



# CS 3110

## Testing

Nate Foster

Spring 2020

Today's music: Wrecking Ball by Miley Cyrus

# **CLICKER QUESTION 1**

# Review

## Previously in 3110:

- Modules
- Specification (functions, modules)

## Today:

- Validation
- Testing
  - Black box
  - Glass box

# Validation

- **Validation:** does program behave as intended?
- **Testing:** a process for validation
- **Debugging:** determining cause of unintended behavior
- **Defensive programming:** implementation techniques for making validation and debugging easier

# Approaches to validation

- Social
  - Code reviews
  - Extreme/Pair programming
- Methodological
  - Test-driven development
  - Version control
  - Bug tracking
- Technological
  - Static analysis (“lint” tools, FindBugs, ...)
  - Fuzzers
- Mathematical
  - Type systems
  - Formal verification



Less formal: Techniques may miss problems in programs

All of these methods should be used!

Even the most formal can still have holes:

- did you prove the right thing?
- do your assumptions match reality?

More formal: eliminate *with certainty* as many problems as possible.

# Testing vs. Verification

## Testing:

- Cost effective
- Guarantee that program is correct on **tested** inputs and in **tested** environments

## Verification:

- Expensive
- Guarantee that program is correct on **all** inputs and in **all** environments

# Edsger W. Dijkstra



(1930-2002)

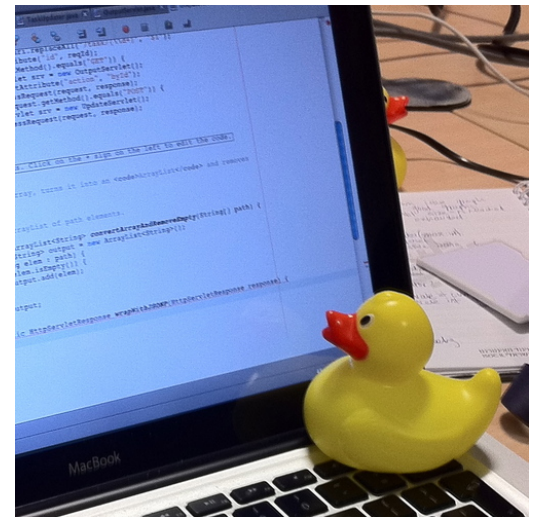
**Turing Award Winner (1972)**

*For eloquent insistence and practical demonstration that programs should be composed correctly, not just debugged into correctness*

"Program testing can at best show the presence of errors but never their absence."

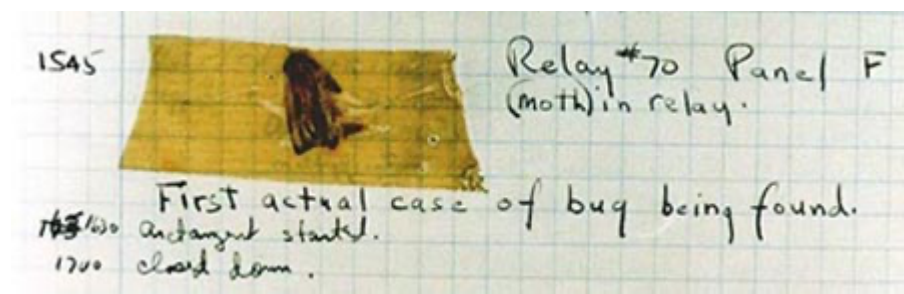
(more in recitation)

# DEBUGGING ADVICE





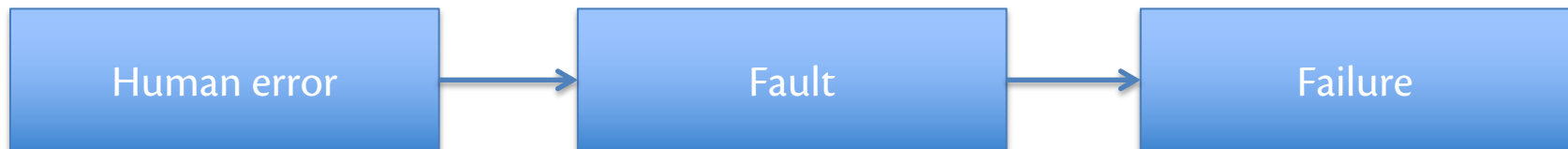
# Bugs



"bug": suggests something just wandered in

[IEEE 729]

