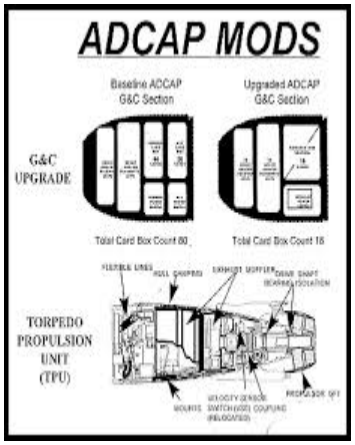


A Systems Approach to Safety and Cybersecurity

Nancy Leveson

MIT




ADCAP MODS
Baseline ADCAP G&C Section
Upgraded ADCAP G&C Section
G&C UPGRADE
Total Card Box Count 80
Total Card Box Count 16
TORPEDO PROPULSION UNIT (TPU)

General Definition of “Safety”

- Accident = Loss: Any undesired and unplanned event that results in a loss
  - loss of human life or injury,
  - property damage,
  - environmental pollution,
  - mission loss,
  - negative business impact (damage to reputation, etc.), product launch delay, legal entanglements
- Includes inadvertent and intentional
- System goals vs. constraints (limits on how can achieve the goals)
- Hazard/vulnerability: A system state or set of conditions that, together with worst-case environmental conditions, will lead to a loss

Understanding The Problem

“It’s never what we don’t know that stops us. It’s what we do know that just ain’t so.”

Our current tools are all 50-65 years old but our technology is very different today

1940 1950 1960 1970 1980 1990 2000 2010 2020

FMEA

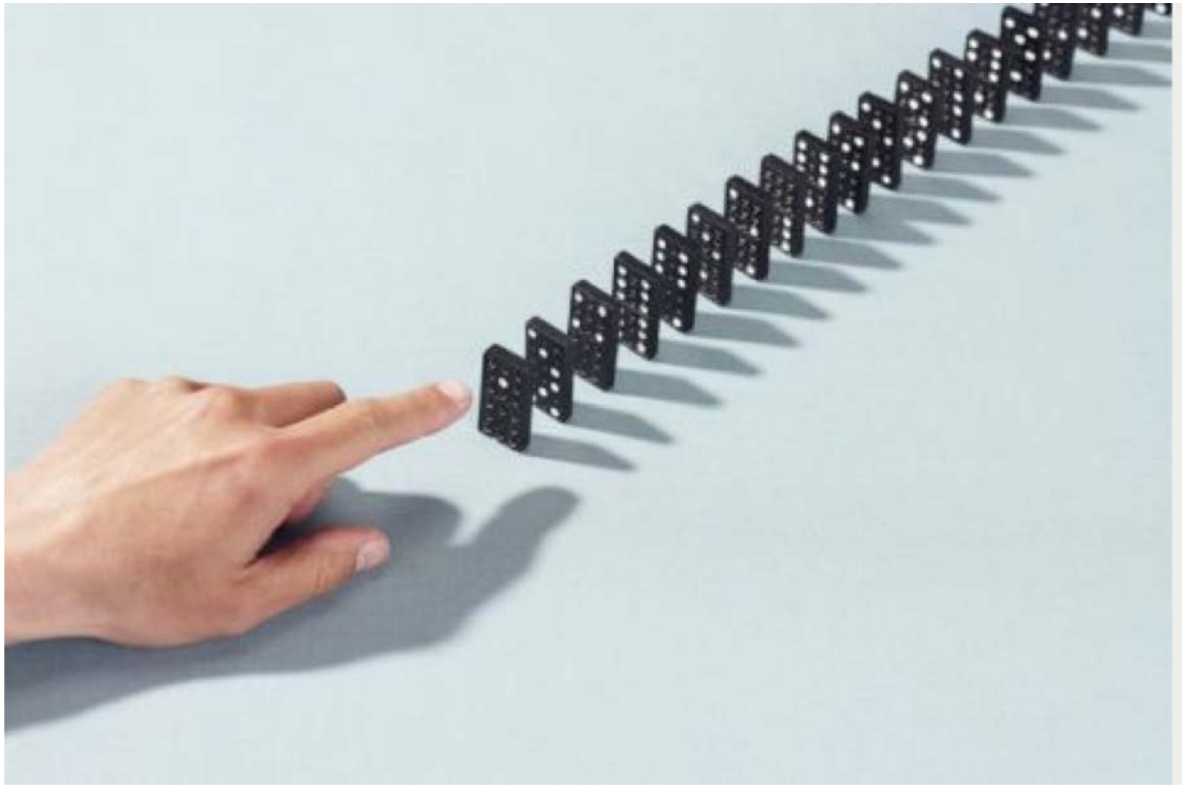
FTA HAZOP ETA

Bow Tie (CCA) FTA + ETA

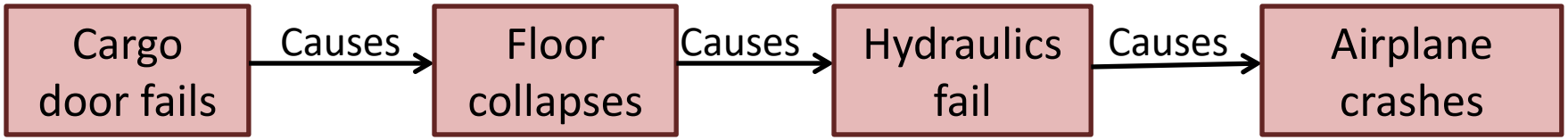
- Introduction of computer control
- Exponential increases in complexity
- New technology
- Changes in human roles

Assumes accidents caused by component failures

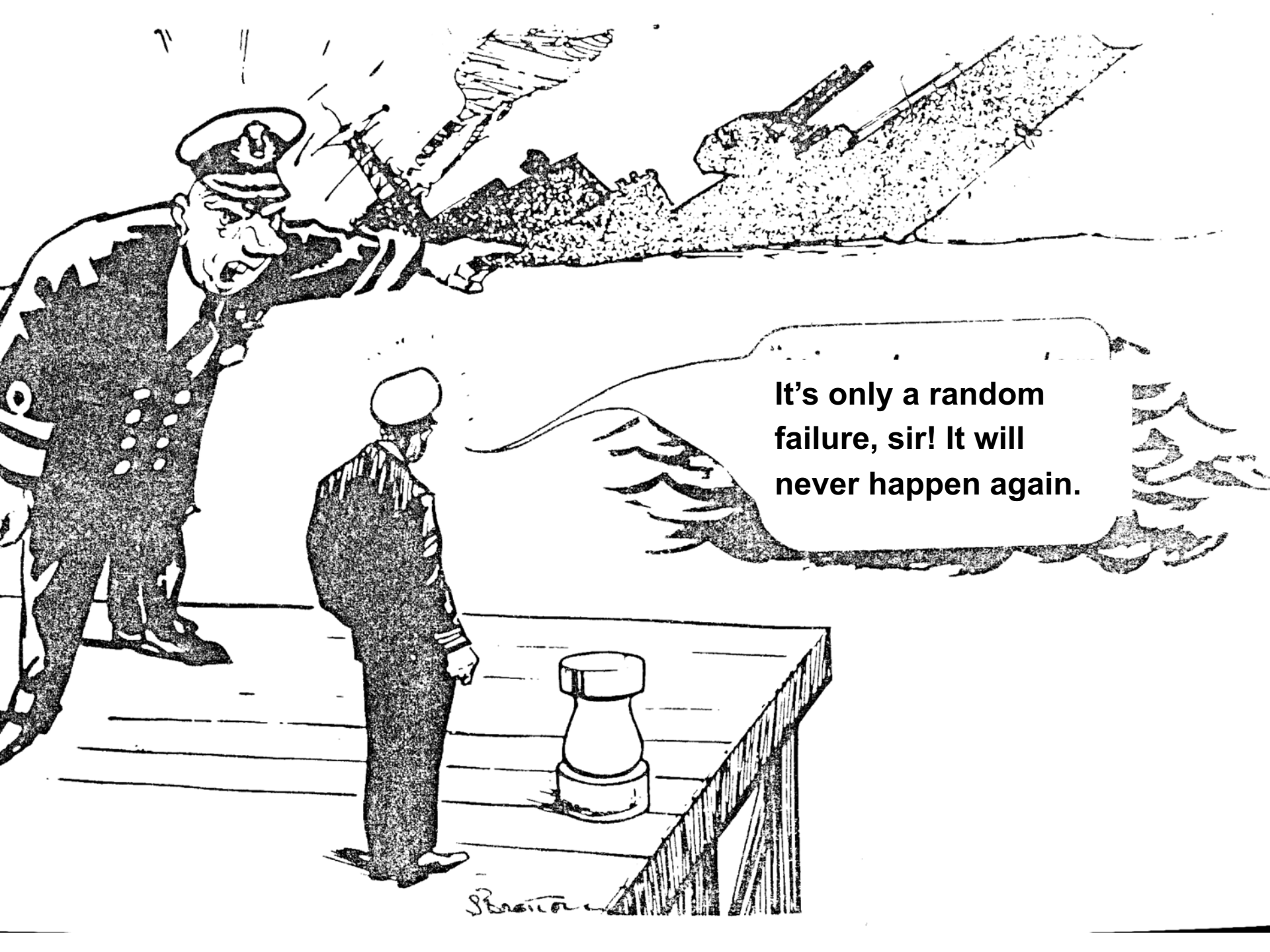
Domino “Chain of events” Model



DC-10:



Failure Event-Based


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

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.

Boeing 787 Lithium Battery Fires

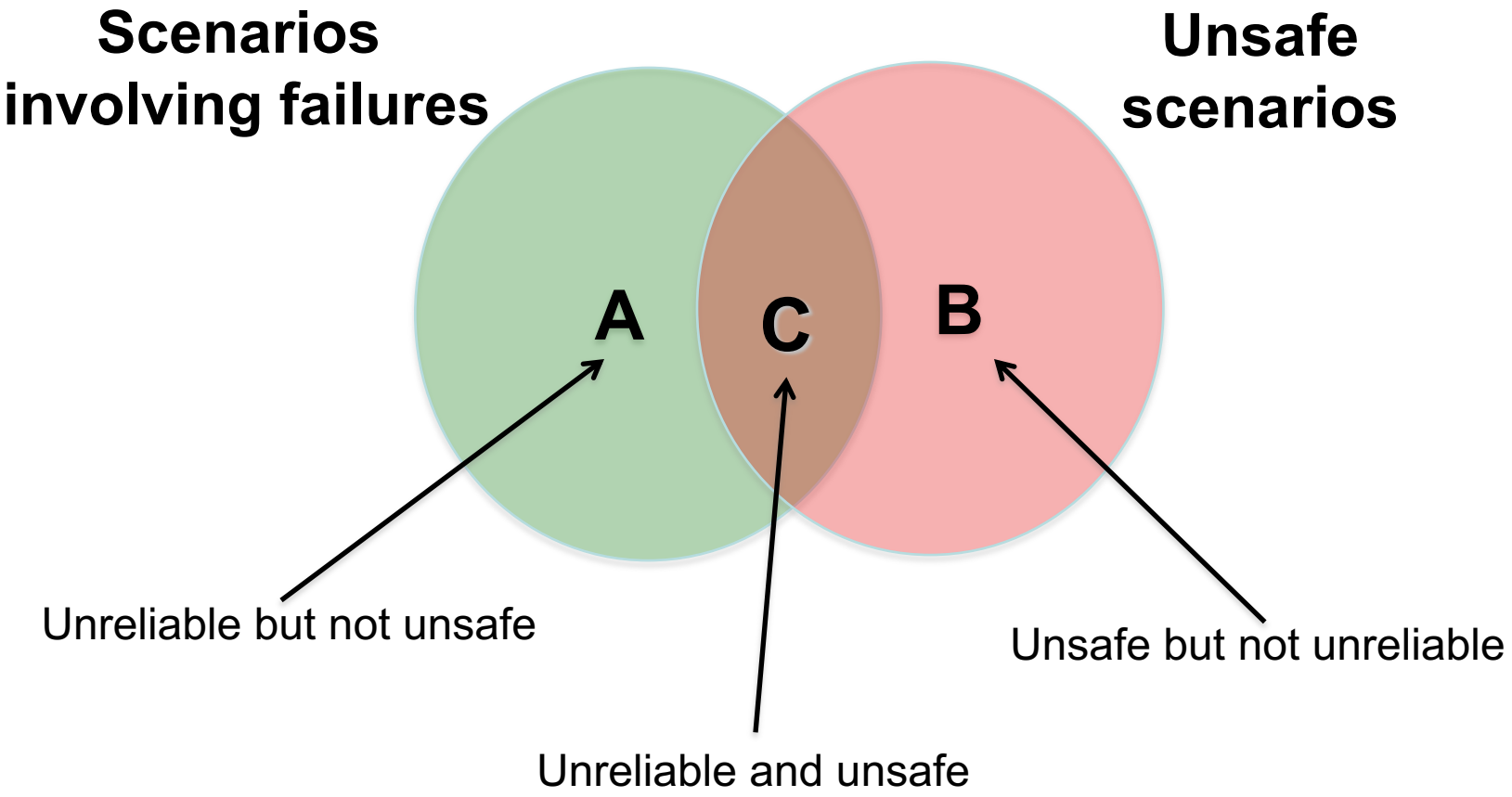
Models predicted 787 battery thermal problems would occur once in 10 million flight hours…but two batteries overheated in just two weeks in 2013

