

A True Sociotechnical Approach to Safety in Complex Systems

Prof. Nancy G. Leveson

Aeronautics and Astronautics
MIT



General Definition of “Safety”

- Accident = Mishap = Loss: Any undesired and unplanned event that results in a loss
  - Including loss of human life or injury, property damage, environmental pollution, mission loss, negative business impact (damage to reputation, etc.), product launch delay, legal entanglements, etc.
  - Includes inadvertent and intentional losses (security)
- System goals vs. constraints (limits on how can achieve the goals)
- Safety: Absence of losses

The Problem

- Need a “true” sociotechnical approach, not an empty term
  - Humans do not work in a vacuum
    - To understand and change human behavior, need to look at “system” in which they are working
    - Human error is a symptom of a system that needs to be redesigned
  - To get there, will have to change the way we think and to create new models and holistic system tools
    - The tools we have will not get us there
    - And most new tools are simply variations of what we have already
    - Need a paradigm change

What do we need to get there?

- An enhanced causality model
- New collaborative tools that allow people with different backgrounds to analyze systems together
- People willing to learn something new (perhaps the hardest)

What do we need to get there?

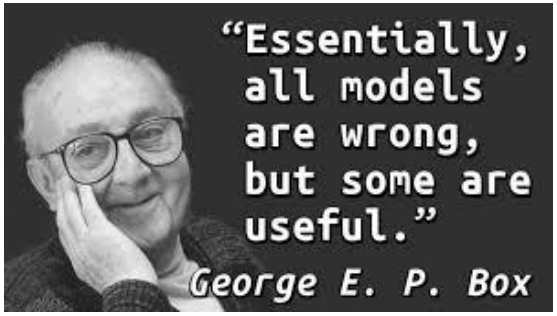
- **An enhanced causality model**

- New collaborative tools that allow people with different backgrounds to analyze systems together

- People willing to learn something new (perhaps the hardest)

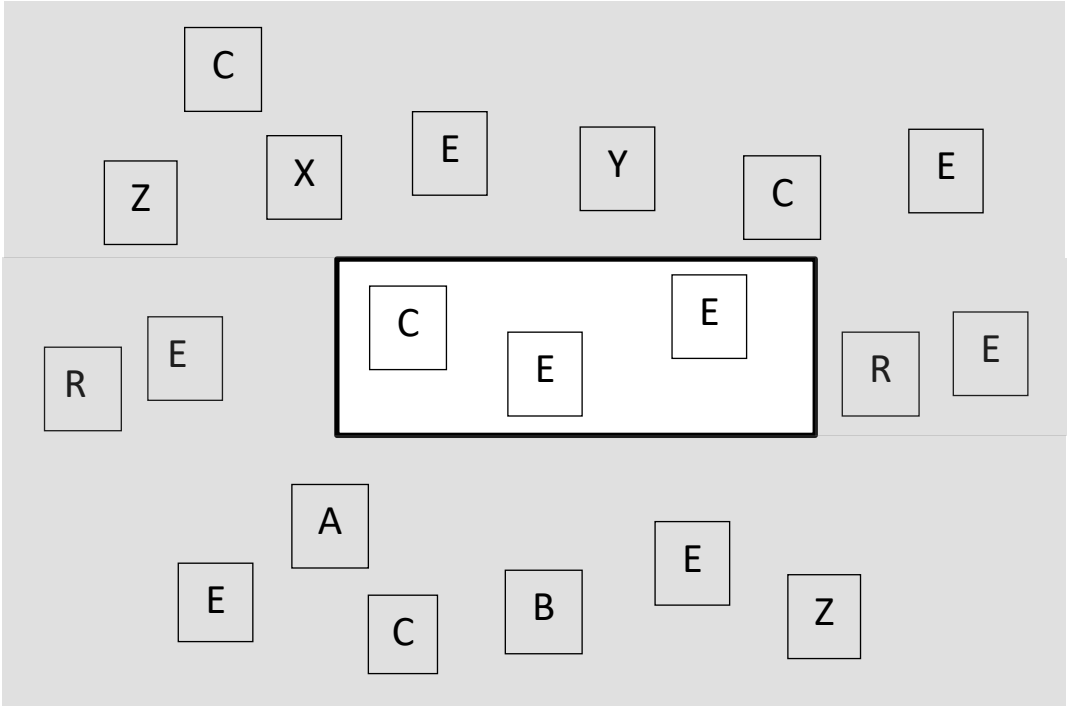
What is a Causality Model?

