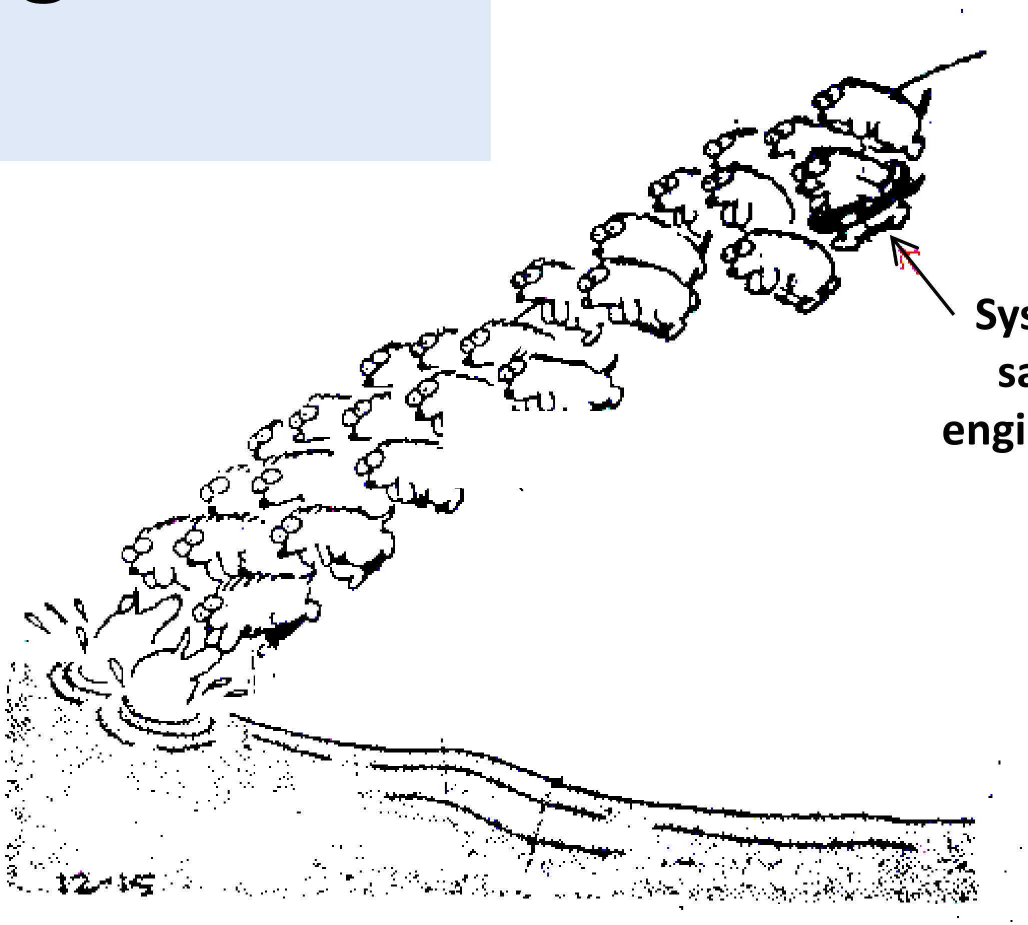
My 36 Years in System Safety: Looking Backward, Looking Forward

Nancy Leveson

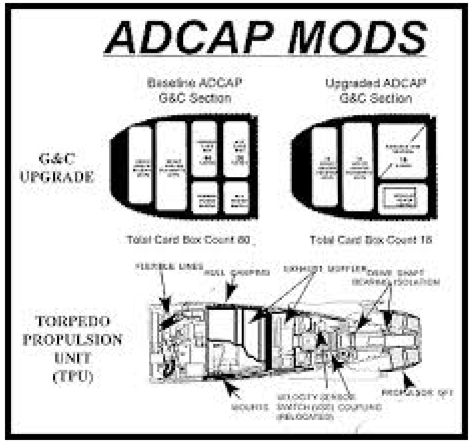


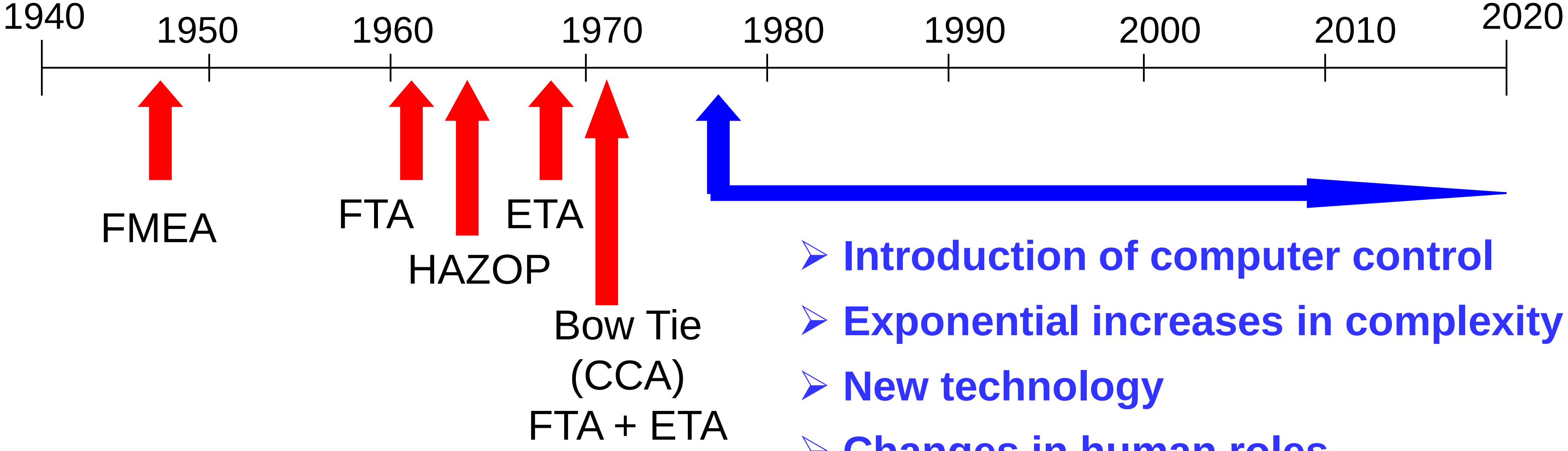
System safety engineer

(Gary Larsen, *The Far Side*)

Topics

- How I Got Started – Looking Backward
- A Systems Approach to Safety
- STAMP Today
- Where to From Here? – Looking Forward
- Lessons Learned Along the Way


ADCAP MODS
Baseline ADCAP G&C Section
Upgraded ADCAP G&C Section
G&C UPGRADE
Total Card Box Count 80
Total Card Box Count 18
TORPEDO PROPULSION UNIT (TPU)
FLEXIBLE LINES
HULL DAMPING
EXHAUST MUFFLER
DRIVE SHAFT BEARING ISOLATION
MOUNTS
VELOCITY SENSOR SWITCH (VSS) COUPLING (RELOCATION)
PROPULSOR KIT



- ➢ **Introduction of computer control**
- ➢ **Exponential increases in complexity**
- ➢ **New technology**
- ➢ **Changes in human roles**

Assumes accidents caused by component failures

Changes in the Past 36 Years

- New causes of accidents created by use of software
- Role of humans in systems and in accidents has changed
- Increased recognition of importance of organizational and social factors in accidents
- Faster pace of technological change
  - Learning from experience (“fly-fix-fly”) no longer as effective
  - Introduces “unknowns” and new paths to accidents
  - Less exhaustive testing is possible
- Increasing complexity
- Decreasing tolerance for single accidents