Boeing 787 Lithium Battery Fires

- A module monitors for smoke in the battery bay, controls fans and ducts to exhaust smoke overboard.
- Power unit monitors for low battery voltage, shut down various electronics, including ventilation
- Smoke could not be redirected outside cabin

All software requirements were satisfied!
The requirements were unsafe

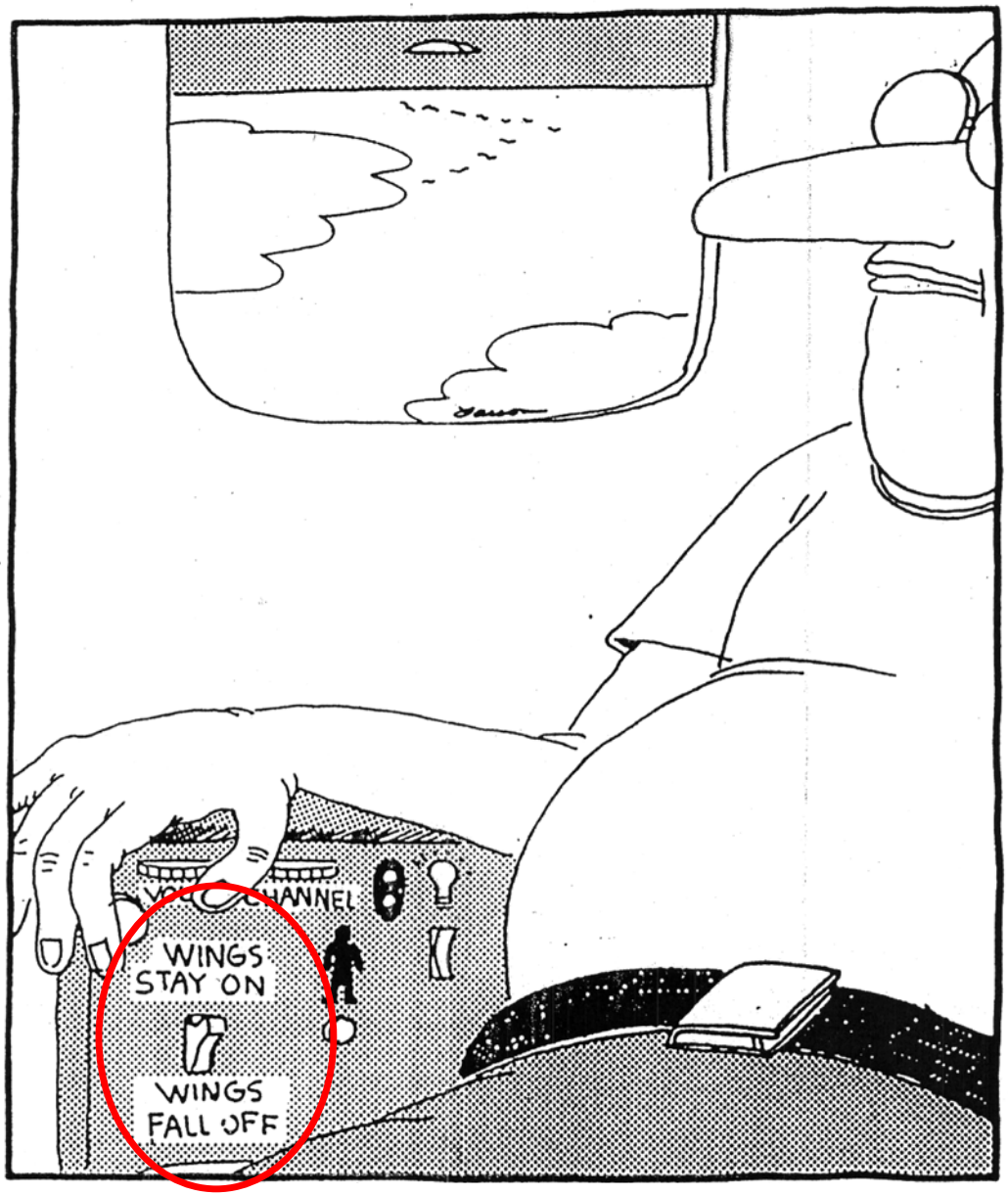
Safety and Reliability are Different



Preventing Component or Functional Failures is Not Enough

The Problem is Complexity

- Systems are becoming more complex
  - Accidents often result from interactions among components
  - Too complex to anticipate all potential interactions (“unknown unknowns”)
- We can no longer:
  - Plan, understand, anticipate, and guard against all undesired system behavior
  - Exhaustively test to get out all design errors
- Design of automation is creating new types of human “error”



Fumbling for his recline button Ted unwittingly instigates a disaster

Change in the Way We Conceive of Human Error

Traditional Approach:

- Operators/pilots responsible for most accidents
- So fire, train them not to make mistakes, or add more automation (which marginalizes the pilot and causes more and different errors)

Systems Approach:

- Human behavior always affected by the context in which it occurs
  - We are designing systems in which human error inevitable
  - **Human error is a symptom of a system that needs to be redesigned**.
- To eliminate human errors, need to change the system design

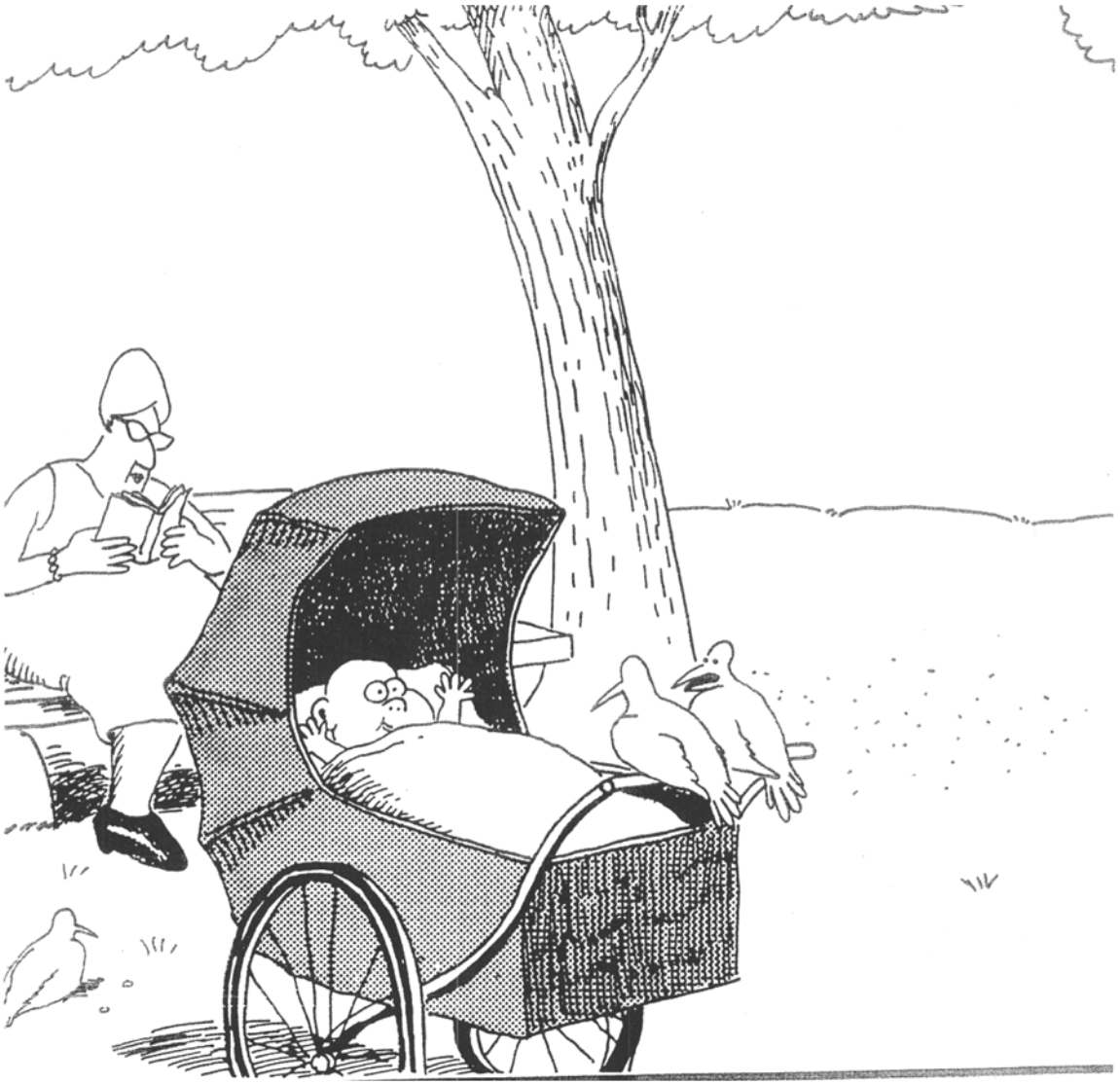
Bottom Line: We Need Something New

- Two approaches being taken now:

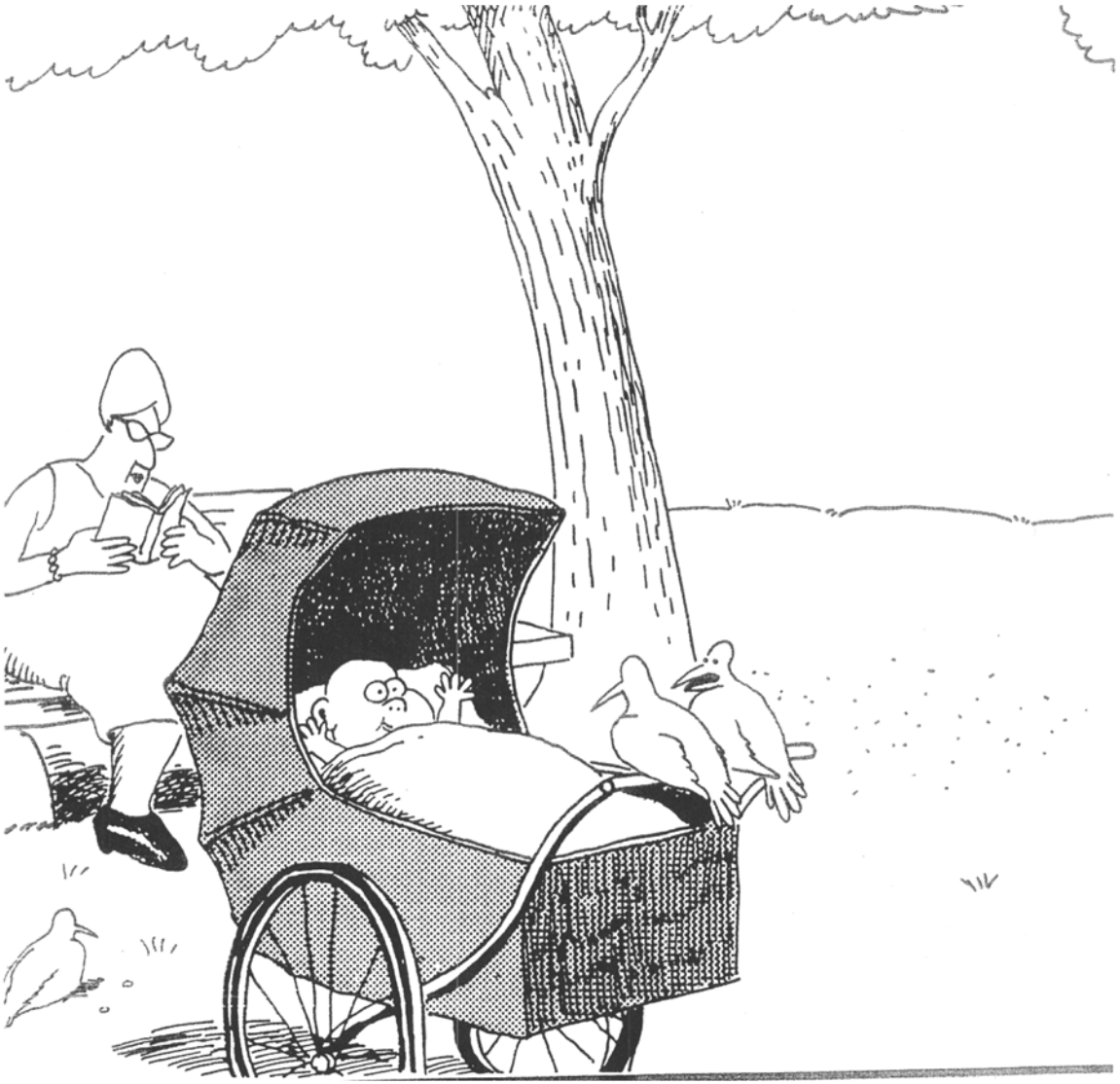
Pretend there is no problem

Shoehorn new technology and new levels of complexity into old methods

New levels of complexity are creating new problems that cannot be solved using traditional techniques.



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



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

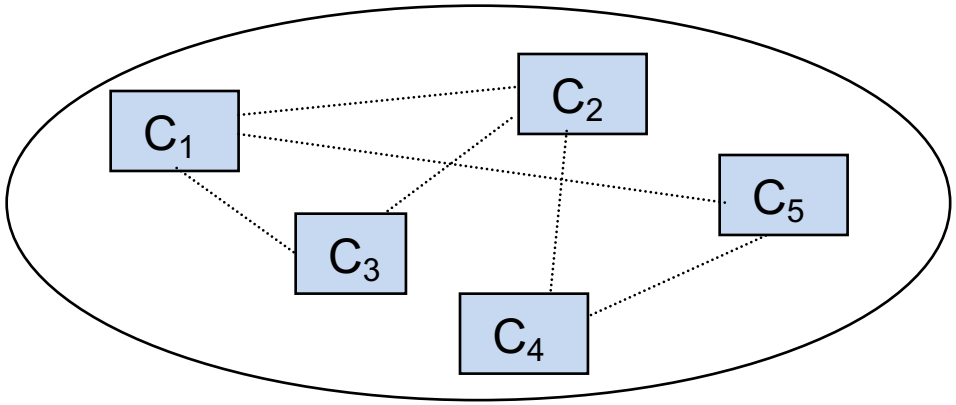
We Need New Tools for the New Problems

Traditional Approach to Coping with Complexity

Analytic Decomposition ("Divide and Conquer")

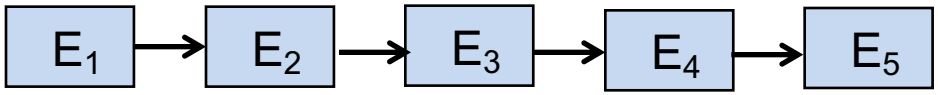
1. Divide system into separate parts

Physical/Functional: Separate into distinct components



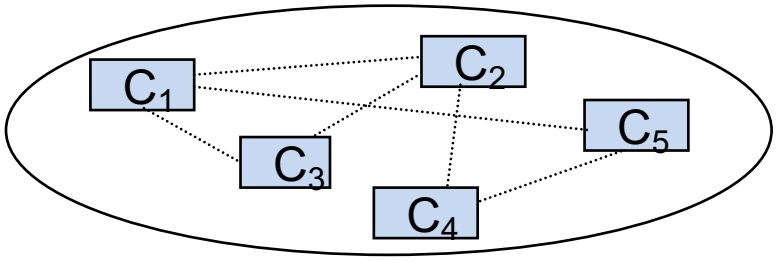
Components interact
In direct ways

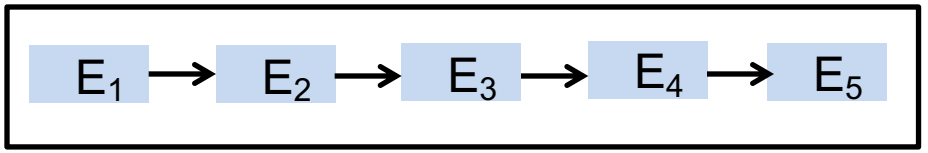
Behavior: Separate into events over time



Each event is the direct
result of the preceding event

Analytic Decomposition (2)





2. Analyze/examine pieces separately and combine results

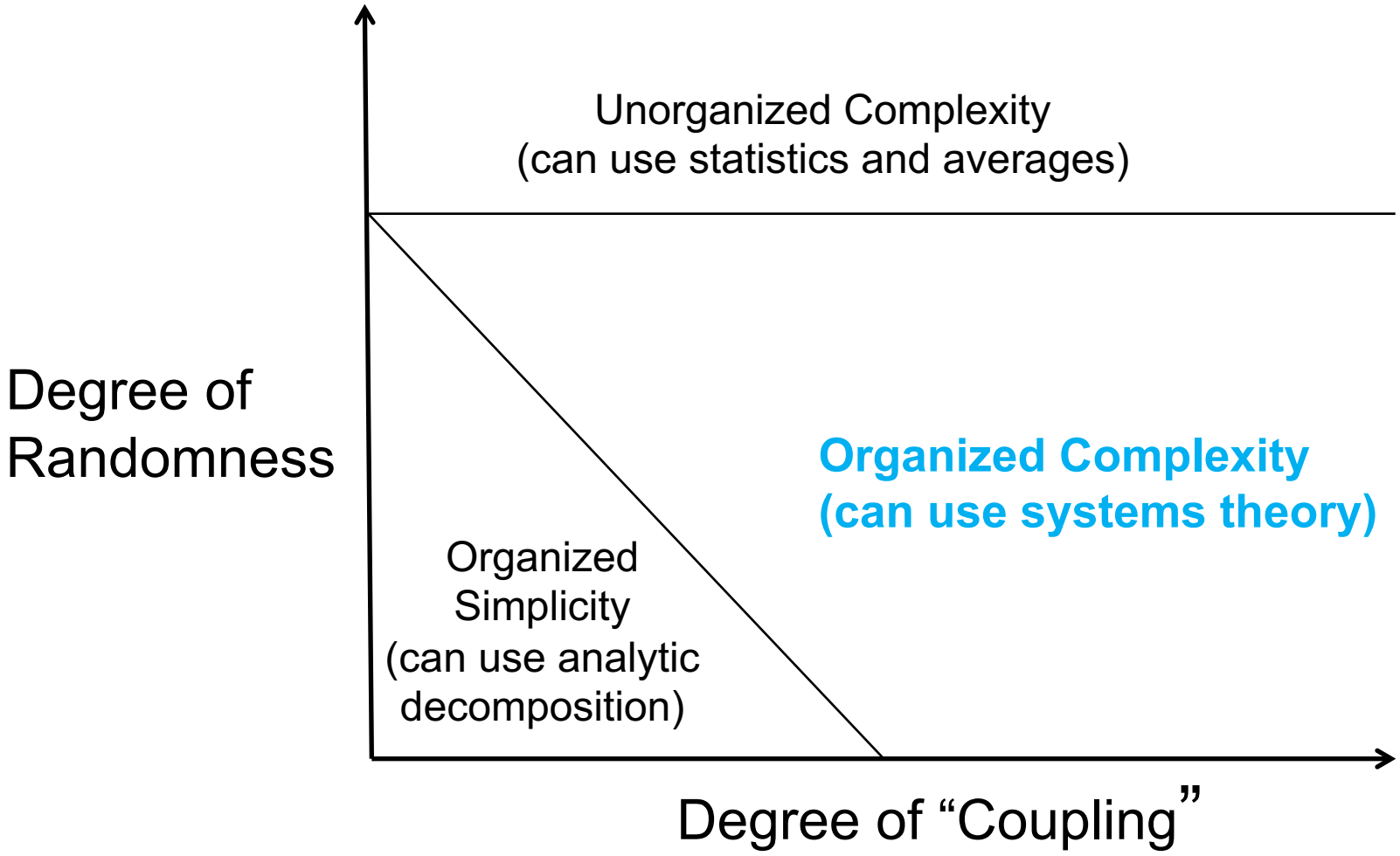
- Assumes such separation does not distort phenomenon
  - ✓ Each component or subsystem operates independently
  - ✓ Components act the same when examined singly as when playing their part in the whole
  - ✓ Components/events not subject to feedback loops and non-linear interactions
  - ✓ Interactions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

Bottom Line

- These assumptions are no longer true in our
  - Tightly coupled
  - Software intensive
  - Highly automated
  - Interconnected

  engineered systems today

- Need a new theoretical basis for discovering new approaches and tools
  - *System theory* can provide it



(Gerald Weinberg, An Introduction to General Systems Theory, 1975)

A Potential Way Forward

A Systems Theoretic View of Safety and Security

Systems Theory

- Developed for systems that are
  - Too complex for complete analysis
    - Separation into (interacting) subsystems distorts the results
    - The most important properties are emergent
  - Too organized for statistics
    - Too much underlying structure that distorts the statistics
    - New technology and designs have no historical information
- First used on ICBM systems of 1950s/1960s

System Theory was created to provide a more powerful way to deal with complexity

Systems Theory (2)