- Explain how things work and help predict how they will behave in the future
- No right or wrong model, only comparative effectiveness and usefulness



- Models help us deal with a messy world

Models filter out “irrelevant” information (for problem being solved)

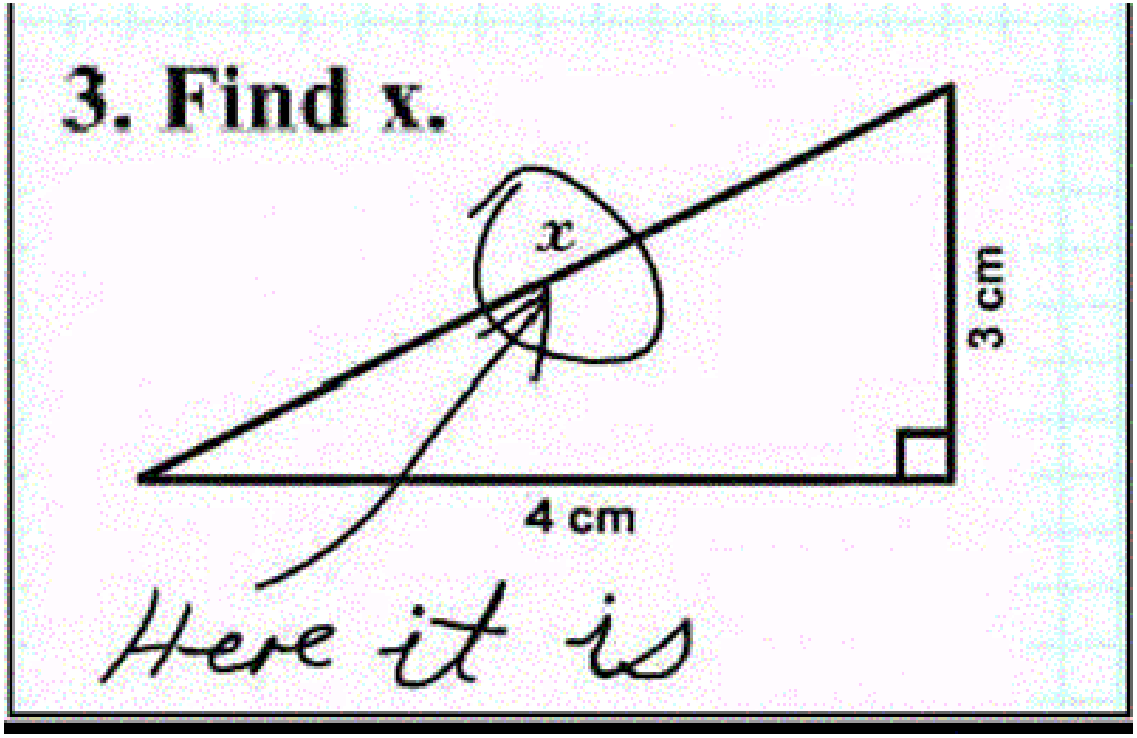


But what is “irrelevant” ?

How do we keep from filtering out important information?

We want simple answers to complex questions.