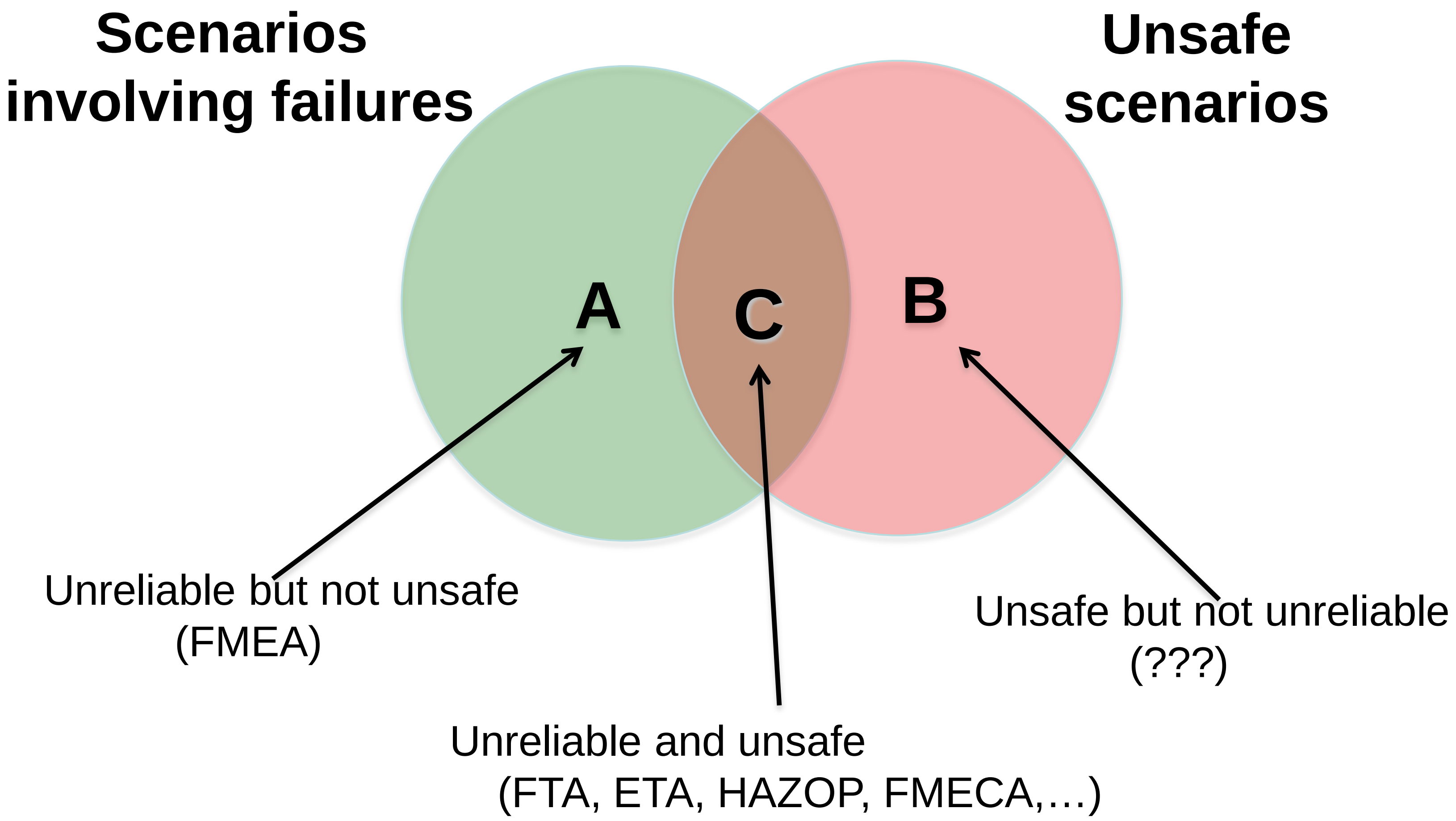
Reliability Engineering Approach to Safety

- Examples: fail-safe, defense in depth
- Many accidents occur without any component “failure”
  - Caused by equipment operation outside parameters and time limits upon which reliability analyses are based.
  - Caused by interactions of components all operating according to specification.
- Highly reliable components are not necessarily safe

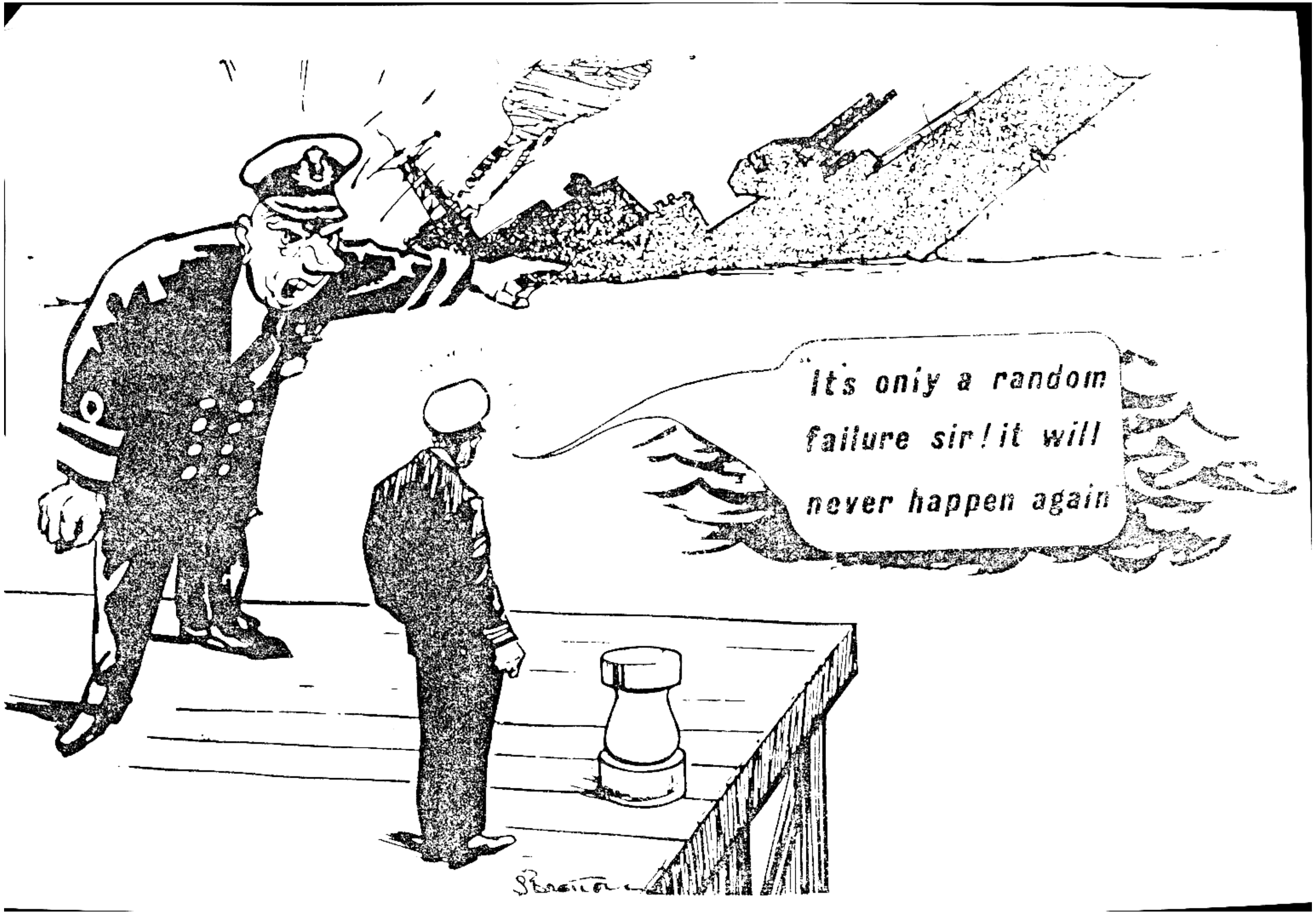
Reliability is NOT equal to safety in complex systems

What “Failed” Here?

