*** from the top of the stack.
- Only the top of the stack can be accessed; no other objects in the stack are visible.
- Main operations:
 - **push(value)**: add an element to the top of the stack
 - **pop()**: remove and return the top element in the stack
 - **peek()**: return (but do not remove) the top element in the stack

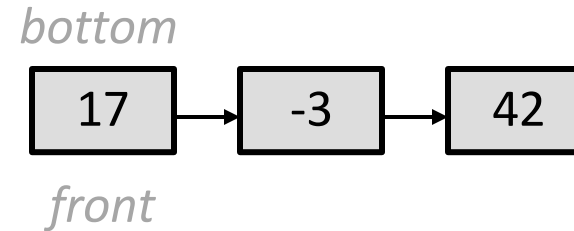


Refresher: Stacks

A stack is often implemented using a **linked list** internally.

- "bottom" = tail of linked list
- "top" = head of linked list

```
Stack<int> s;  
s.push(42);  
s.push(-3);  
s.push(17);
```



Problem: C is not object-oriented! We can't call methods on variables.

**What modifications are
necessary to make a
generic stack?**

Stack Structs

```
typedef struct int_node {  
    struct int_node *next;  
    int data;  
} int_node;
```

```
typedef struct int_stack {  
    int nelems;  
    int_node *top;  
} int_stack;
```

How might we modify the Stack data representation itself to be generic?

Stack Structs

```
typedef struct int_node {  
    struct int_node *next;  
    int data;  
} int_node;
```

```
typedef struct int_stack {  
    int nelems;  
    int_node *top;  
} int_stack;
```

Problem: each node can no longer store the data itself, because it could be any size!

Generic Stack Structs

```
typedef struct node {  
    struct node *next;  
    void *data;  
} node;
```

```
typedef struct stack {  
    int nelems;  
    node *top;  
} stack;
```

Solution: each node stores a pointer, which is always 8 bytes, to the data somewhere else. We must also store the data size, where should that go?

Generic Stack Structs

```
typedef struct node {  
    struct node *next;  
    void *data;  
} node;
```

```
typedef struct stack {  
    int nelems;  
    int elem_size_bytes;  
    node *top;  
} stack;
```

Solution: each node stores a pointer, which is always 8 bytes, to the data somewhere else. We must also store the data size in the Stack struct.

Stack Functions

- **`int_stack_create()`**: creates a new stack on the heap and returns a pointer to it
- **`int_stack_push(int_stack *s, int data)`**: pushes data onto the stack
- **`int_stack_pop(int_stack *s)`**: pops and returns topmost stack element

int_stack_create

```
int_stack *int_stack_create() {  
    int_stack *s = malloc(sizeof(int_stack));  
    s->nelems = 0;  
    s->top = NULL;  
    return s;  
}
```

How might we modify this function to be generic?

From previous slide:

```
typedef struct stack {  
    int nelems;  
    int elem_size_bytes;  
    node *top;  
} stack;
```

Generic stack_create

```
stack *stack_create(int elem_size_bytes) {  
    stack *s = malloc(sizeof(stack));  
    s->nelems = 0;  
    s->top = NULL;  
    s->elem_size_bytes = elem_size_bytes;  
    return s;  
}
```

From previous slide:

```
typedef struct stack {  
    int nelems;  
    int elem_size_bytes;  
    node *top;  
} stack;
```

Generic stack_push

```
int main() {  
    stack *int_stack = stack_create(sizeof(int));  
    add_one(int_stack);  
    // now stack stores pointer to invalid memory for 7!  
}  
  
void add_one(stack *s) {  
    int num = 7;  
    stack_push(s, &num);  
}
```

int_stack_push

```
void int_stack_push(int_stack *s, int data) {  
    int_node *new_node = malloc(sizeof(int_node));  
    new_node->data = data;  
  
    new_node->next = s->top;  
    s->top = new_node;  
    s->nelems++;  
}
```

How might we modify this function to be generic?

From previous slide:

```
typedef struct stack {  
    int nelems;  
    int elem_size_bytes;  
    node *top;  
} stack;
```

```
typedef struct node {  
    struct node *next;  
    void *data;  
} node;
```

Generic stack_push

```
void int_stack_push(int_stack *s, int data) {  
    int_node *new_node = malloc(sizeof(int_node));  
    new_node->data = data;  
  
    new_node->next = s->top;  
    s->top = new_node;  
    s->nelems++;  
}
```

Problem 1: we can no longer pass the data itself as a parameter, because it could be any size!