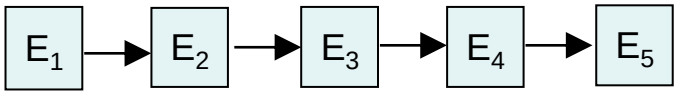
And simple models



So we get simple (but not useful) answers

Chain of (Failure) Events (COE) Model

- Assumes linear causality is enough to understand the world



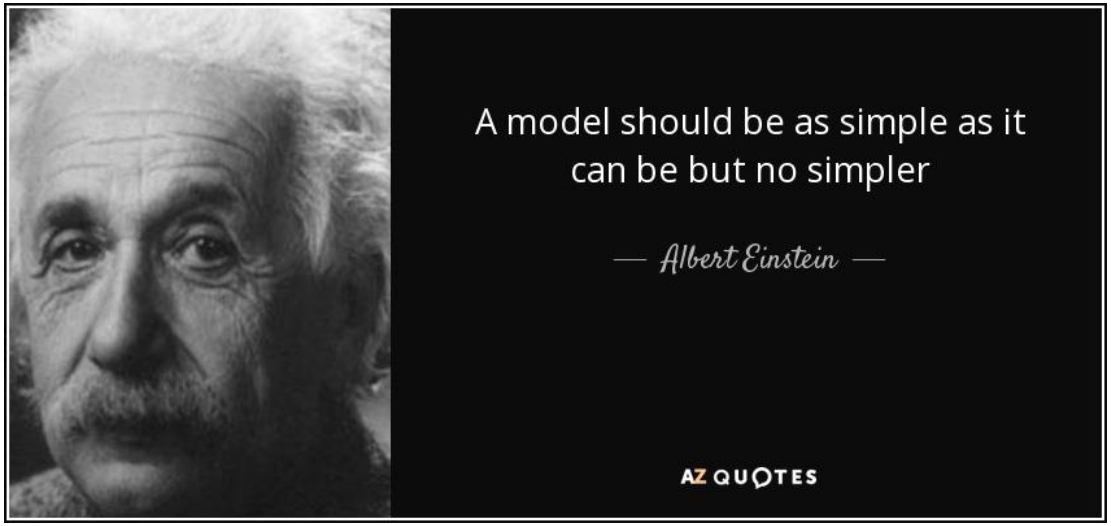
→ means “is necessary and sufficient for”

E_2 happens if and only if E_1 did

- One event is root or probable cause of final loss event
- Root and contributory causes are assumed to be in event chain

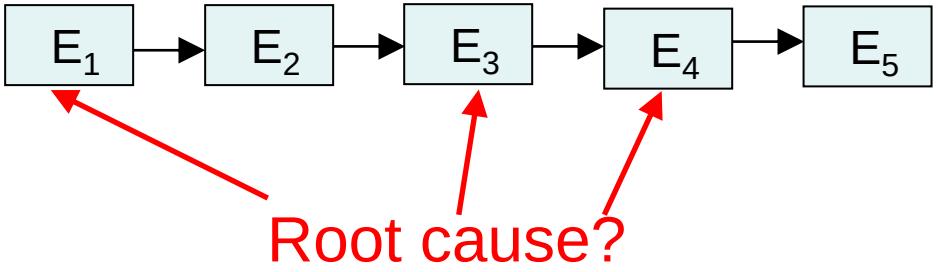
Chain of Events Model of Causality

- The chain of events model is very simple. But is it still useful?



- Does it leave out important causal factors in today's world?

Selection of a "Root Cause" is Arbitrary



- We like the concept of a "root cause"
  - Usually focus on the operator or on physical failures
  - Ignore system-related, management factors (not in the events)
  - What "event" is involved in design of aircraft, design of pilot-vehicle interface, competitive or productivity pressures?
- "Root Cause Seduction" (John Carroll)
  - We want a root cause so we make up a convenient one. Why?
  - Provides an illusion of control
  - So fix symptoms but not process that led to those symptoms

Focus on Pilot/Operator Error ("Failures")

- Pilots/operators almost always in COE for an accident
  - So can always select something they did as the root cause
  - After a while, becomes established that they cause most accidents
  - But human behavior always affected by the context in which it occurs
    - We are designing systems in which human error inevitable
    - But blame the human, not the design
  - Need to understand
    - WHY pilots/operators behaved the way they did
    - Reasons behind why the events occurred

Focus on Identifying a Root or Probable Cause

- May be used to deflect attention from powerful interests.
  - What is declared to be root cause is arbitrary so want to direct attention to someone else.

  - Easy to accomplish when only direct or simple relationships included in chain
  - Sometimes argue that because not everyone made a mistake when presented with same circumstances, those circumstances cannot be the cause.
    - Other pilots flew 737 MAX before crashes and they overcame design flaws so design flaws cannot be "cause" of the accident

Focus on Identifying a Root or Probable Cause

- The cause of all accidents is not the events but why the events occurred
- B737 MAX

  Quote from Muilenberg (CEO of Boeing):

  - "Accidents always involve a chain of events"
  - "Pilots were in chain of events as was MCAS"
  - "MCAS added to workload of pilots"
  - "We can break chain of events that led to both crashes by developing a software fix that would limit the potency of that stabilization system"

- Is that really the "root" cause of the B737 MAX accidents?
- Are we missing deeper issues --- why the events occurred ---that then are never eliminated?