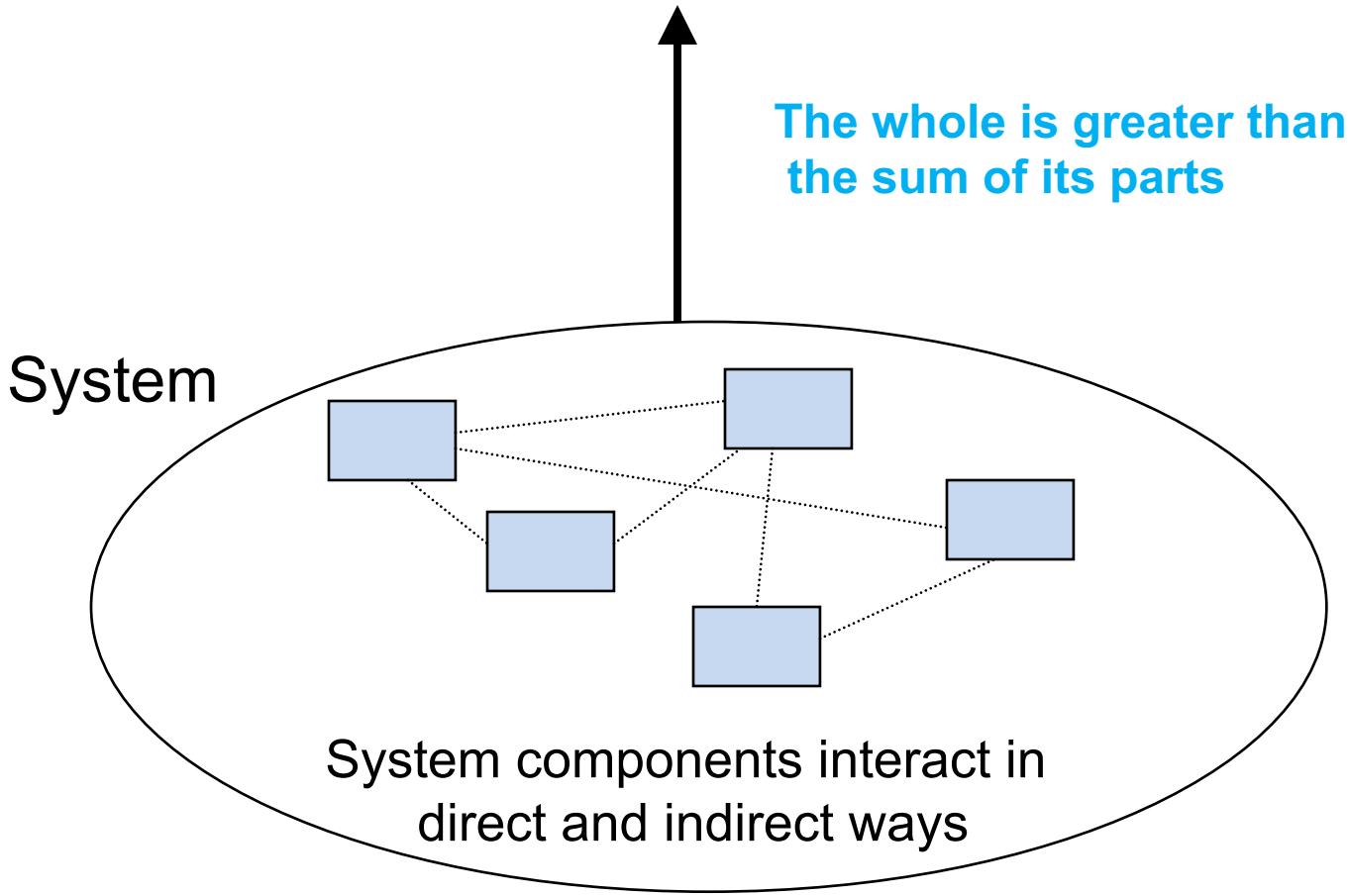
- Focuses on systems taken as a whole, not on parts taken separately
- Emergent properties
  - Some properties can only be treated adequately in their entirety, taking into account all social and technical aspects

    "The whole is greater than the sum of the parts"

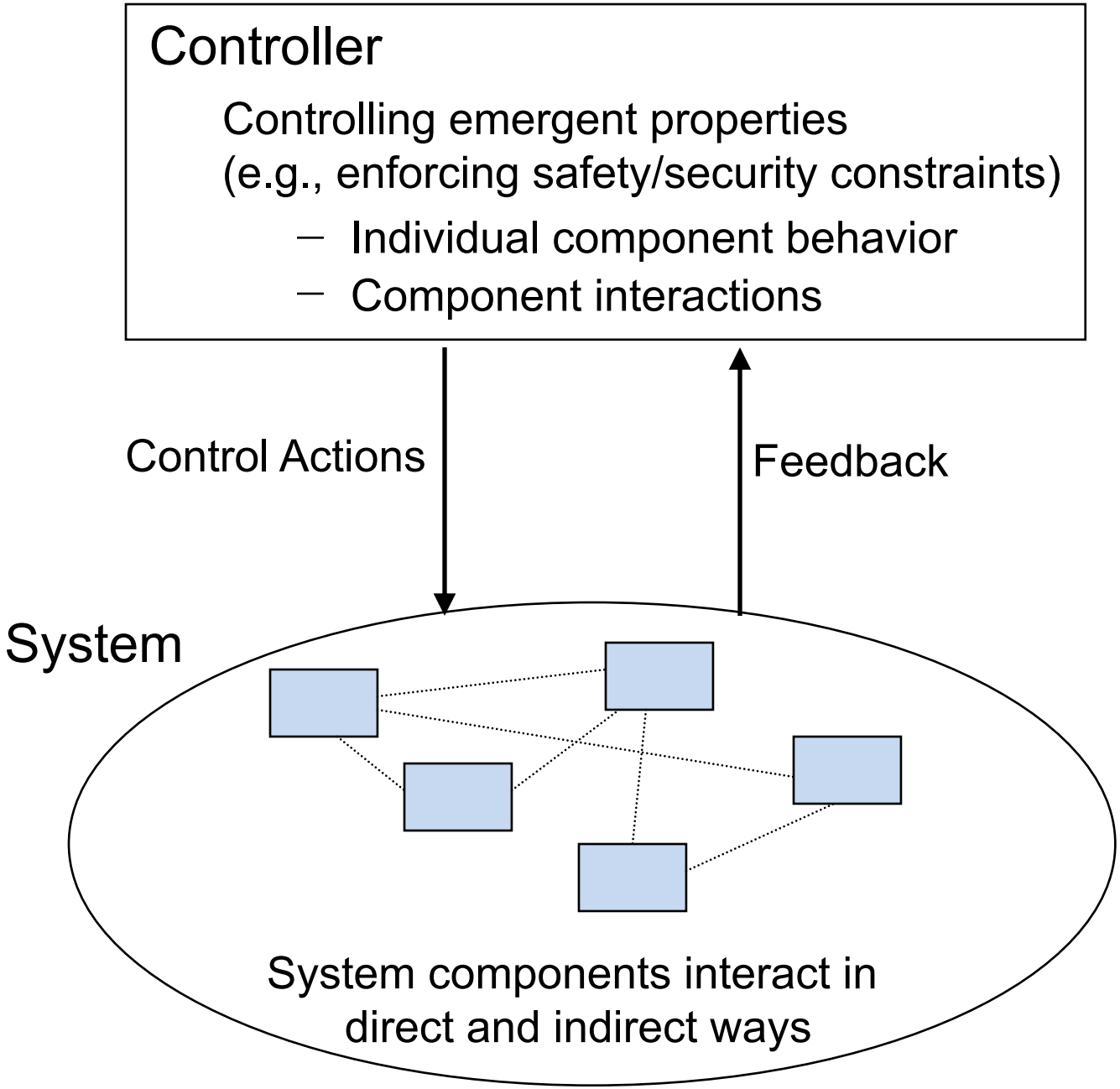
  - These properties arise from relationships among the parts of the system

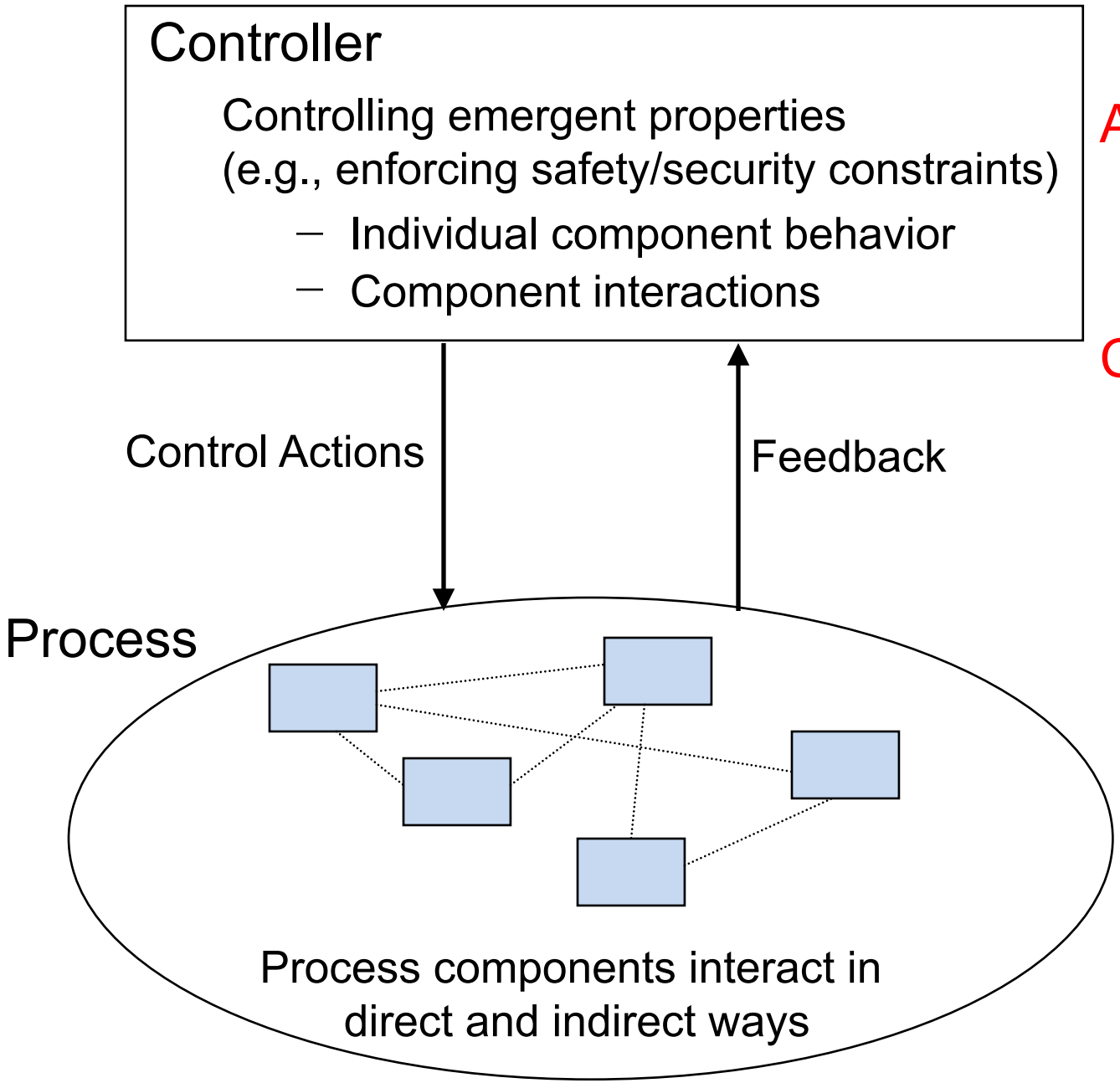
    How they interact and fit together

Emergent properties
(arise from complex interactions)



Safety and security are emergent properties


Controller
Controlling emergent properties
(e.g., enforcing safety/security constraints)
– Individual component behavior
– Component interactions
Control Actions
Feedback
System
System components interact in
direct and indirect ways