- **Fault:** result of human error in software system
  - E.g., implementation doesn't match design, or design doesn't match requirements
  - Might never appear to end user
- **Failure:** violation of requirement
  - Something goes wrong for end user



# Testing

- Goal is to expose existence of faults, so that they can be fixed
- **Unit testing:** isolated components
- **Integration testing:** combined components
- **System testing:** functionality, performance, acceptance, installation

# Regression testing

- **Regression:** a previously fixed fault is reintroduced into the code
- **Regression testing:** running tests against new version of software to ensure no regressions
- If you ever find and fix a fault...
  - Put a test case into your suite for it
  - Run suite frequently to detect regressions

## **CLICKER QUESTION 2**

# Testing

When do you stop testing?

- **Bad answer:** when time is up
- **Bad answer:** what all tests pass

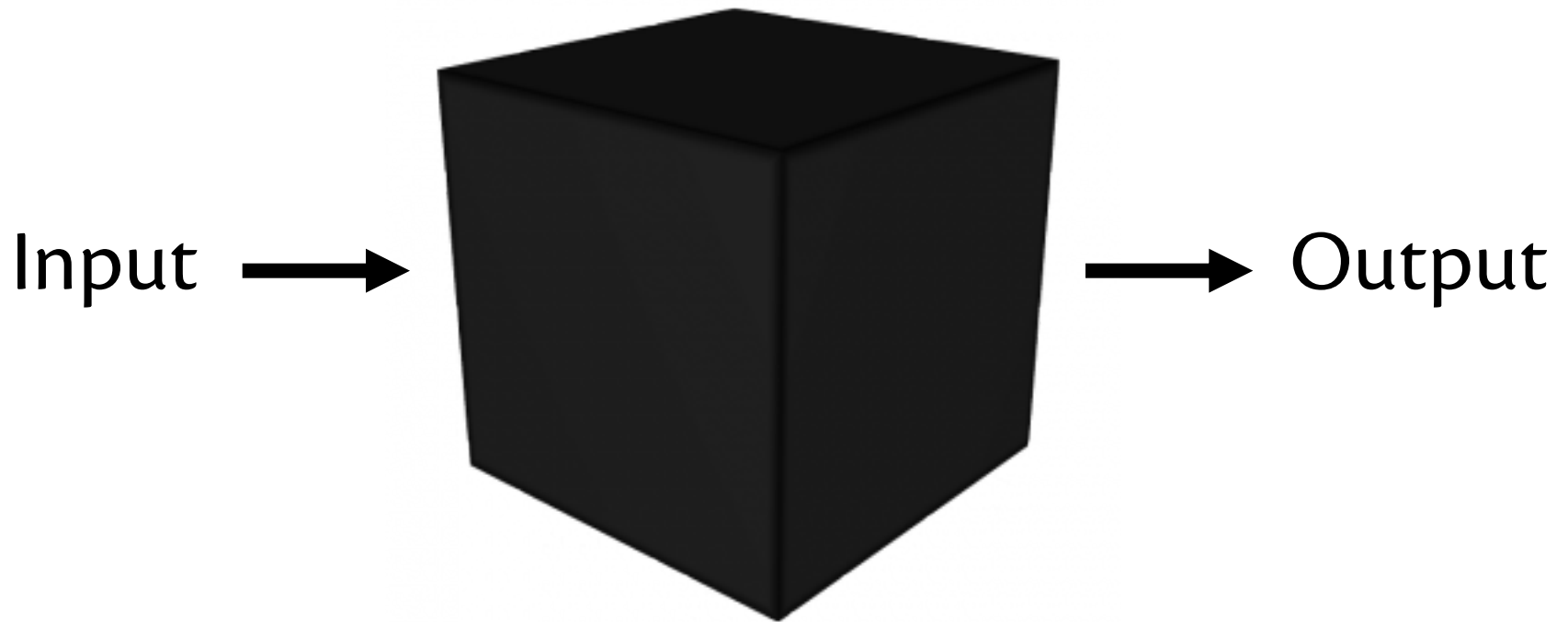
# Testing

When do you stop testing?