Focus on Identifying a Root or Probable Cause

- While software needs to be fixed, are there not deeper causes that also were involved?
  - Impact of competitive pressures with Airbus A320neo on Boeing management decision making?
  - Was lack of redundancy in AOA sensor simply a random mistake of a design engineer?
  - What was the impact of certification procedures?
  - Inadequate resources of FAA?
  - Changes in regulatory policies and procedures that changed over time to give Boeing more autonomy?
  - Role of system engineering processes and procedures?

Disclaimer: Accident reports are not available. I am speculating like everyone else.

Chain of Events Model is Too Simple

- Implies only direct causes exist and are important, not indirect ones

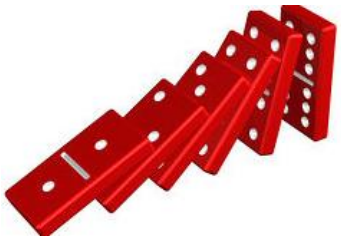
  For example, can argue that smoking does not “cause” lung cancer.

  - Not everyone who smokes gets lung cancer
  - Not everyone who gets lung cancer smokes
  - Ergo, using linear causality arguments, smoking does not cause lung cancer

- Misses systemic problems (sociotechnical problems)

Various Incarnations of COE Model

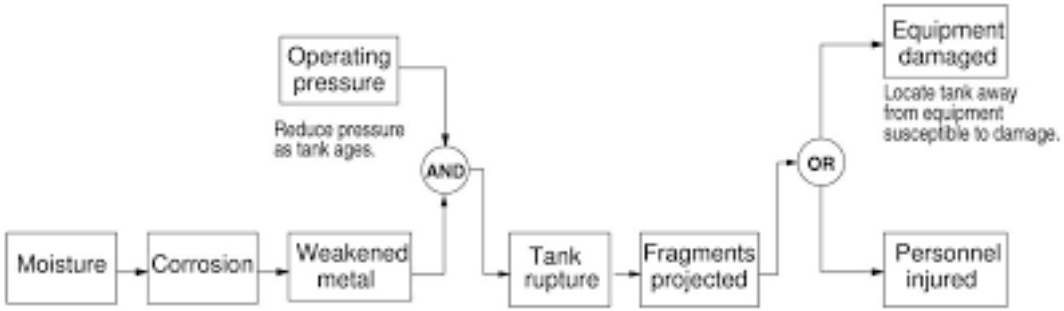
- All use different real world analogies for same thing
  - Bow ties[1], dominoes, cheese slices, etc.
  - Different names and graphical notations for same thing

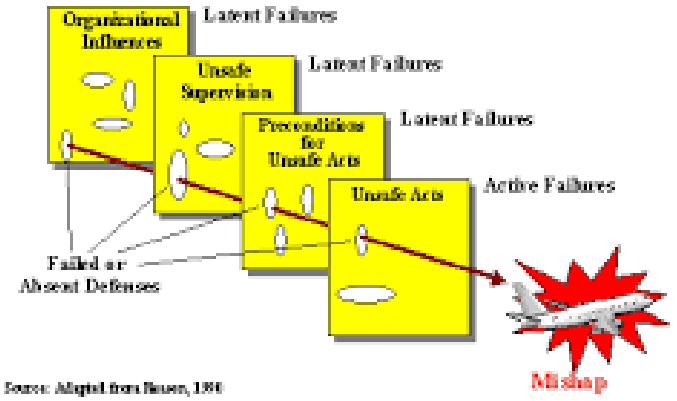


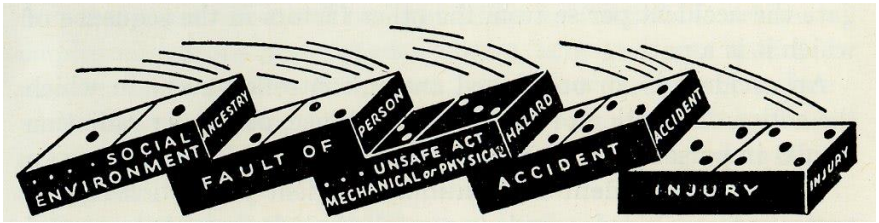
- Easily understood but is the COE model too simple for today's increasingly complex world (technical and social)?
  - Question is <u>not</u> whether the COE model is right or wrong
  - Question <u>is</u> whether it provides the most useful explanation for the goals of accident causal analysis and prevention.