Generic stack_push

```
void int_stack_push(int_stack *s, const void *data) {  
    int_node *new_node = malloc(sizeof(int_node));  
    new_node->data = data;  
  
    new_node->next = s->top;  
    s->top = new_node;  
    s->nelems++;  
}
```

Solution 1: pass a pointer to the data as a parameter instead.

Generic stack_push

```
void int_stack_push(int_stack *s, const void *data) {  
    int_node *new_node = malloc(sizeof(int_node));  
    new_node->data = data;  
  
    new_node->next = s->top;  
    s->top = new_node;  
    s->nelems++;  
}
```

Problem 2: we cannot copy the existing data pointer into new_node. The data structure must manage its own copy that exists for its entire lifetime. The provided copy may go away!

Generic stack_push

```
void stack_push(stack *s, const void *data) {  
    node *new_node = malloc(sizeof(node));  
    new_node->data = malloc(s->elem_size_bytes);  
    memcpy(new_node->data, data, s->elem_size_bytes);  
  
    new_node->next = s->top;  
    s->top = new_node;  
    s->nelems++;  
}
```

Solution 2: make a heap-allocated copy of the data that the node points to.

int_stack_pop

```
int int_stack_pop(int_stack *s) {  
    if (s->nelems == 0) {  
        error(1, 0, "Cannot pop from empty stack");  
    }  
    int_node *n = s->top;  
    int value = n->data;  
  
    s->top = n->next;  
  
    free(n);  
    s->nelems--;  
  
    return value;  
}
```

How might we modify this function to be generic?

From previous slide:

```
typedef struct stack {  
    int nelems;  
    int elem_size_bytes;  
    node *top;  
} stack;
```

```
typedef struct node {  
    struct node *next;  
    void *data;  
} node;
```

Generic stack_pop

```
int int_stack_pop(int_stack *s) {  
    if (s->nelems == 0) {  
        error(1, 0, "Cannot pop from empty stack");  
    }  
    int_node *n = s->top;  
    int value = n->data;  
  
    s->top = n->next;  
  
    free(n);  
    s->nelems--;  
  
    return value;  
}
```

Problem: we can no longer return the data itself, because it could be any size!

Generic stack_pop

```
void *int_stack_pop(int_stack *s) {  
    if (s->nelems == 0) {  
        error(1, 0, "Cannot pop from empty stack");  
    }  
    int_node *n = s->top;  
    void *value = n->data;  
  
    s->top = n->next;  
  
    free(n);  
    s->nelems--;  
  
    return value;  
}
```

While it's possible to return the heap address of the element, this means the client would be responsible for freeing it. Ideally, the data structure should manage its own memory here.

Generic stack_pop

```
void stack_pop(stack *s, void *addr) {  
    if (s->nelems == 0) {  
        error(1, 0, "Cannot pop from empty stack");  
    }  
    node *n = s->top;  
    memcpy(addr, n->data, s->elem_size_bytes);  
    s->top = n->next;  
  
    free(n->data);  
    free(n);  
    s->nelems--;  
}
```

Solution: have the caller pass a memory location as a parameter and copy the data to that location.

Using Generic Stack

```
int_stack *intstack = int_stack_create();  
for (int i = 0; i < TEST_STACK_SIZE; i++) {  
    int_stack_push(intstack, i);  
}
```

We must now pass the *address* of an element to push onto the stack, rather than the element itself.

Using Generic Stack

```
stack *intstack = stack_create(sizeof(int));  
for (int i = 0; i < TEST_STACK_SIZE; i++) {  
    stack_push(intstack, &i);  
}
```

We must now pass the *address* of an element to push onto the stack, rather than the element itself.

Using Generic Stack

```
int_stack *intstack = int_stack_create();  
int_stack_push(intstack, 7);
```

We must now pass the *address* of an element to push onto the stack, rather than the element itself.

Using Generic Stack

```
stack *intstack = stack_create(sizeof(int));  
int num = 7;  
stack_push(intstack, &num);
```

We must now pass the *address* of an element to push onto the stack, rather than the element itself.