- Navy aircraft were ferrying missiles from one location to another.
- One pilot executed a planned test by aiming at aircraft in front and firing a dummy missile.
- Nobody involved knew that the software was designed to substitute a different missile if the one that was commanded to be fired was not in a good position.
- In this case, there was an antenna between the dummy missile and the target so the software decided to fire a live missile located in a different (better) position instead.


Scenarios involving failures
Unsafe scenarios
A
C
B
Unreliable but not unsafe (FMEA)
Unsafe but not unreliable (???)
Unreliable and unsafe (FTA, ETA, HAZOP, FMECA,…)

Preventing Component or Functional Failures NOT Enough


It's only a random failure sir! it will never happen again

Why Our Efforts are Often Not Cost-Effective (1)

- Efforts superficial, isolated, or misdirected
  - Often isolated from engineering design
  - Spend too much time and effort on assurance, not building safety in from the beginning
    - Focusing on making arguments that systems are safe rather than making them safe
    - “Safety/Assurance cases”: Subject to confirmation bias
    - Traditional system safety tries to prove the system is unsafe (looks for paths to hazards), not that it is safe
    - Safety must be built in, it cannot be “assured in” or argued in

Why Our Efforts are Often Not Cost-Effective (2)

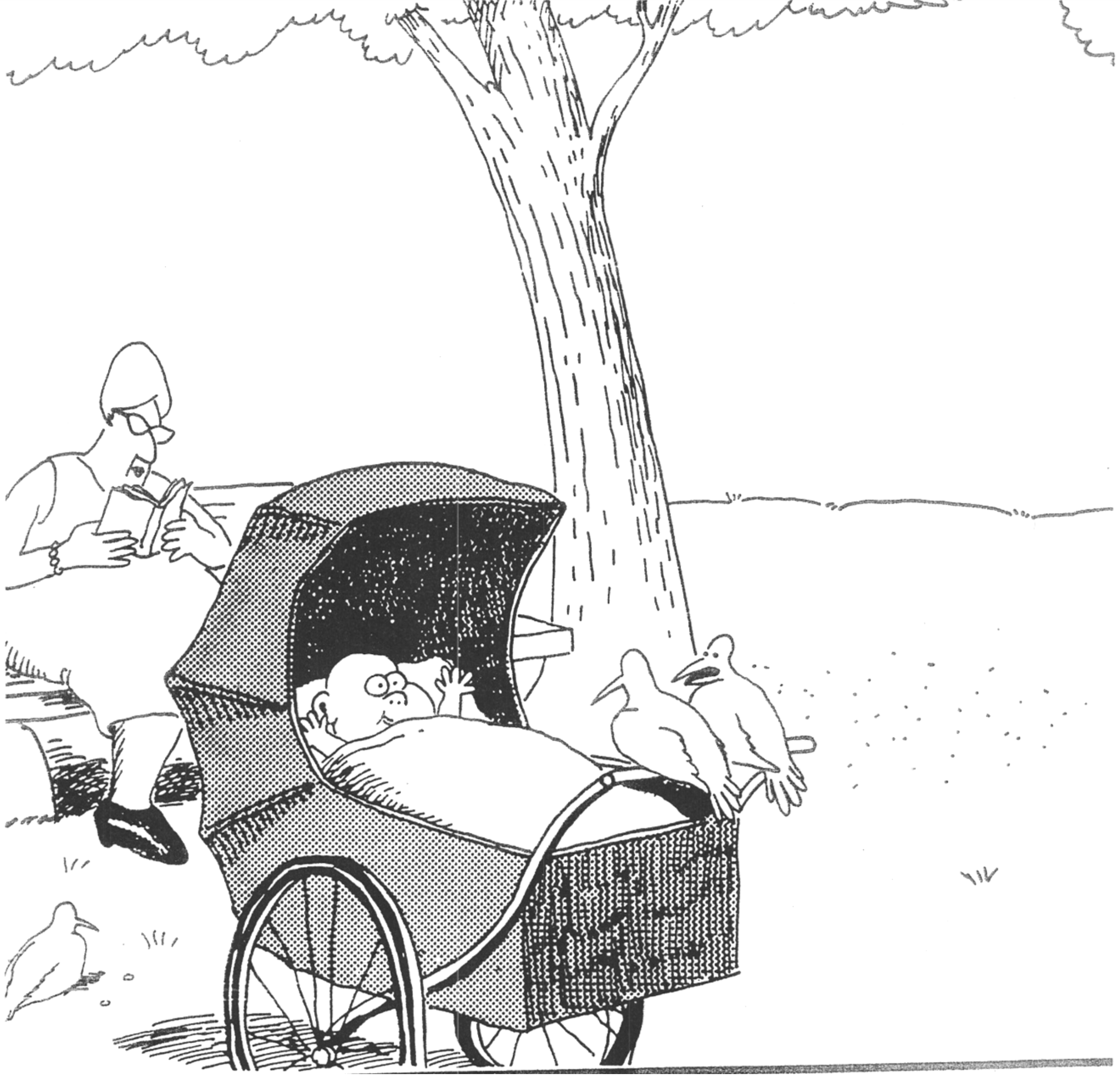
- Safety efforts start too late
    - 80-90% of safety-critical decisions made in early system concept formation (C.O. Miller)
    - Cannot “add” safety to an unsafe design
    - Most of our techniques require a relatively complete design to work
- Focus efforts only on technical components of systems
    - Ignore or only superficially handle
        - Management decision making
        - Operator error (and operations in general)
        - Safety culture
    - Focus on development and often ignore operations

Why Our Efforts are Often Not Cost-Effective (3)

- Using inappropriate analysis techniques for systems built today
- Need new, more powerful safety engineering approaches to deal with complexity and new causes of accidents
- Inadequate risk assessment
    - Applying probabilistic risk analysis for events that are not random
        - Software errors are design errors, not random failures
        - Human error is not random (slips vs. mistakes)
        - Component interaction accidents (system design errors) are not random
        - End up either leaving things out or making up numbers
    - Need better ways to assess and communicate risk

Why Our Efforts are Often Not Cost-Effective (4)

- Limited learning from events
  - Oversimplification of accident causation
  - “Blame is the enemy of safety”
    - Focus on “who” and not “why”
  - “Root cause” seduction
    - Believing in a “root cause” appeals to our desire for control
    - Leads to a sophisticated “whack a mole” game
      - Fix symptoms but not process that led to those symptoms
      - In continual fire-fighting mode
      - Having the same accident over and over



It’s still hungry … and I’ve been stuffing worms into it all day.

Summary

- Doing things that require great effort and resources but demonstrably do not work
  - Don’t seem to notice
  - Almost no evaluations of old techniques

The Problem is Complexity

- How do we traditionally deal with complexity?

1. Analytic Reduction
2. Statistics

[Recommended reading: Peter Checkland, “Systems Thinking, Systems Practice,” John Wiley, 1981]

Analytic Reduction

- Divide system into distinct parts for analysis

  Physical aspects → Separate physical or functional components

  Behavior → Events over time

- Examine parts separately and later combine analysis results

- Assumes such separation does not distort phenomenon
  - Each component or subsystem operates independently
  - Components act the same when examined singly as when playing their part in the whole
  - Events not subject to feedback loops and non-linear interactions

Statistics

- Treat system as a structureless mass with interchangeable parts
- Use Law of Large Numbers to describe behavior in terms of averages
- Assumes components are sufficiently regular and random in their behavior that they can be studied statistically

Traditional Approach to Safety

- Uses Analytic Reduction and Statistics
- Divide system into components
  - Assume accidents are caused by component failure
  - Identify chains of directly related physical or logical (functional) component failures that can lead to a loss
  - Evaluate reliability of components separately and later combine analysis results into a system reliability value

Note: Assume randomness in the failure events so can derive probabilities for a loss

Software and humans do not satisfy this assumption

Accidents are Treated as Chains of Failure Events

- Forms the basis for most safety engineering and reliability engineering analysis:

  FTA, PRA, FMEA/FMECA, Event Trees, FHA, etc.