- **Good answer:** when testing methodology is complete
- **Future answer:** statistical estimation says *Pr[undetected faults]* is low enough (active research)

**TESTING**

# Black box testing



tester knows nothing about internals of functionality being tested

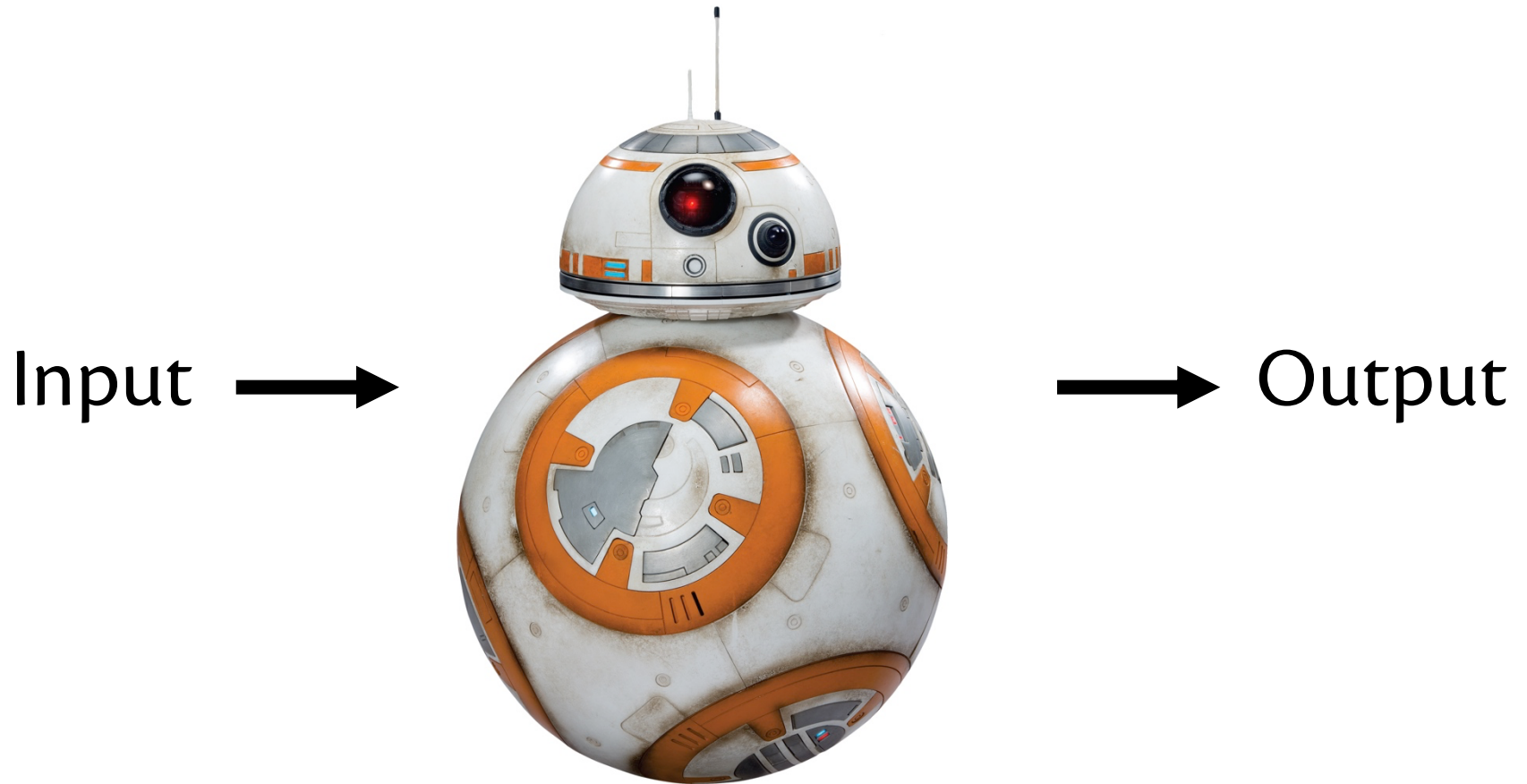


# Glass box testing



tester knows internals of functionality being tested

# Black box testing



tester knows nothing about internals of functionality being tested

# Glass box testing

Input



Output

tester knows internals of functionality being tested

# **BLACK BOX TESTING**

# Black box testing

- Tests are based on the **specification**
- **Advantages:**
  - Tester is not biased by assumptions made in implementation
  - Tests are robust w.r.t. changes in implementation
  - Tests can be read and evaluated by reviewers who are not implementers
- **Main kinds of black box tests:**
  - Example inputs provided by spec
  - Typical inputs
  - Boundary cases
  - Paths through spec

# Typical inputs

- Common, simple values of a type
  - **int**: small integers like 1 or 10
  - **char**: alphabetic letters, digits
  - **string**: whose length is a small integer and whose characters are typical
  - '**a list**': a small integer number of elements, each of which is a typical value of type '**a**'
  - **records/tuples**: each field/component with a typical value
  - **variants**: typical constructors, if there is such a thing

# Boundary cases



**Bill Sempf**

@sempf

Follow



QA Engineer walks into a bar. Orders a beer.  
Orders 0 beers. Orders 9999999999 beers.  
Orders a lizard. Orders -1 beers. Orders a  
sfdeljknesv.