Air Traffic Control:
Safety
Throughput

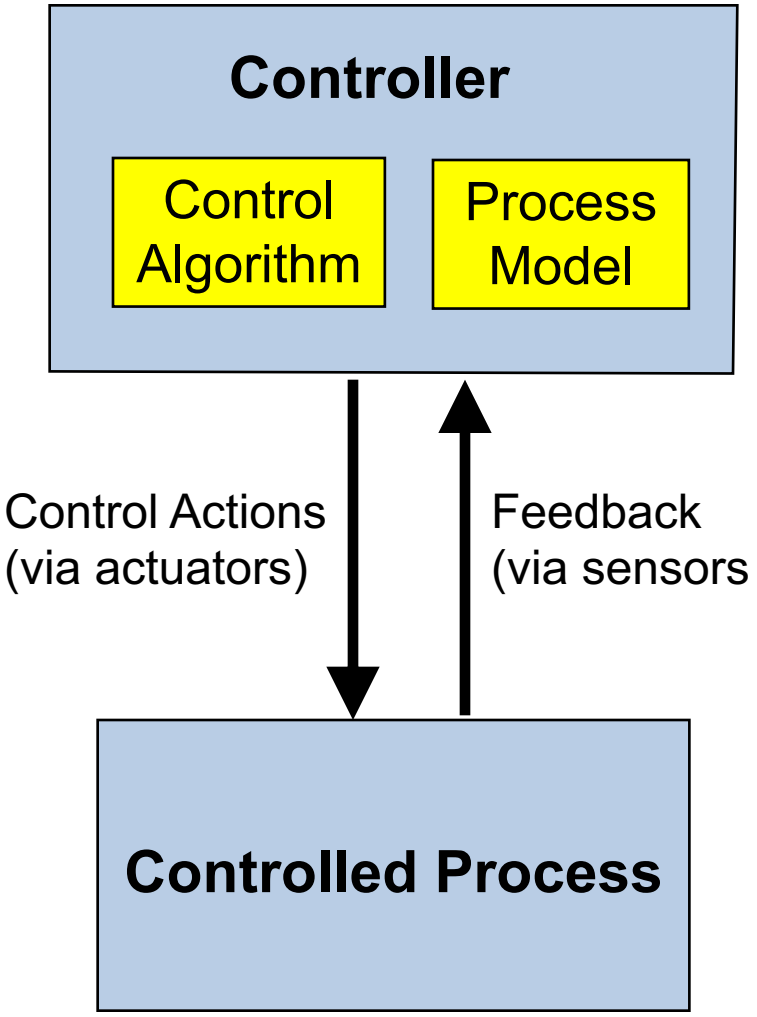
Controlling flow
over internet

Controls/Controllers Enforce Safety/Security Constraints

- Two aircraft/automobiles must not violate minimum separation
- Aircraft must maintain sufficient lift to remain airborne
- Weapons must not target friendly forces
- Toxic chemicals/radiation must not be released from the plant
- Nuclear materials must never get into the wrong hands
- Weapons must never be detonated inadvertently

Note: Functional Security vs. Information Security

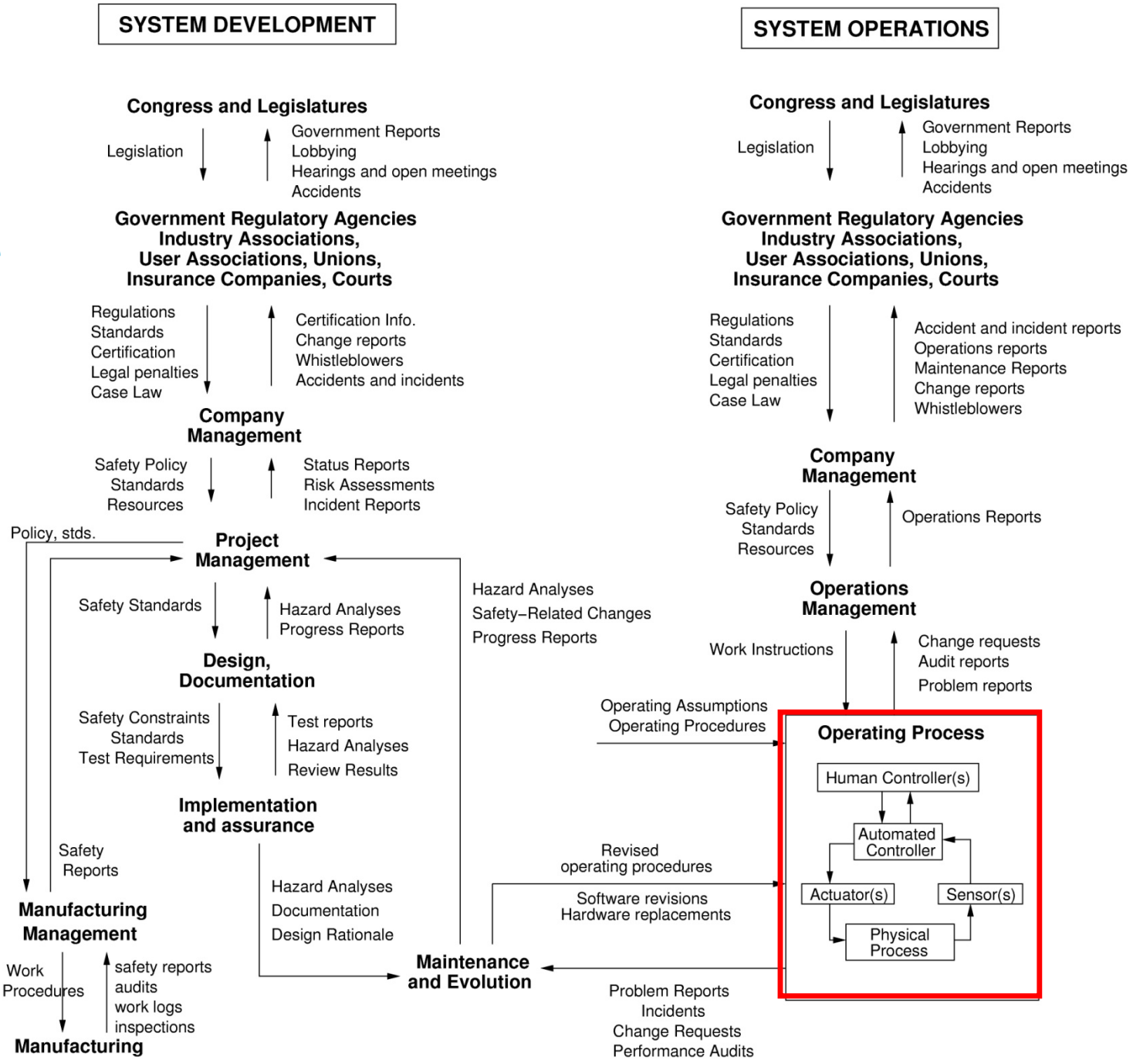
Safety as a Control Problem (vs. Failure Problem)