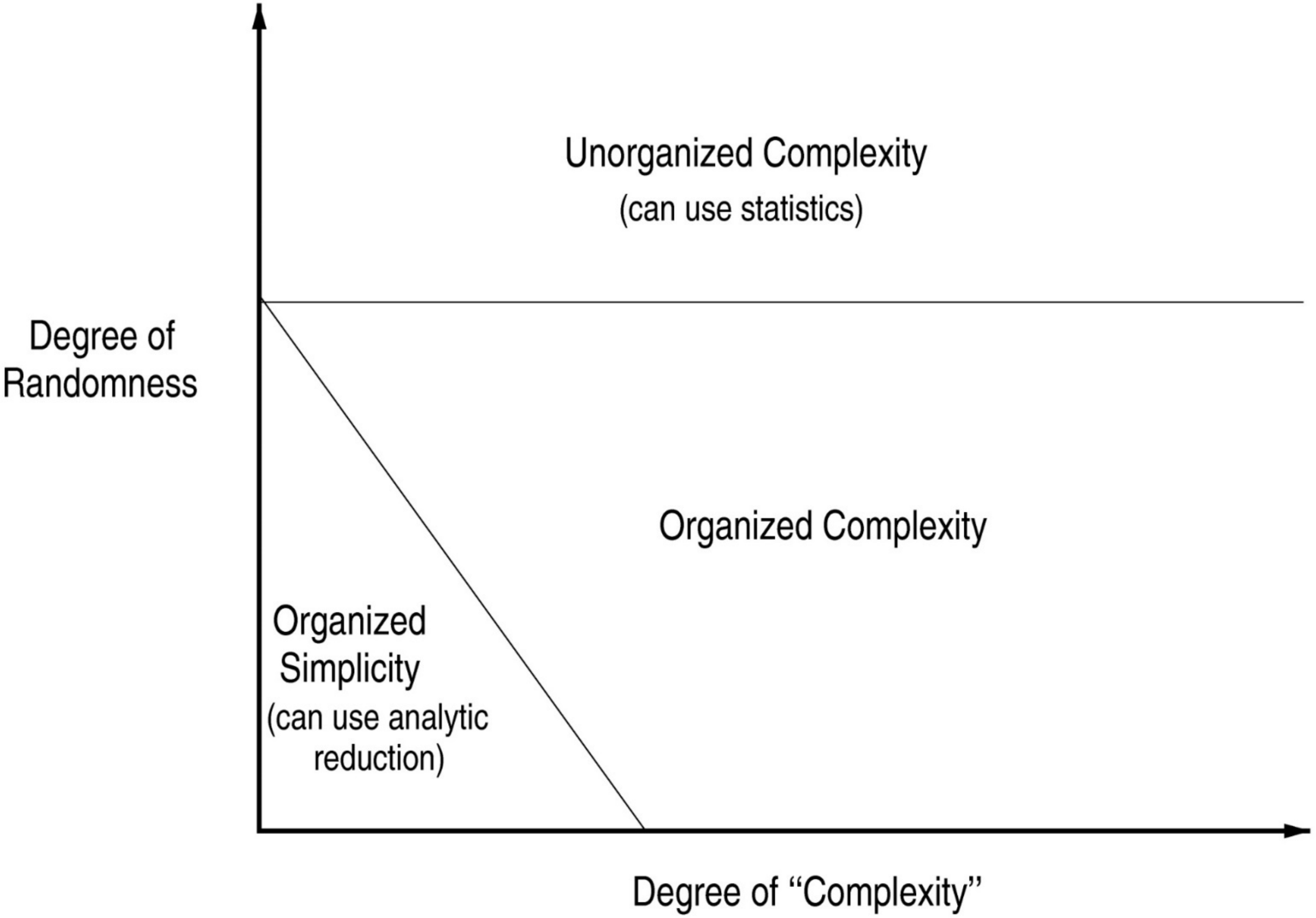
  and design (concentrate on dealing with component failure):

  Redundancy and barriers (to prevent failure propagation),

  High component integrity and overdesign,

  Fail-safe design,

  (humans) Operational procedures, checklists, training, ….



(Gerald Weinberg, An Introduction to General Systems Thinking)

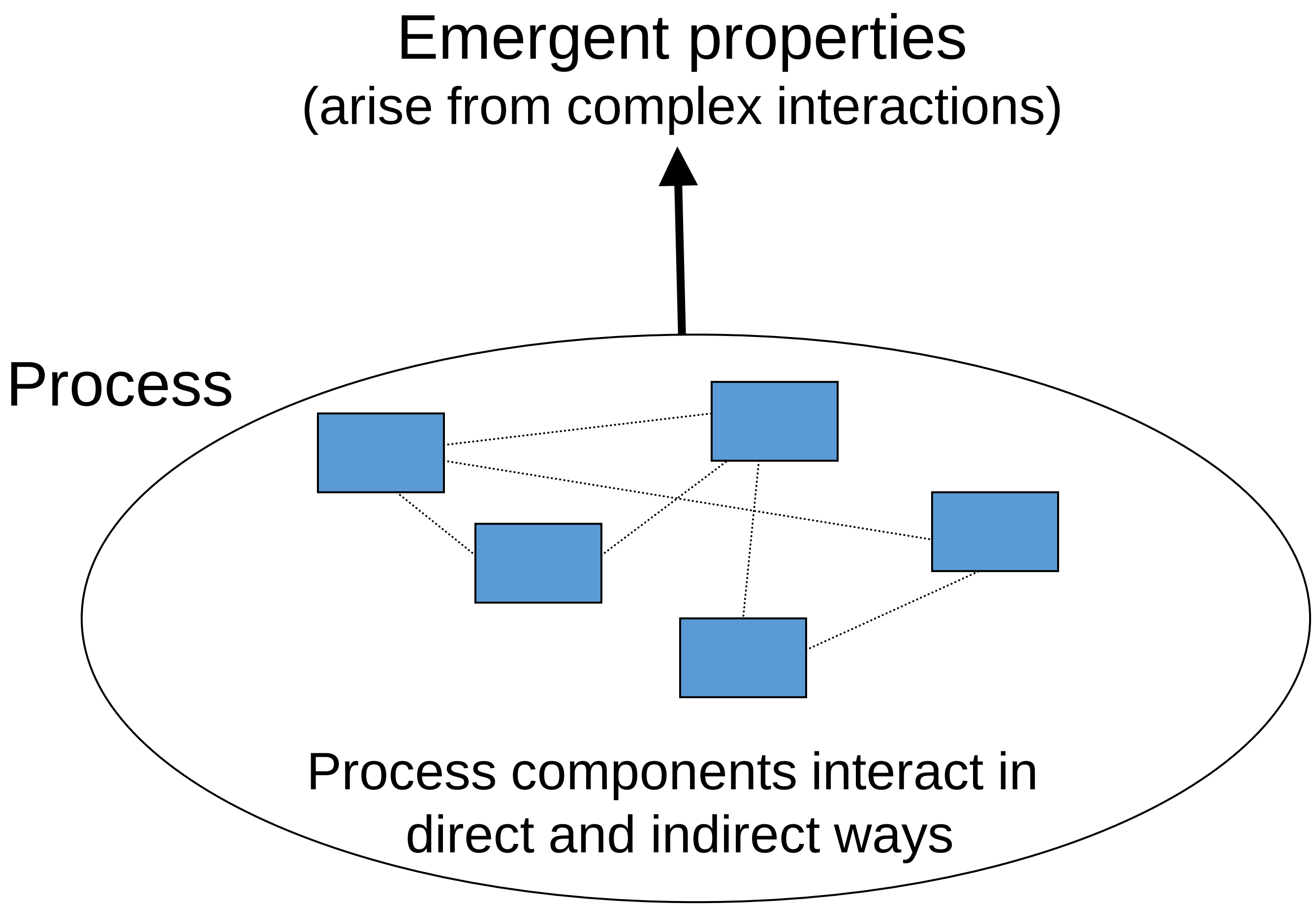
Applying Systems Thinking to Safety (STAMP)