[1]Nancy Leveson, *Shortcomings of the Bow Tie and Other Safety Tools Based on Linear Causality*, July 2019

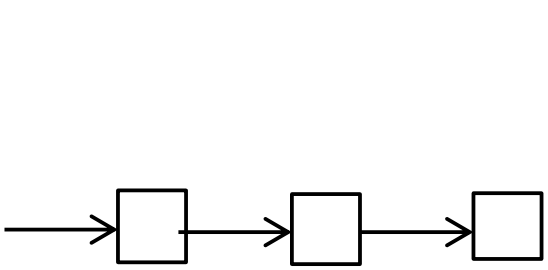
Going Beyond our Current Accident Models



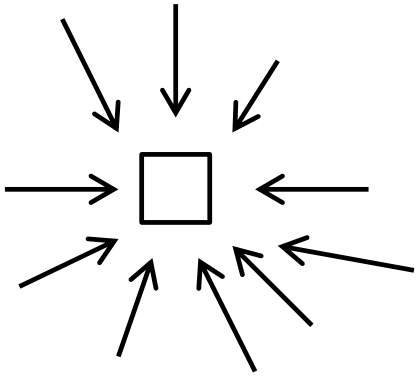




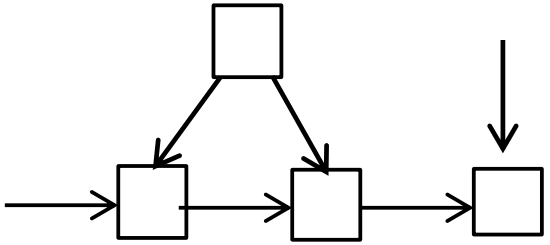
Need to Capture More Types of Causality than Linear

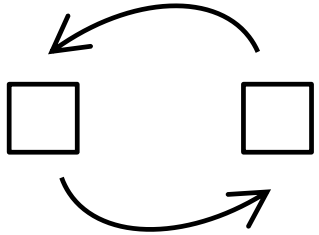


(a)



(b)





Herald of Free Enterprise

Deckhand overslept	Deckhand did not close doors	Captain in hurry to leave	Bosun did not check doors closed	Ferry leaves dock	Ferry capsizes