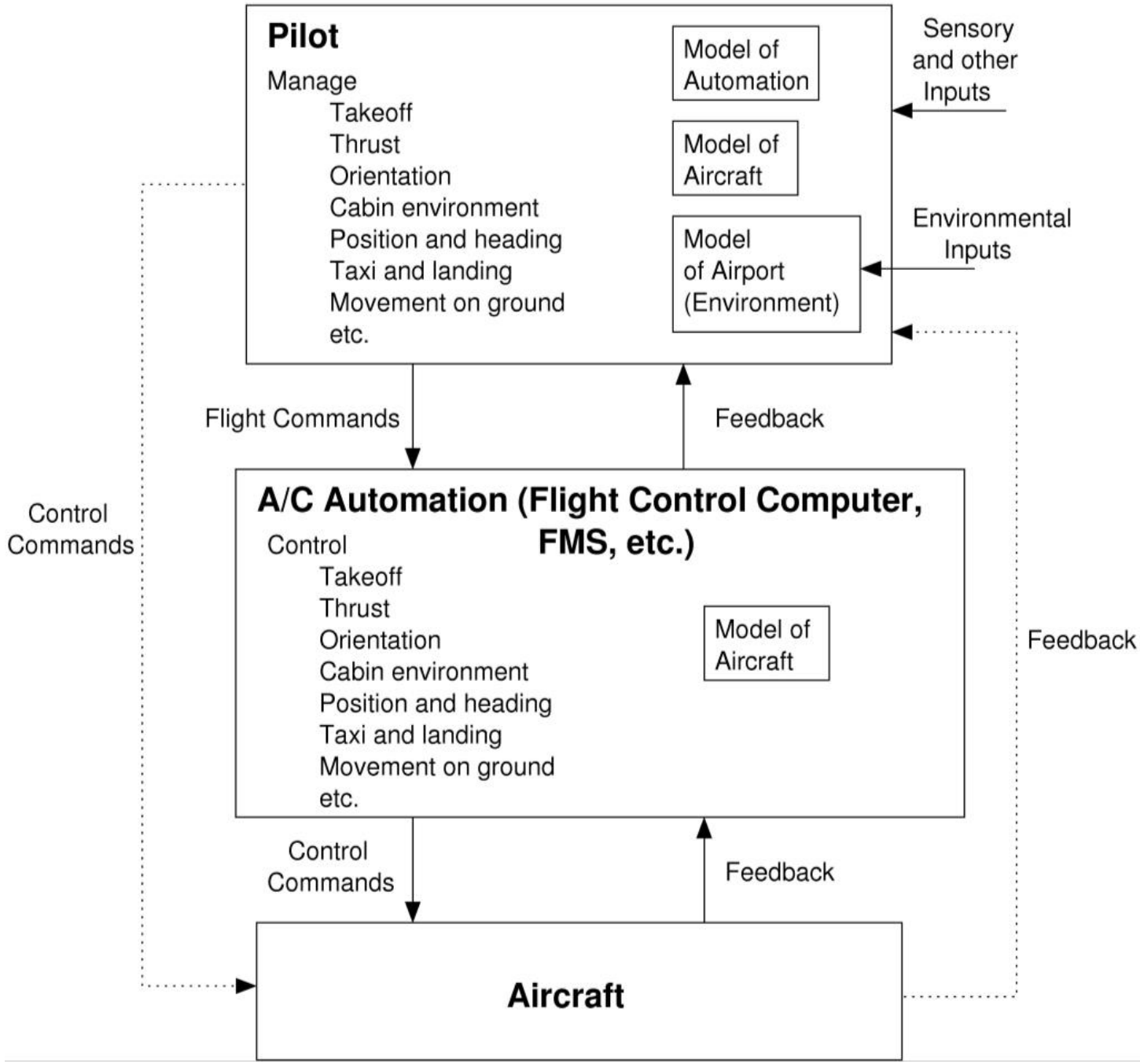


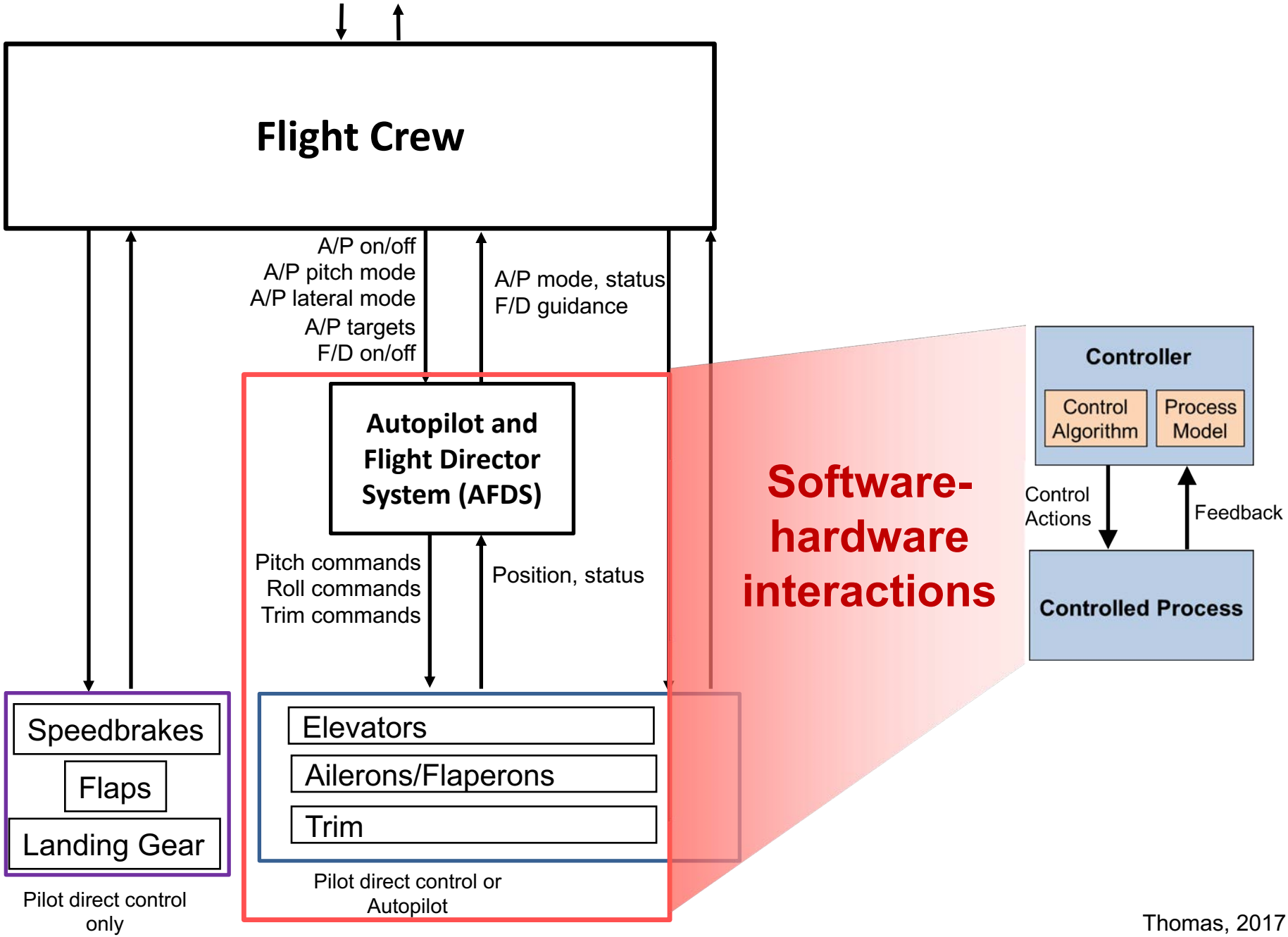
- Controllers use a **process model** to determine control actions
- Software/human related accidents often occur when the process model is incorrect
- Captures software errors, human errors, flawed requirements …

Treat safety as a control problem, not a failure problem

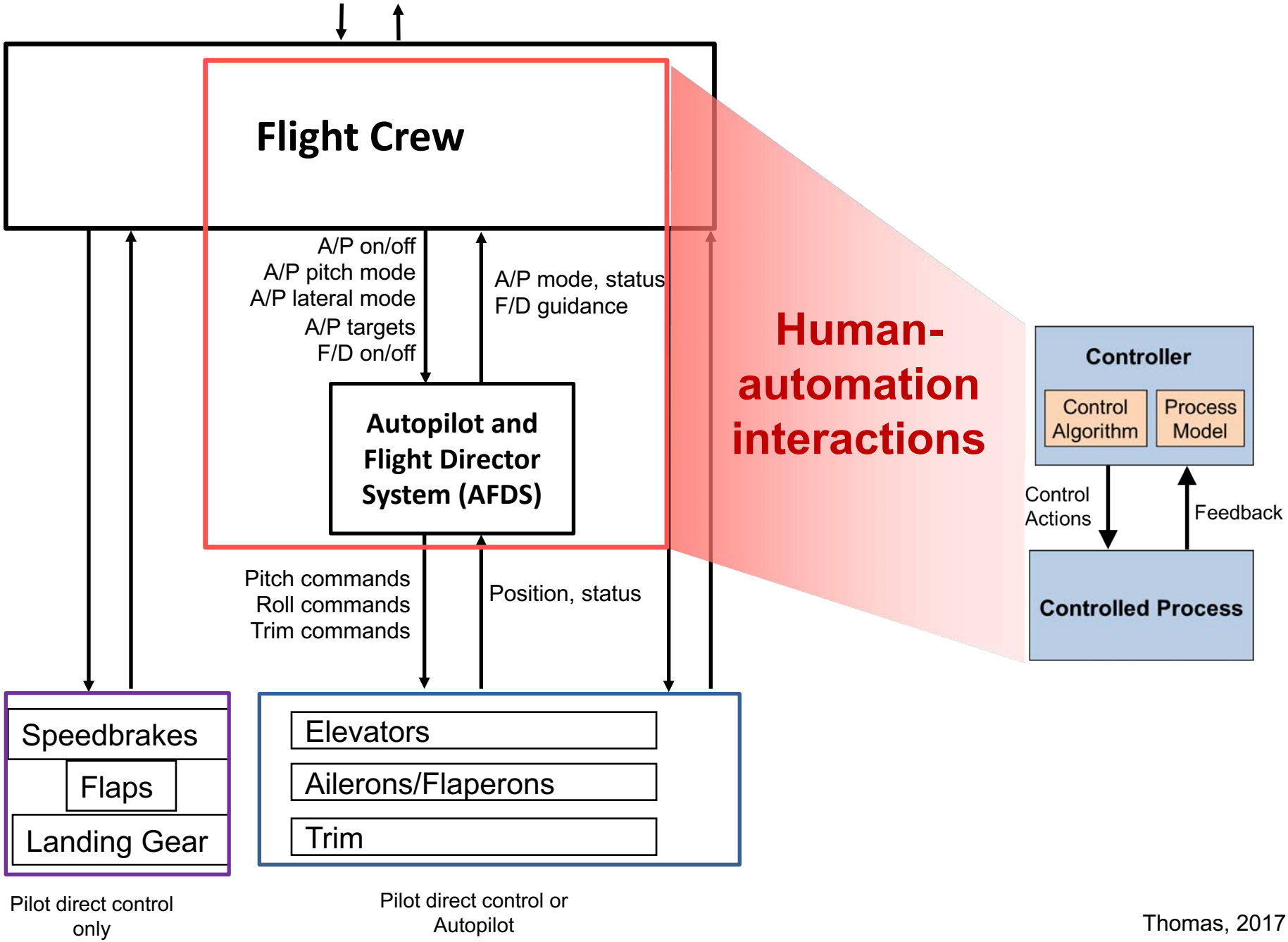
Example Safety Control Structure (SMS)



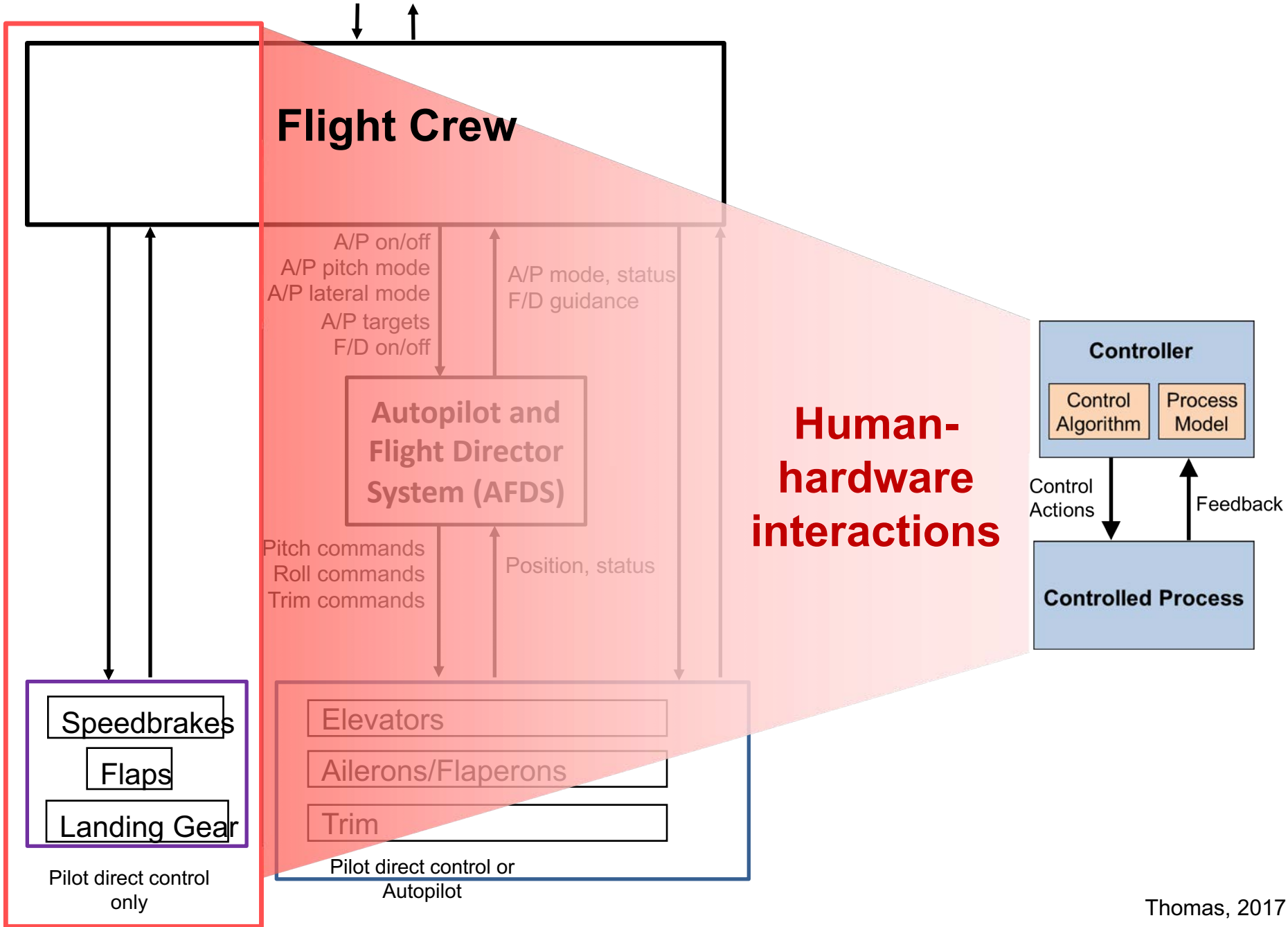

Pilot
Manage
Takeoff
Thrust
Orientation
Cabin environment
Position and heading
Taxi and landing
Movement on ground
etc.
Model of Automation
Model of Aircraft
Model of Airport (Environment)
Sensory and other Inputs
Environmental Inputs
Flight Commands
Feedback
A/C Automation (Flight Control Computer, FMS, etc.)
Control
Takeoff
Thrust
Orientation
Cabin environment
Position and heading
Taxi and landing
Movement on ground
etc.
Model of Aircraft
Control Commands
Feedback
Control Commands
Feedback
Aircraft