- Accidents involve a complex, dynamic “process”
  - Not simply chains of failure events
  - Arise in interactions among humans, machines and the environment

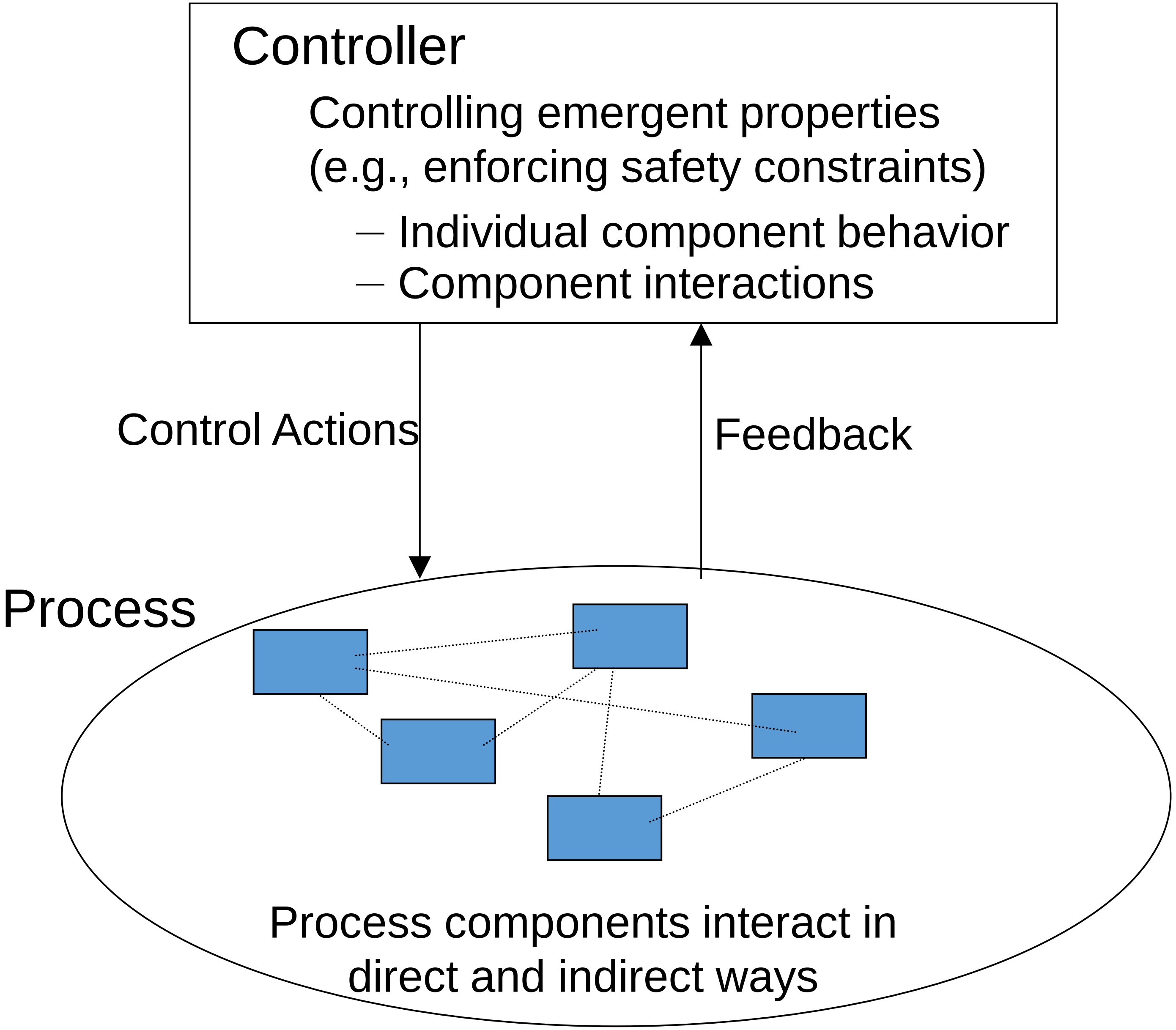
- Treat safety as a dynamic control problem
  - Safety requires enforcing a set of constraints on system behavior
  - Accidents occur when interactions among system components violate those constraints
  - Safety becomes a control problem rather than just a reliability problem

Examples of Safety Constraints

- Power must never be on when access door open
- Two aircraft must not violate minimum separation
- Aircraft must maintain sufficient lift to remain airborne
- Public health system must prevent exposure of public to contaminated water and food products
- Pressure in a deep water well must be controlled
- Runway incursions and operations on wrong runways or taxiways must be prevented


Emergent properties
(arise from complex interactions)
Process
Process components interact in direct and indirect ways

- **“The whole is greater than the sum of its parts”**
- **Safety and security are examples of emergent properties**



Controller

Controlling emergent properties (e.g., enforcing safety constraints)

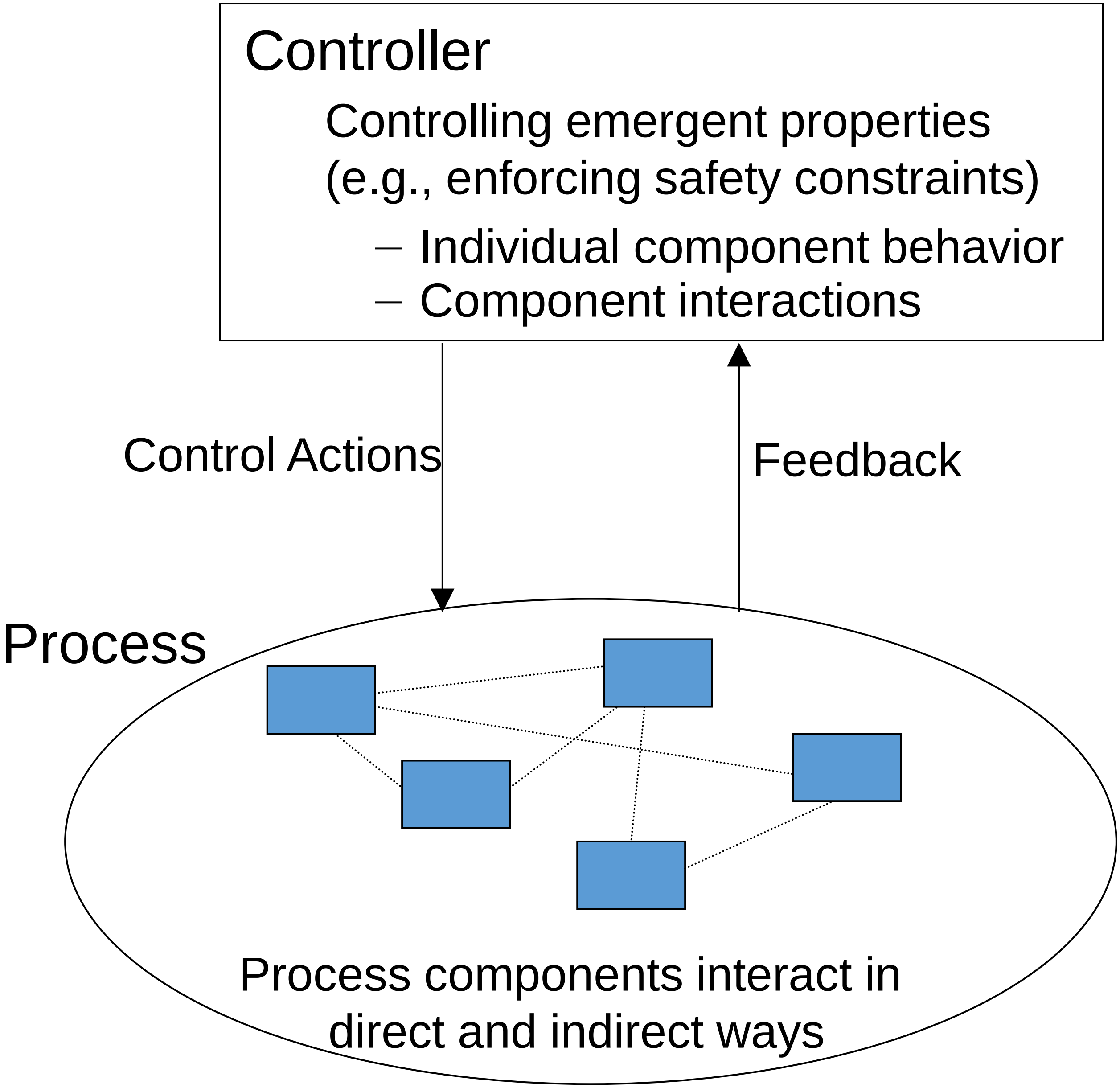
- Individual component behavior
- Component interactions