Using Generic Stack

```
// Pop off all elements
while (intstack->nelems > 0) {
    printf("%d\n", int_stack_pop(intstack));
}
```

We must now pass the *address* of where we would like to store the popped element, rather than getting it directly as a return value.

Using Generic Stack

```
// Pop off all elements
int popped_int;
while (intstack->nelems > 0) {
    stack_pop(intstack, &popped_int);
    printf("%d\n", popped_int);
}
```

We must now pass the *address* of where we would like to store the popped element, rather than getting it directly as a return value.

Recap

- **void *** is a variable type that represents a generic pointer “to something”.
- We cannot perform pointer arithmetic with or dereference a **void ***.
- We can use **memcpy** or **memmove** to copy data from one memory location to another.
- To do pointer arithmetic with a **void ***, we must first cast it to a **char ***.
- **void *** and generics are powerful but dangerous because of the lack of type checking, so we must be extra careful when working with generic memory.

Generic Lecture Plan

- Generic Swap
- Generics Pitfalls
- Generic Swap Ends
- Generic Stack
- **Array Rotation**

Bonus Exercise: Array Rotation

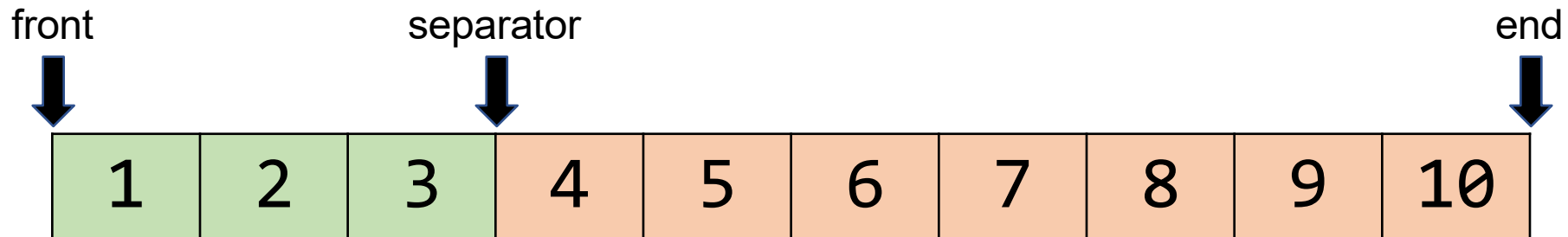
Exercise: You're asked to provide an implementation for a function called **rotate** with the following prototype:

```
void rotate(void *front, void *separator, void *end);
```

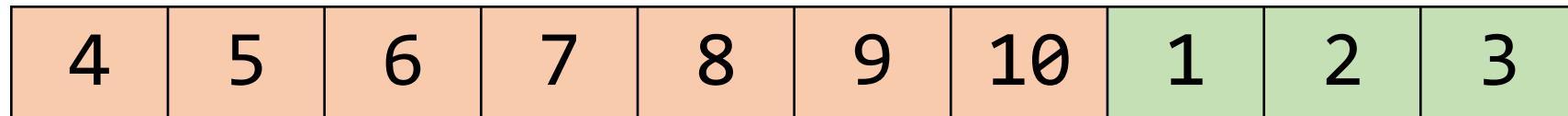
The expectation is that **front** is the base address of an array, **end** is the past-the-end address of the array, and **separator** is the address of some element in between. **rotate** moves all elements in between **front** and **separator** to the end of the array, and all elements between **separator** and **end** move to the front.

Bonus Exercise: Array Rotation

```
int array[10] = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10};  
rotate(array, array + 3, array + 10);
```



