
Why Tests Don't Pass

CAST July 14, 2009

Douglas Hoffman
Software Quality Methods, LLC.
Doug.Hoffman@acm.org
www.SoftwareQualityMethods.com

Definitions

- **Test**
 - An exercise designed to surface bugs
 - Includes a validation mechanism
- **Pass**
 - A test that behaves as expected per validation
- **Fail**
 - A test where validation fails
 - A test with unexpected behavior
- **Oracle**
 - Principle or mechanism by which we recognize a problem
 - Answers “why” when unexpected things happen

Deterministic Test Result Possibilities

Situation / Test Result	No Bug in SUT	Bug in SUT
As Expected (Pass)	Correct Pass	Silent Miss
Abnormal (Fail)	False Alarm	Caught it!

Pass?

- What results are we checking?
- How do we know the expected outcome?
- Are we checking all of the results?
- What outcomes are we not checking?
- How would we know misbehaviors for the outcomes we are not checking?

A Simple Challenge

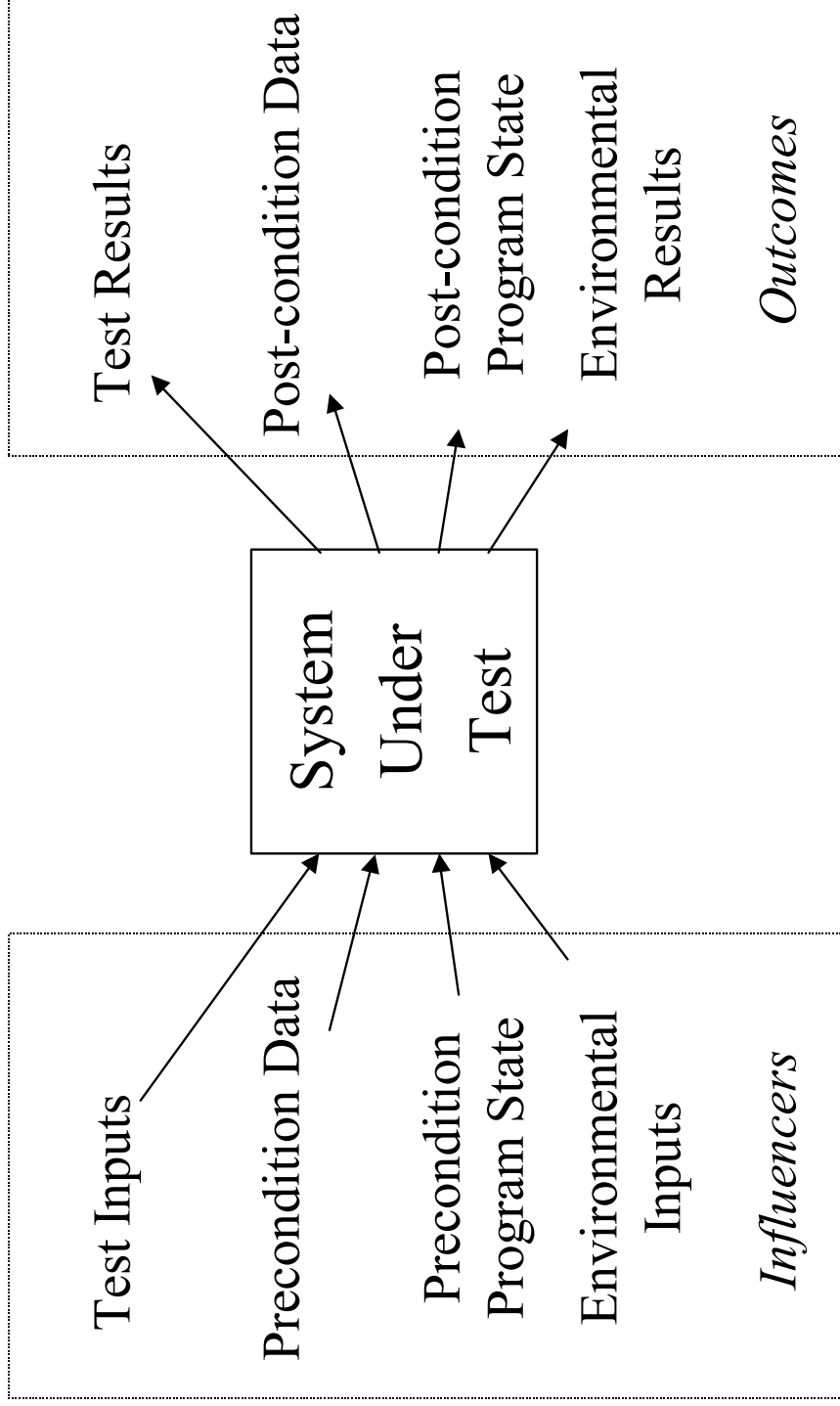
Given a program and a set of tests for it, modify the code [to do anything] in a way that the tests don't detect.

This change is an undetected bug.

Fail?

- The test finds the bug it's designed for!
- Something unexpected happens while the test is running.
- Possible false alarms.
- Possible unnoticed failures.

Software Test Execution Model



Copyright © 2000-2009 SQM, LLC.

Questions to Ask Ourselves

- Are we controlling or monitoring the most important influencers?
- Are we checking the most important outcomes?
- What values/conditions should influence the SUT?
How do we set/cause them?
- How do we know the expected outcomes?
- What influencers do we know we are not controlling or monitoring?
- What outcomes do we know we are not checking?
- What gives us confidence the test isn't missing bugs?

So What

- Be open minded about (skeptical of) “Pass”
- Understand that a “Fail” could mask errors
- We aren’t checking all of the results
- We don’t know the outcomes from arbitrary errors

Pass/Fail Metrics

- Count of passes overstates goodness
- Count of fails overstates the failures
- Passes will never be corrected
- Failures will only be correct after investigation
- Pass/Fail metrics don't really give us interesting information

Summary

- The SUT doesn't really pass or fail
 - “Pass” means we noticed only expected behaviors
 - “Fail” means we noticed something not according to plan
- There are lots of other possible reasons for “passing” or “failing” other than a bug in the SUT