Control Actions

Feedback

Process

Process components interact in direct and indirect ways



Controller

Controlling emergent properties (e.g., enforcing safety constraints)

– Individual component behavior
– Component interactions

Control Actions

Feedback

Process

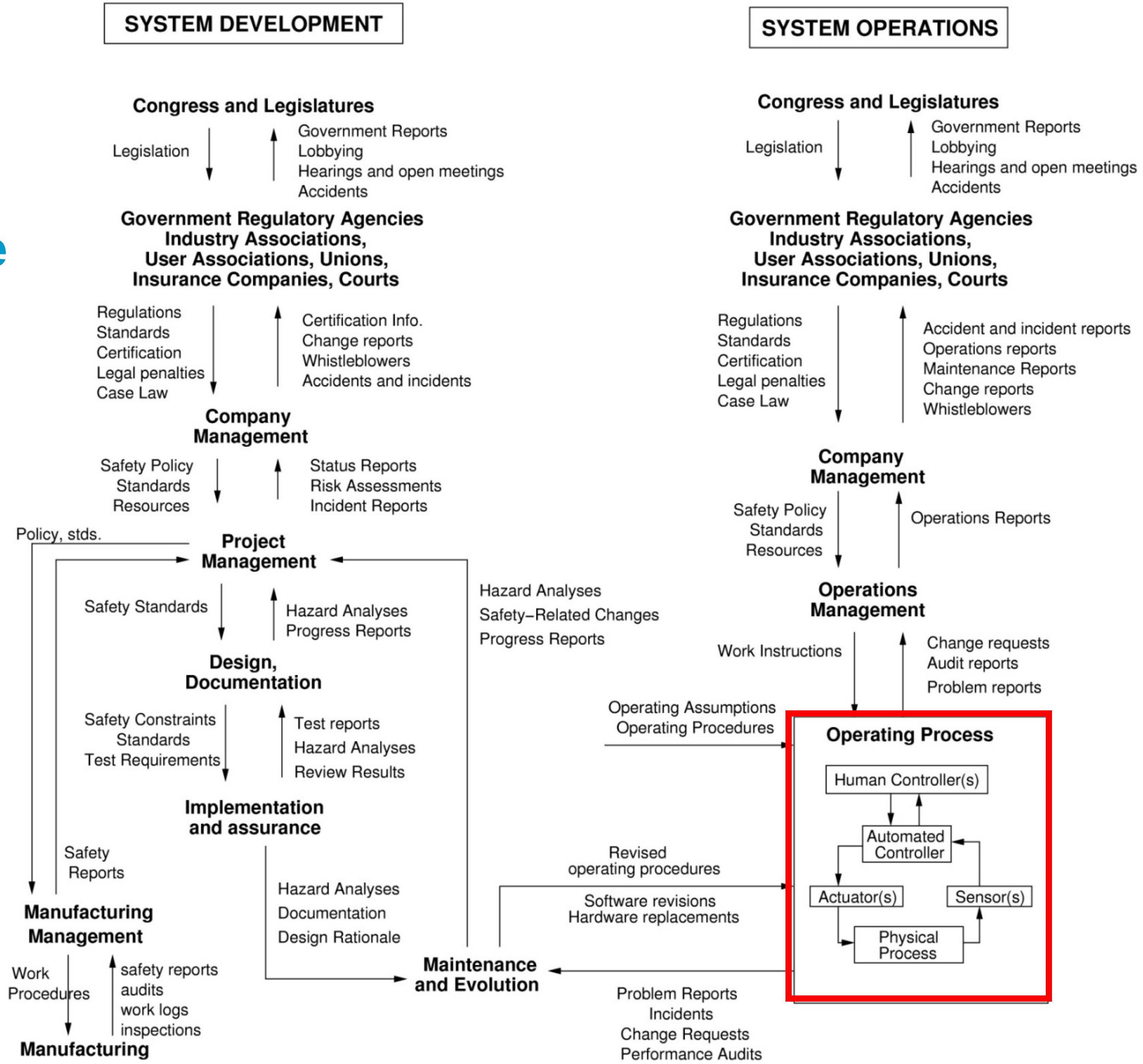
Process components interact in direct and indirect ways

Air Traffic Control:

Safety

Throughput

Example Safety Control Structure



SYSTEM DEVELOPMENT

Congress and Legislatures

Legislation

Government Reports
Lobbying
Hearings and open meetings
Accidents

Government Regulatory Agencies
Industry Associations,
User Associations, Unions,
Insurance Companies, Courts

Regulations
Standards
Certification
Legal penalties
Case Law

Certification Info.
Change reports
Whistleblowers
Accidents and incidents

Company
Management

Safety Policy
Standards
Resources

Status Reports
Risk Assessments
Incident Reports

Policy, stds.

Project
Management

Safety Standards

Hazard Analyses
Progress Reports

Hazard Analyses
Safety–Related Changes
Progress Reports

Design,
Documentation

Safety Constraints
Standards
Test Requirements

Test reports
Hazard Analyses
Review Results

Implementation
and assurance

Safety
Reports

Hazard Analyses
Documentation
Design Rationale

Manufacturing
Management

Work
Procedures

safety reports
audits
work logs
inspections

Manufacturing

Maintenance
and Evolution

SYSTEM OPERATIONS

Congress and Legislatures

Legislation

Government Reports
Lobbying
Hearings and open meetings
Accidents

Government Regulatory Agencies
Industry Associations,
User Associations, Unions,
Insurance Companies, Courts

Regulations
Standards
Certification
Legal penalties
Case Law

Accident and incident reports
Operations reports
Maintenance Reports
Change reports
Whistleblowers

Company
Management

Safety Policy
Standards
Resources

Operations Reports

Operations
Management

Work Instructions

Change requests
Audit reports
Problem reports

Operating Assumptions
Operating Procedures

Operating Process

Human Controller(s)

Automated
Controller

Actuator(s)

Sensor(s)

Physical
Process

Revised
operating procedures

Software revisions
Hardware replacements

Problem Reports
Incidents
Change Requests
Performance Audits

Safety Treated as a Dynamic Control Problem

- **Goal: Design an effective control structure that eliminates or reduces adverse events**.
  - Need clear definition of expectations, responsibilities, authority, and accountability at all levels of safety control structure
  - Need appropriate feedback
  - Entire control structure must together enforce the system safety property (constraints)
    - Physical design (inherent safety)
    - Operations
    - Management
    - Social interactions and culture

A Broad View of "Control"

Component failures and unsafe interactions may be "controlled" through design

(e.g., redundancy, interlocks, fail-safe design)

or through process

- Manufacturing processes and procedures
- For humans, change the context in which they are operating
- Maintenance processes
- Operations

or through social controls (e.g., regulatory, insurance, legal, culture, or individual self-interest)

- For humans, change the context in which they are operating
- Human error is a symptom of a system that needs to be redesigned.