Flight Crew
A/P on/off
A/P pitch mode
A/P lateral mode
A/P targets
F/D on/off
A/P mode, status
F/D guidance
Autopilot and Flight Director System (AFDS)
Pitch commands
Roll commands
Trim commands
Position, status
Speedbrakes
Flaps
Landing Gear
Elevators
Ailerons/Flaperons
Trim
Pilot direct control only
Pilot direct control or Autopilot
Software-hardware interactions
Controller
Control Algorithm
Process Model
Control Actions
Feedback
Controlled Process

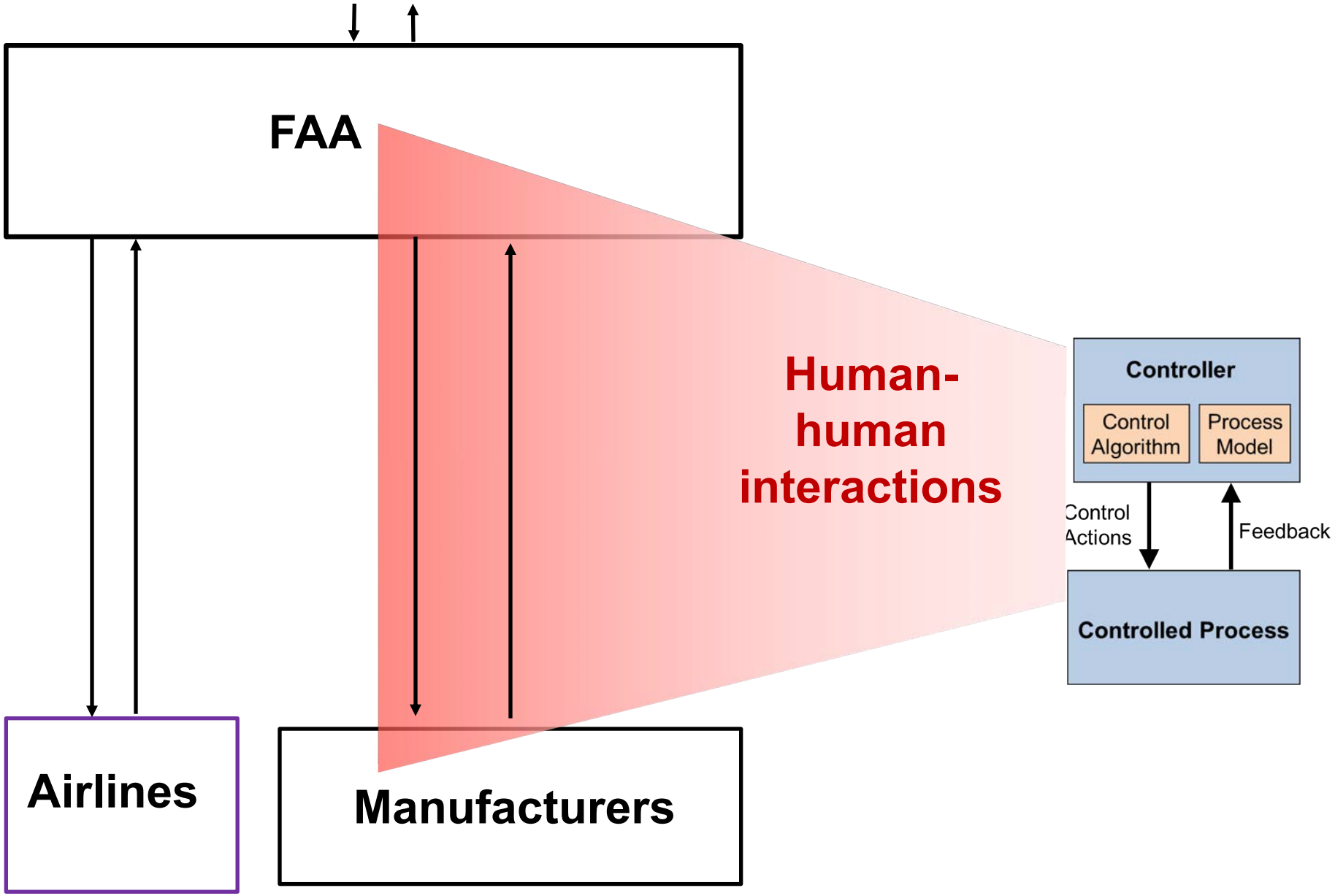
Thomas, 2017


Flight Crew
A/P on/off
A/P pitch mode
A/P lateral mode
A/P targets
F/D on/off
A/P mode, status
F/D guidance
Autopilot and Flight Director System (AFDS)
Pitch commands
Roll commands
Trim commands
Position, status
Human-automation interactions
Controller
Control Algorithm
Process Model
Control Actions
Feedback
Controlled Process
Speedbrakes
Flaps
Landing Gear
Elevators
Ailerons/Flaperons
Trim
Pilot direct control only
Pilot direct control or Autopilot

Thomas, 2017


Flight Crew
A/P on/off
A/P pitch mode
A/P lateral mode
A/P targets
F/D on/off
A/P mode, status
F/D guidance
Autopilot and Flight Director System (AFDS)
Pitch commands
Roll commands
Trim commands
Position, status
Speedbrakes
Flaps
Landing Gear
Pilot direct control only
Elevators
Ailerons/Flaperons
Trim
Pilot direct control or Autopilot
Human-hardware interactions
Controller
Control Algorithm
Process Model
Control Actions
Feedback
Controlled Process

Thomas, 2017


FAA
Human-
human
interactions
Controller
Control
Algorithm
Process
Model
Control
Actions
Feedback
Controlled Process
Airlines
Manufacturers

Thomas, 2017

STAMP
(System-Theoretic Accident Model and Processes)

- A new, more powerful accident/loss causality model
- Based on systems theory, not reliability theory
- Treats accidents/losses 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

A Broad View of "Control"

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

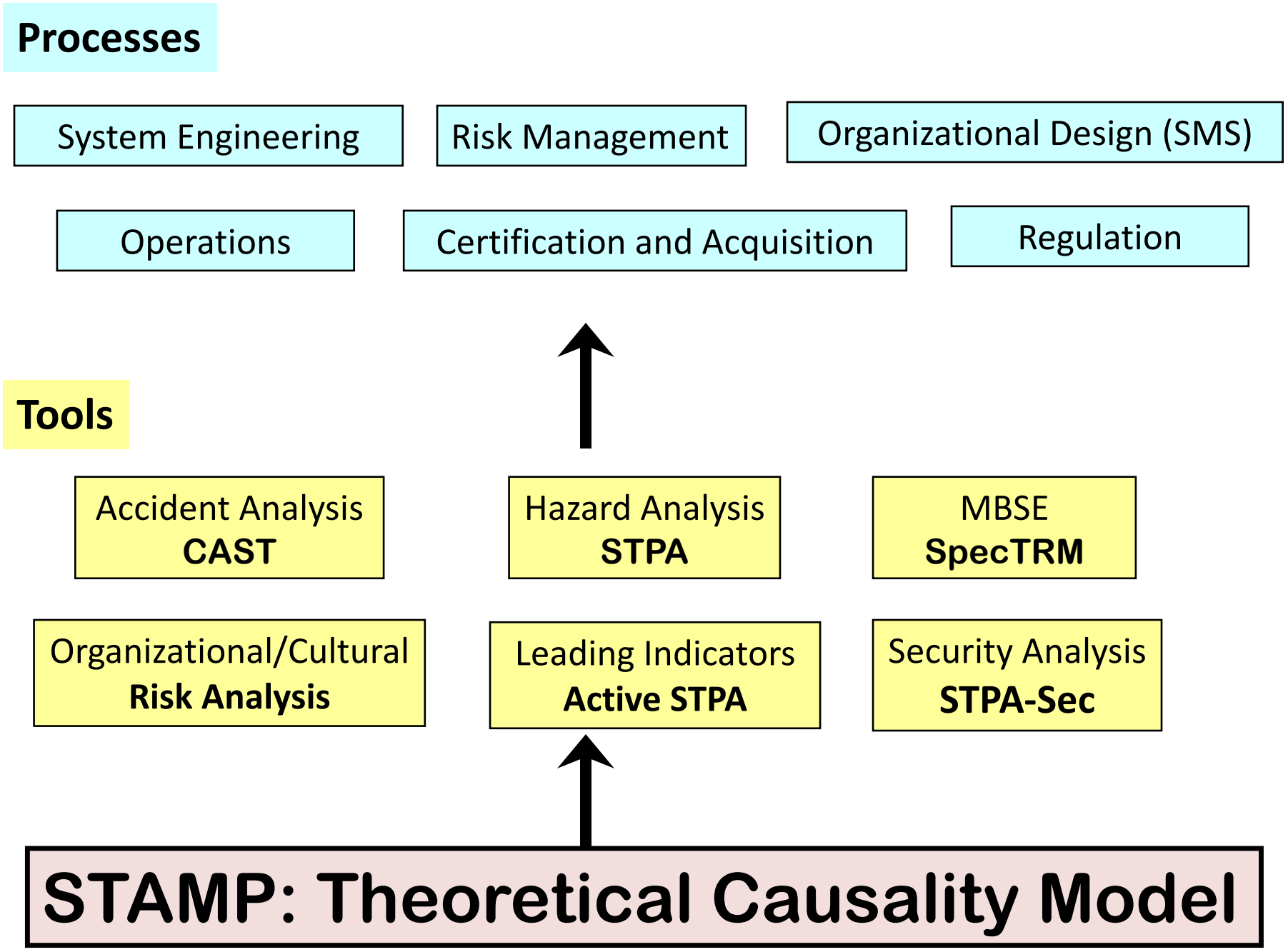
(e.g., redundancy, interlocks, fail-safe design)

or through process

- Manufacturing processes and procedures
- Maintenance processes
- Operations

or through social controls

- Governmental or regulatory
- Culture
- Insurance
- Law and the courts
- Individual self-interest (incentive structure)


Processes
System Engineering
Risk Management
Organizational Design (SMS)
Operations
Certification and Acquisition
Regulation
Tools
Accident Analysis
CAST
Hazard Analysis
STPA
MBSE
SpecTRM
Organizational/Cultural
Risk Analysis
Leading Indicators
Active STPA
Security Analysis
STPA-Sec
STAMP: Theoretical Causality Model

Integrated Approach to Safety and Security (Col. Bill Young)

- Safety: prevent losses due to unintentional actions by benevolent actors
- Security: prevent losses due to intentional actions by malevolent actors
- Key difference is intent but usually doesn't matter in prevention
- Common goal: loss prevention
  - Ensure that critical functions and services provided by networks and services are maintained
  - New paradigm for safety will work for security too
    - May have to add new causes, but rest of process is the same
  - A top-down, system engineering approach to designing safety and security into systems

Example: Stuxnet

- Loss: Damage to reactor (in this case centrifuges)
- Hazard/Vulnerability: Centrifuges are damaged by spinning too fast
- Constraint to be Enforced: Centrifuges must never spin above maximum speed
- Hazardous control action: Issuing *increase speed* command when already spinning at maximum speed
- One potential causal scenario:
  - Incorrect process model: thinks spinning at less than maximum speed
    - Could be inadvertent or deliberate
- Potential controls:
  - Mechanical limiters (interlock), Analog RPM gauge

Focus on preventing hazardous state
(not keeping intruders out)

Is it Practical?