10:56 AM - 23 Sep 2014

# Boundary cases

- aka *corner cases* or *edge cases*
- Atypical or extremal values of a type, and values nearby
  - **int**: 0, 1, -1, **min\_int**, **max\_int**
  - **char**: '**\000**', '**\255**', '**\032**' (space), '**\127**' (delete)
  - **string**: empty string, string with a single character, unreasonably long string
  - '**a list**': empty list, list with a single element, list with enough elements to trigger stack overflow on non-tail-recursive functions
  - **records/tuples**: combinations of atypical values
  - **variants**: all constructors



# Paths through spec

## Representative inputs for classes of outputs

```
(* [is_prime n] is true iff [n] is prime *)  
val is_prime: int -> bool
```

two classes of output:

- true: representative input: n=13
- false: representative input: n=42

other examples:

- **compare** functions have three classes of output
- functions that return variants have several classes of output

# Paths through spec

Representative inputs for each way of satisfying the precondition

```
(* [sqrt x n] is the square root of [x]  
  * computed to an accuracy of [n]  
  * significant digits  
  * requires: x >= 0 and n >= 1 *)  
val sqrt : float -> int -> float
```

(i) x=0.0, n=1, (ii) x=1.0, n=1,  
(iii) x=0.0, n=2, (iv) x=1.0, n=2

# Paths through spec

Representative inputs for each way of raising and not raising exception

```
(* [pos x lst] is the 0-based position of  
 * the first element of [lst] that equals [x].  
 * raises:  Not_found if [x] is not in [lst].  
 *)
```

```
val pos: 'a -> 'a list -> int
```

(i)  $x=1, lst=[1]$ , (ii)  $x=0, lst=[1]$

# **GLASS BOX TESTING**

# Glass box testing

- aka *white box testing*
- **Advantages:**
  - can determine whether a new test case really yields additional information about correctness of implementation
  - can address likely errors that are not apparent from specification
- **Supplements** black-box testing; does not **replace** examination of specification

# Glass box testing

- Goal is to **cover** entire program with test cases: ensure entire program exercised by tests
- Branches (if, match, Boolean ops, exceptions, loops, etc.) make it challenging
- Exact definition of **coverage** is flexible; could attempt to:
  - Evaluate every expression
  - Evaluate every Boolean/pattern match to each possible value
  - Cause every possible execution path through program to occur
  - Classically those are called **statement, condition, and path coverage**

# Coverage

```
let max3 x y z =  
  if x>y then  
    if x>z then x else z  
  else  
    if y>z then y else z
```

Testing according to black-box specification might lead to all kinds of inputs

But there are really only four paths through implementation!  
Representatives: (i) 3 2 1, (ii) 3, 2, 4, (iii) 1, 2, 1, (iv) 1, 2, 3

# Achieving good coverage

- Include test cases for:
  - each branch of each (nested) if expression
  - each branch of each (nested) pattern match
- Particularly watch out for:
  - base cases of recursive function
  - recursive calls in recursive function
  - every place where an exception might be raised



# Bisect

- OCaml tool for glass-box testing (statement and condition coverage)
- Tutorial in textbook
- You will use it on A4

# **RANDOMIZED (“FUZZ”) TESTING**

# Randomized testing

- *"It was a dark and stormy night..."*
- Generate **random inputs** and feed them to programs:
  - Crash? hang? terminate normally?
  - Of ~90 utilities in '89, crashed about 25-33% in various Unixes
  - Crash => buffer overflow potential
- Since then, "fuzzing" has become a standard practice for security testing
- Results have been repeated for X-windows system, Windows NT, Mac OS X
  - **Results keep getting worse in GUIs** but better on command line

# Upcoming events

- [last night] A2 due
- [tomorrow] February break!

*This is saving your code from being rekt.*

**THIS IS 3110**