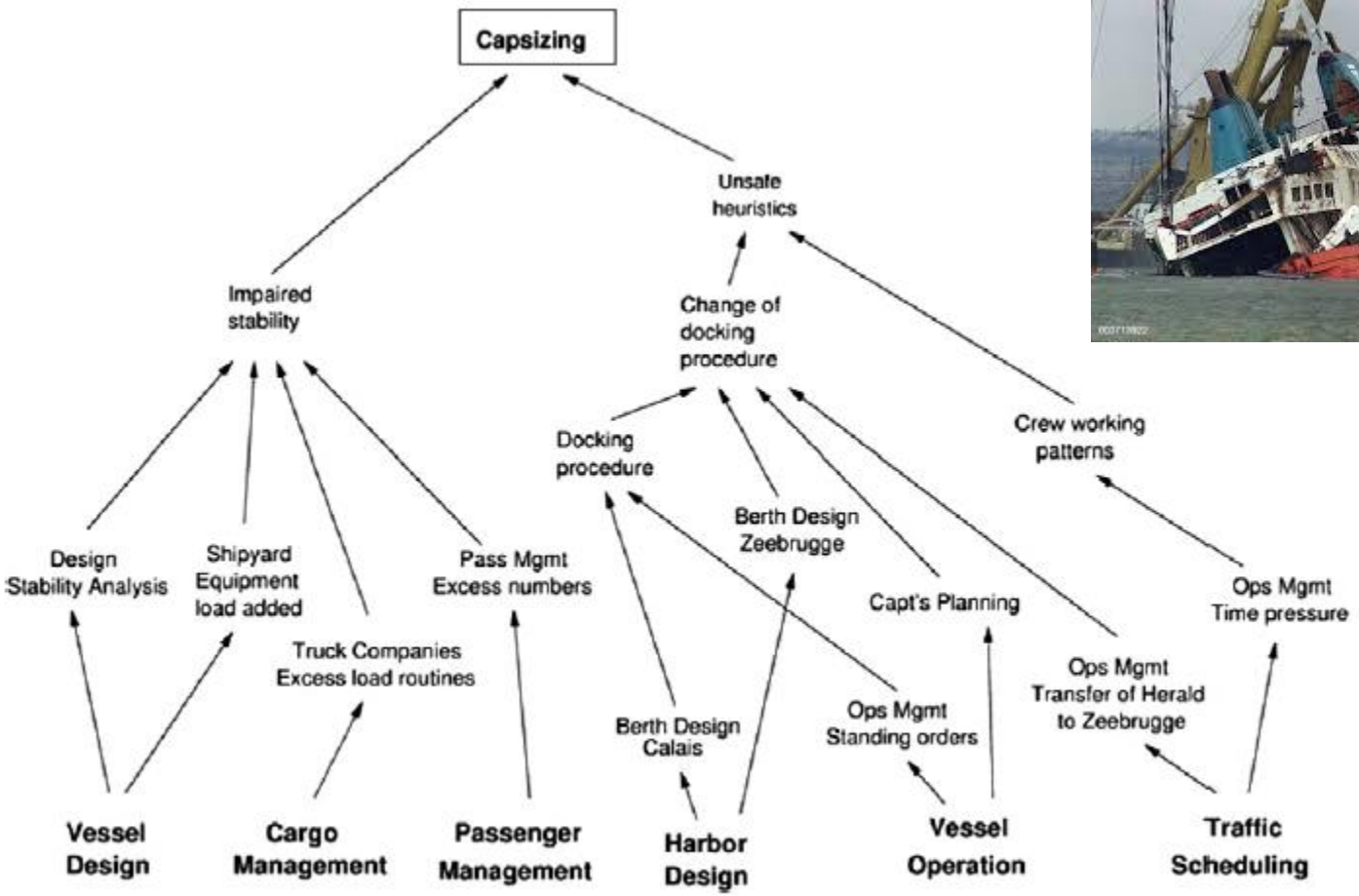


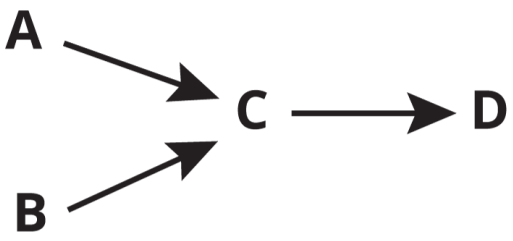
Fig. 1. The complex interactions in the Zeebrugge ferry accident (adapted from Rasmussen, (1997)).

“Reality is made up of circles, but we see straight lines”

Peter Senge, *Fifth Discipline* (p. 73)

Event Oriented Thinking

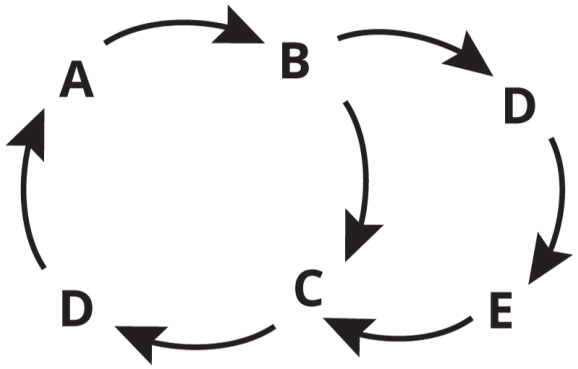
Thinks in straight lines



In event oriented thinking everything can be explained by causal chains of events. From this perspective the **root causes** are the events starting the chains of cause and effect, such as A and B.