- STPA has been or is being used in a large variety of industries
  - Automobiles (>80% use)
  - Aircraft and Spacecraft (extensive use and growing)
  - Defense systems
  - UAVs (RPAs)
  - Air Traffic Control
  - Medical Devices and Hospital Safety
  - Chemical plants
  - Oil and Gas
  - Nuclear and Electric Power
  - Robotic Manufacturing / Workplace Safety
  - Finance
- New international standards (autos) created, in development, or STPA already satisfies (MIL-STD-882)

Evaluations and Estimates of ROI

- Hundreds of evaluations and comparison with traditional approaches used now
  - Controlled scientific and empirical (in industry)
  - All show STPA is better (identifies more critical requirements or design flaws)
  - All (that measured) show STPA requires orders of magnitude fewer resources than traditional techniques
- ROI estimates only beginning but one large defense industry contractor claims they are seeing 15-20% return on investment when using STPA

Ballistic Missile Defense System (MDA)

- Hazard was inadvertent launch
- Analyzed right before deployment and field testing (so done late)
  - 2 people, 5 months (unfamiliar with system)
  - Found so many paths to inadvertent launch that deployment delayed six months
- One of first uses of STPA on a real defense system (2005)

Sea-based sensors on the Aegis platform, upgraded early warning radars (UEWR), the Cobra Dane Upgrade (CDU), Ground-based Midcourse Defense (GMD) Fire Control and Communications (GFC/C), a Command and Control Battle Management and Communications (C2BMC) Element, and Ground-based interceptors (GBI). Future block upgrades were originally planned to introduce additional Elements into the BMDS, including Airborne Laser (ABL) and Terminal High Altitude Area Defense (THAAD).

Example Hazard Scenarios Found

- Missing software requirements, for example:
  - Operator could input a legal (but unanticipated) instruction at same time that radars detect a potential (but not dangerous) threat
  - Could lead to software issuing an instruction to enable firing an interceptor at a non-threat
- Timing conditions that could lead to incorrectly launching an interceptor
- Situations in which simulator data could be taken as real data

Navy Escort Vessels (Lt. Blake Abrecht)

- Dynamic positioning system
- Ran into each other twice during test
- Performed a CAST analysis (on two incidents) and STPA on system as a whole
- STPA found scenarios not found by MIL-STD-882 analysis (fault trees and FMEA)
- Did not implement our findings: “We’ve used PRA for 40 years and it works just fine”
- Put into operation and within 2 months ran into a submarine
- Scenario was one we had found

EPRI Evaluation

- Same design of a nuclear power plant safety system provided to everyone
- Independent and expert teams did: FTA, ETA, FMEA, HAZOP, etc. and we did STPA (two students, two weeks)
- After submitting final analyses, teams were told that there had been a very serious event in plant with that design
- Only STPA found the scenario that had occurred

New EPRI Study

- Learnability (how much time before can find serious problems)
- Found serious design errors in 2-day beginning class

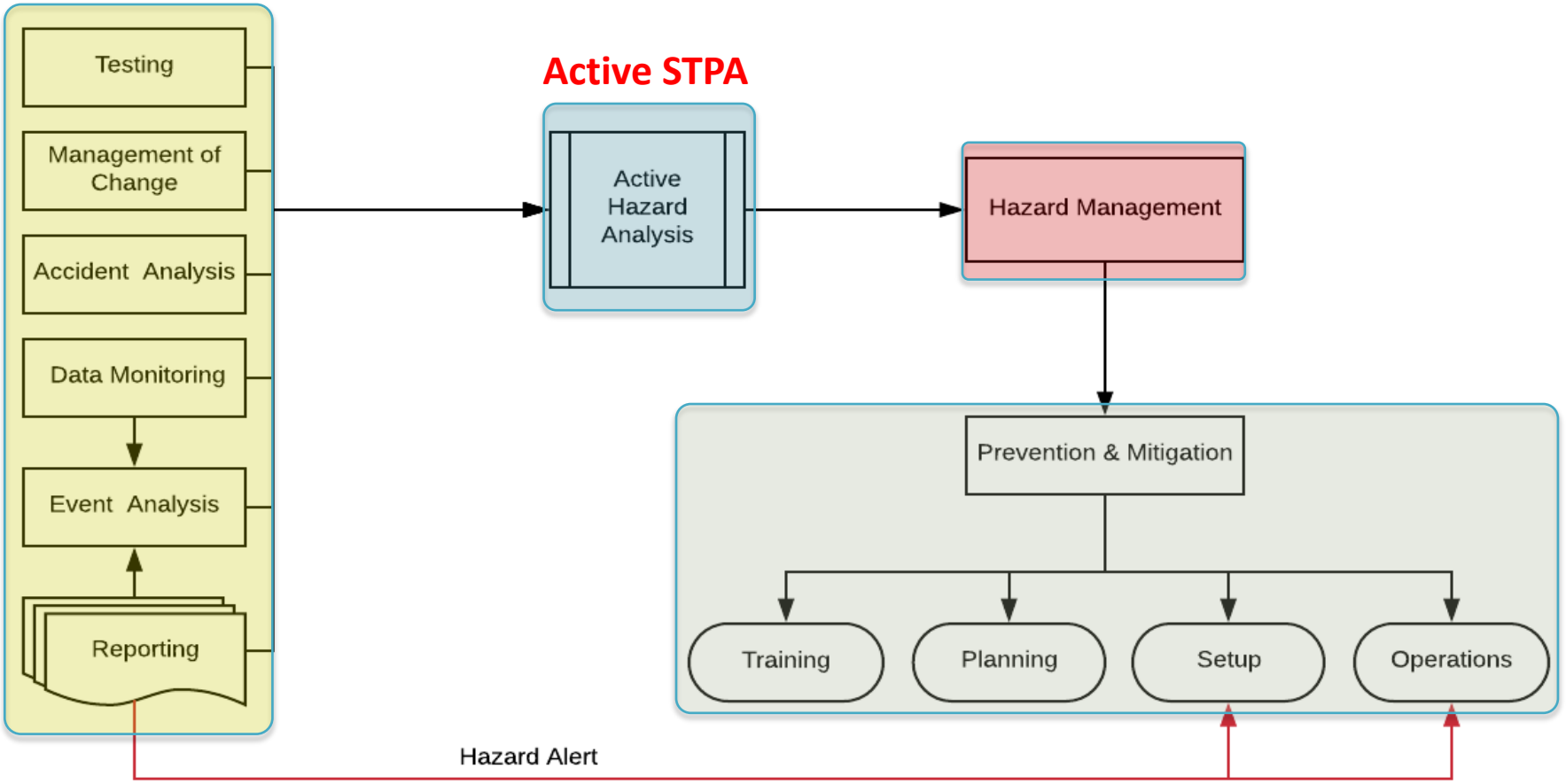
Improving the Standard Risk Matrix

- Use STPA to get better estimates of likelihood

RISK ASSESSMENT MATRIX				
SEVERITY / PROBABILITY	Catastrophic (1)	Critical (2)	Marginal (3)	Negligible (4)
Frequent (A)	High	High	Serious	Medium
Probable (B)	High	High	Serious	Medium
Occasional (C)	High	Serious	Medium	Low
Remote (D)	Serious	Medium	Medium	Low
Improbable (E)	Medium	Medium	Medium	Low
Eliminated (F)	Eliminated			

Risk Management During Operations and Leading Indicators (Lt. Col. Diogo Castilho)

- Systems and their environments are not static
- Goal is to detect when risk is increasing (leading indicators)



Some Other Uses

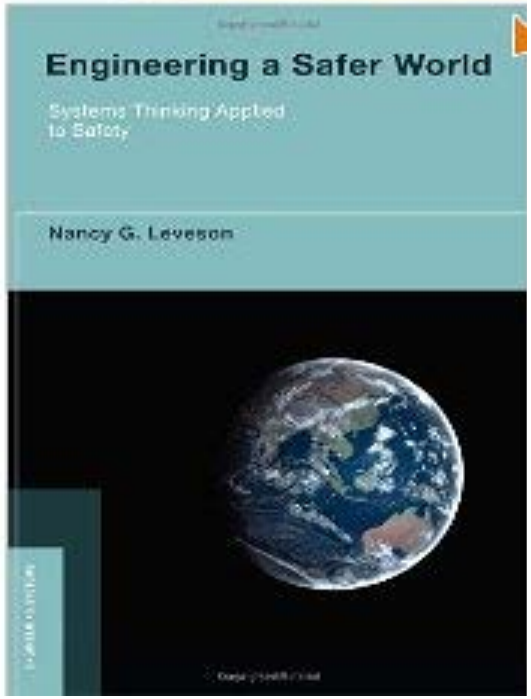
- Organizational risk analysis
- Management system design
- Accounting audit systems
- Workplace safety
- ???

Summary: A Systems Approach to Safety and Security

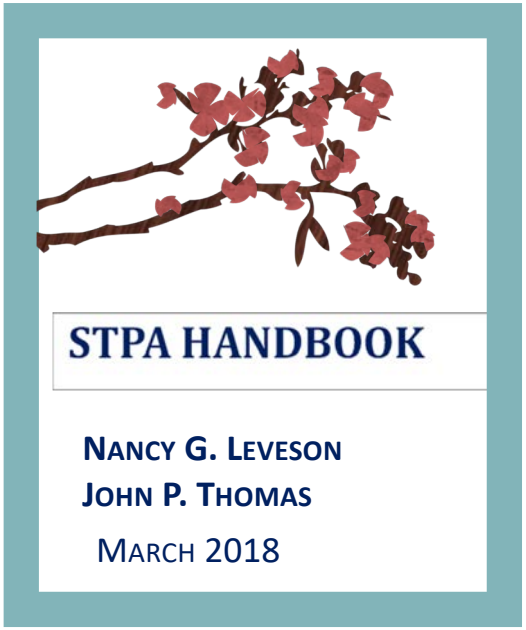
- Emphasizes building in safety rather than measuring it or assuring it
- Looks at system as a whole, not just components (a top-down holistic approach)
- Takes a larger view of causes than just failures
  - Accidents today are not just caused by component failures
  - Includes software and requirements flaws, sophisticated human behavior, design flaws, etc.
- Goal is to use modeling and analysis to design and operate the system to be safe/secure, not to predict the likelihood of a loss or provide after the fact assurance.

More Information

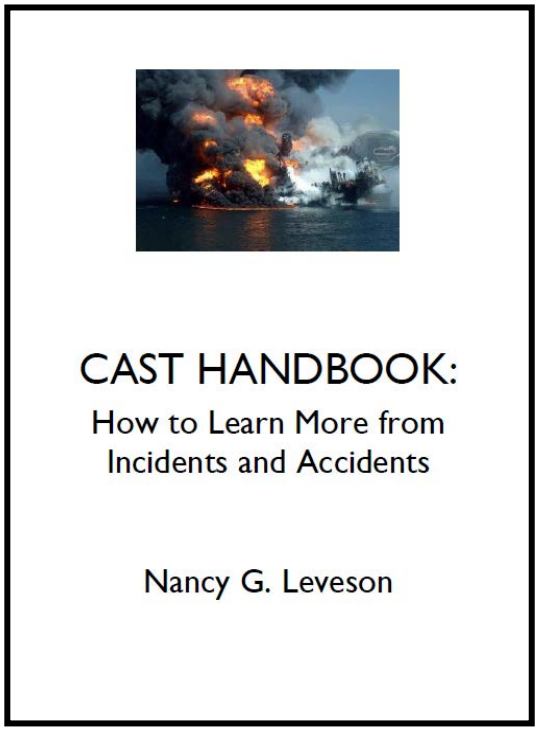
- http://psas.scripts.mit.edu (papers, presentations from conferences, tutorial slides, examples, etc.)



Free download:
http://mitpress.mit.edu/books/engineering-safer-world



http://psas.scripts.mit.edu
(34,000+ downloads in last year
4000+ downloads of Japanese version)
Chinese version (Korean coming)



Free download:
http://sunnyday.mit.edu/CAST-Handbook.pdf