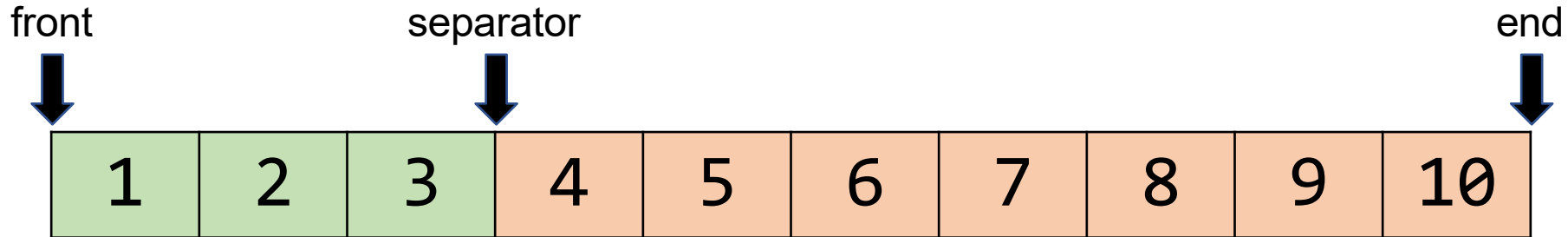
Before:



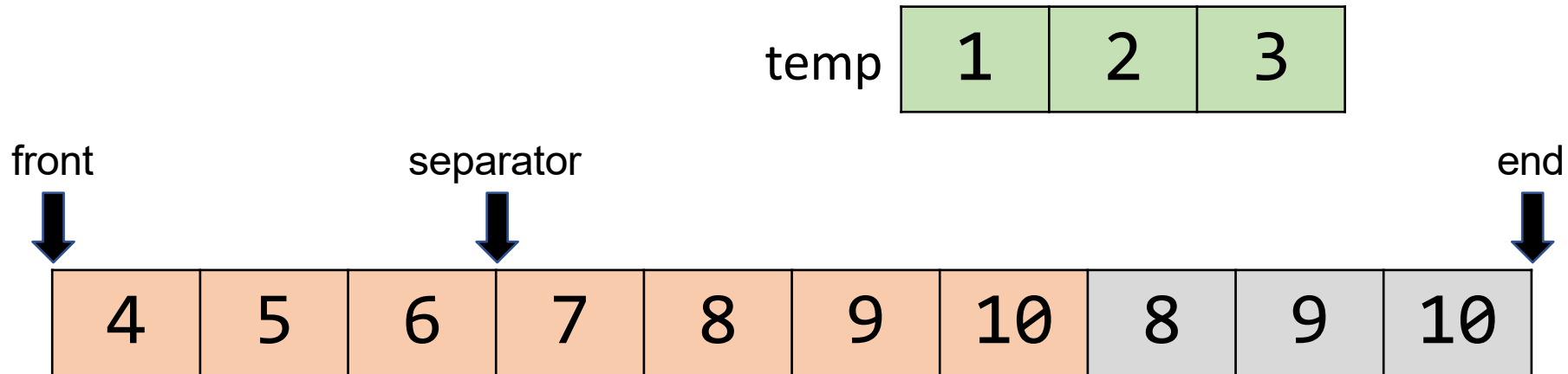
After:

The inner workings of rotate

Before
rotate:



Before
last step:



Bonus Exercise: Array Rotation

Exercise: Implement `rotate` to generate the provided output.

```
int main(int argc, char *argv[]) {
    int array[10] = {1, 2, 3, 4, 5, 6, 7, 8, 9, 10};
    print_int_array(array, 10); // intuit implementation 😊
    rotate(array, array + 5, array + 10);
    print_int_array(array, 10);
    rotate(array, array + 1, array + 10);
    print_int_array(array, 10);
    rotate(array + 4, array + 5, array + 6);
    print_int_array(array, 10);
    return 0;
}
```

Output:

```
myth52:~/lect8$ ./rotate
Array: 1, 2, 3, 4, 5, 6, 7, 8, 9, 10
Array: 6, 7, 8, 9, 10, 1, 2, 3, 4, 5
Array: 7, 8, 9, 10, 1, 2, 3, 4, 5, 6
Array: 7, 8, 9, 10, 2, 1, 3, 4, 5, 6
myth52:~/lect8$
```

Bonus Exercise: Array Rotation

Exercise: A properly implemented **rotate** will prompt the following program to generate the provided output.

And here's that properly implemented function!

```
void rotate(void *front, void *separator, void *end) {
    int width = (char *)end - (char *)front;
    int prefix_width = (char *)separator - (char *)front;
    int suffix_width = width - prefix_width;

    char temp[prefix_width];
    memcpy(temp, front, prefix_width);
    memmove(front, separator, suffix_width);
    memcpy((char *)end - prefix_width, temp, prefix_width);
}
```