STAMP (System Theoretic Accident Model and Processes)

- A new, more powerful accident causality model
- Based on systems theory, not reliability theory
- Treats accidents as a dynamic control problem (vs. a failure problem)
- Includes
  - Entire socio-technical system (not just technical part)
  - Component interaction accidents
  - Software and system design errors
  - Human errors

Paradigm Change

- Does not imply what previously done is wrong and new approach correct

- Einstein:

  “Progress in science (moving from one paradigm to another) is like climbing a mountain”

As move further up, can see farther than on lower points

Paradigm Change (2)

New perspective does not invalidate the old one, but extends and enriches our appreciation of the valleys below

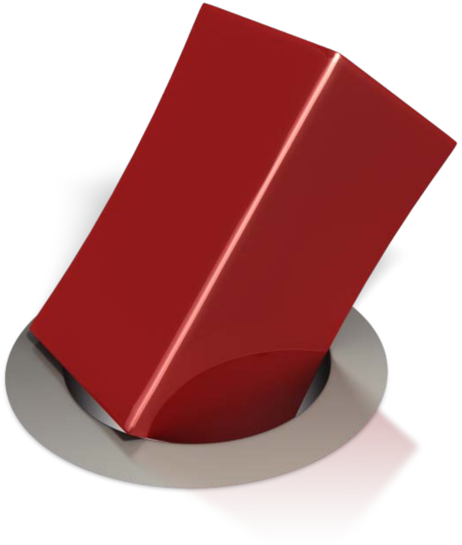
Value of new paradigm often depends on ability to accommodate successes and empirical observations made in old paradigm.

New paradigms offer a broader, rich perspective for interpreting previous answers.

Resist trying to integrate systems thinking with analytic reduction

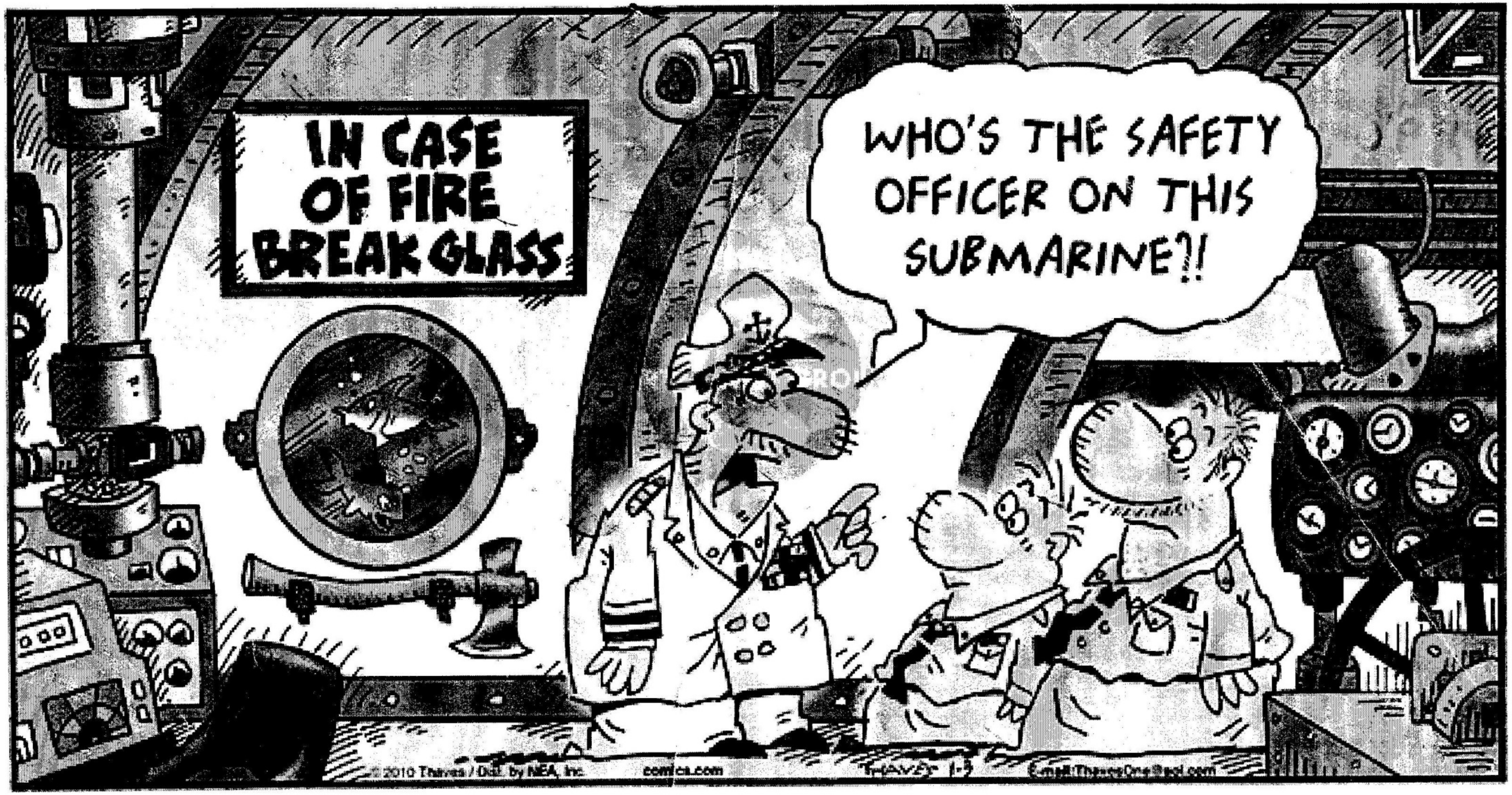
- Trying to shoehorn new technology and new levels of complexity into old methods does not work

- Trying to merge systems thinking into the old models and techniques will not work

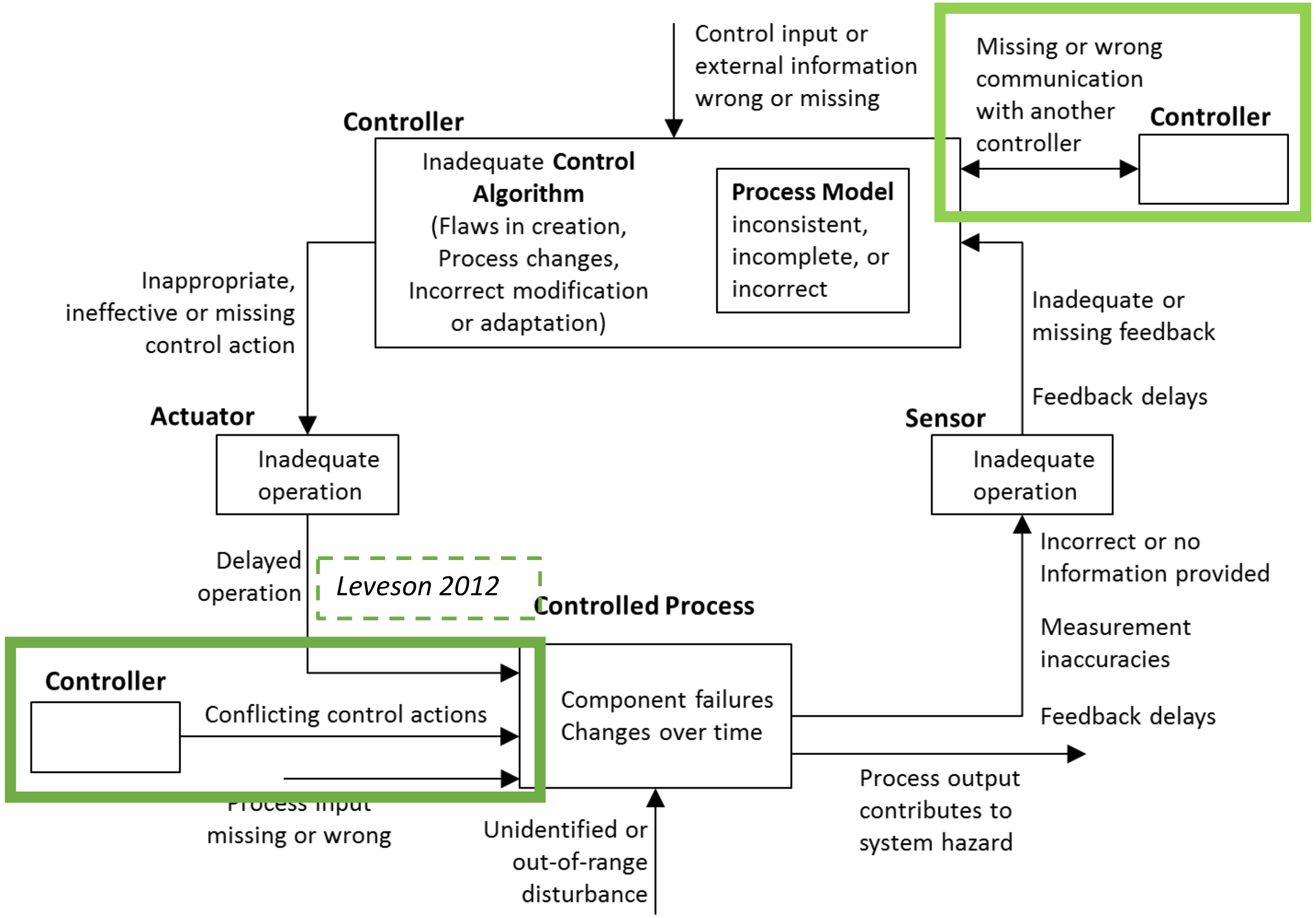


Recent Progress

- Large companies are starting training programs in STPA for their employees
- DoD training program in using STPA for security
- Cited as an example in ISO 26262 draft (out in 2018)
- Recent successes in applying to workplace safety and engineering management
- Lots of evaluations and comparisons with traditional techniques all with STPA finding things that the traditional techniques do not
- Lots of new applications


IN CASE OF FIRE BREAK GLASS
WHO'S THE SAFETY OFFICER ON THIS SUBMARINE?!
THAVES 1-5

Adding Coordination to STPA: Col. Kip Johnson (9/2016)


Control input or external information wrong or missing
Missing or wrong communication with another controller
Controller
Controller
Inadequate Control Algorithm (Flaws in creation, Process changes, Incorrect modification or adaptation)
Process Model inconsistent, incomplete, or incorrect
Inappropriate, ineffective or missing control action
Inadequate or missing feedback
Feedback delays
Actuator
Inadequate operation
Sensor
Inadequate operation
Delayed operation
Leveson 2012
Controlled Process
Incorrect or no Information provided
Measurement inaccuracies
Controller
Conflicting control actions
Component failures
Changes over time
Feedback delays
Process input missing or wrong
Unidentified or out-of-range disturbance
Process output contributes to system hazard

Some Important Research Problems

- Applying STAMP to other properties besides safety and security
- How to integrate into a large company (training, facilitators, how to implement a paradigm change into industries?)
- More help with generating causal scenarios from UCAs
- Generating UCAs (Thomas method complete but harder to teach and maybe do, use to check completeness?)
- Risk assessment without resorting to unknown and unknowable probabilities
- Use in operations
- Controlling unplanned and unsafe changes
- Human factors in STPA and CAST

Lessons I've Learned over 36 Years

- "It is important that students bring a certain ragamuffin barefoot irreverence to their studies. They are here not to worship what is known, but to question it." Jacob Bronowski, *The Ascent of Man*
  - The starting point is to question our assumptions.
  - "It's never what we don't know that stops us. It's what we do know that just ain't so"

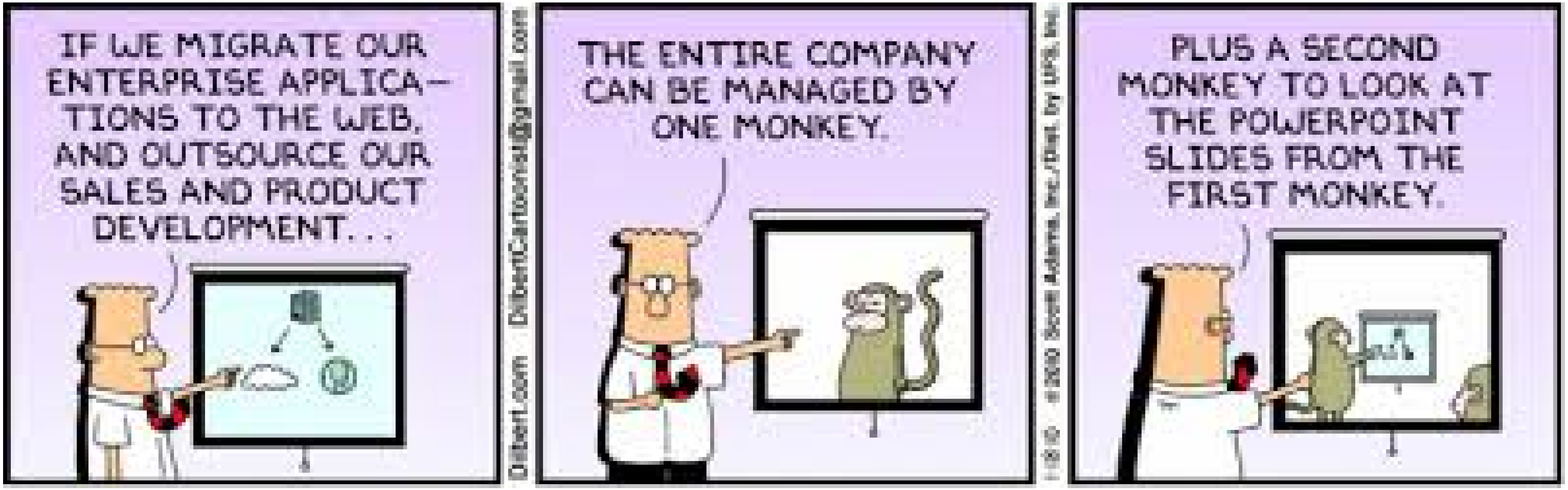
- If you want to make important contributions, work on important problems
  - Pick the problem first, not the solution
  - Understanding a problem is the first step to solving it
  - Don't play "follow the leader" or jump on bandwagons

- Work on problems you care about

  From an anonymous proposal review: "Nancy is passionate about safety, which is her greatest strength and her greatest weakness"

Summary of Where We Need to Go in System Safety

- Expand our accident causation models
- Create new, more powerful and inclusive hazard analysis techniques
- Use new system design techniques
  - Safety-guided design
  - Integrate System Safety more into system engineering
- Improve accident analysis and learning from events
- Improve control of safety during operations
- Improve management decision-making and safety culture


IF WE MIGRATE OUR ENTERPRISE APPLICATIONS TO THE WEB, AND OUTSOURCE OUR SALES AND PRODUCT DEVELOPMENT...
Dilbert.com DilbertCartoonist@gmail.com
THE ENTIRE COMPANY CAN BE MANAGED BY ONE MONKEY.
1-18-10 © 2010 Scott Adams, Inc./Dist. by UFS, Inc.
PLUS A SECOND MONKEY TO LOOK AT THE POWERPOINT SLIDES FROM THE FIRST MONKEY.