Systems Thinking

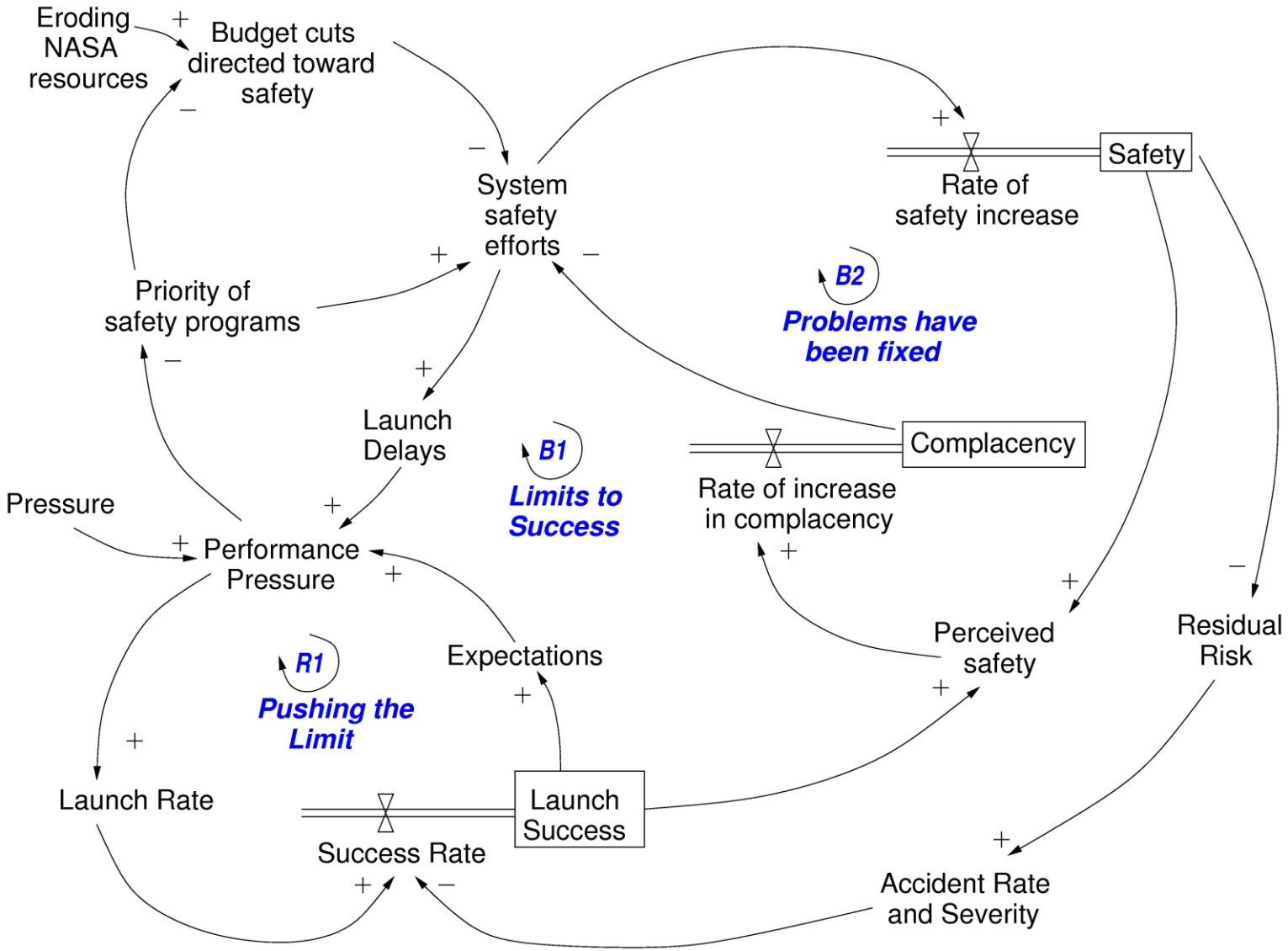
Thinks in loop structure



In systems thinking a system’s behavior emerges from the structure of its feedback loops. **Root causes** are not individual nodes. They are the forces emerging from particular feedback loops.

Created by Thwink.org

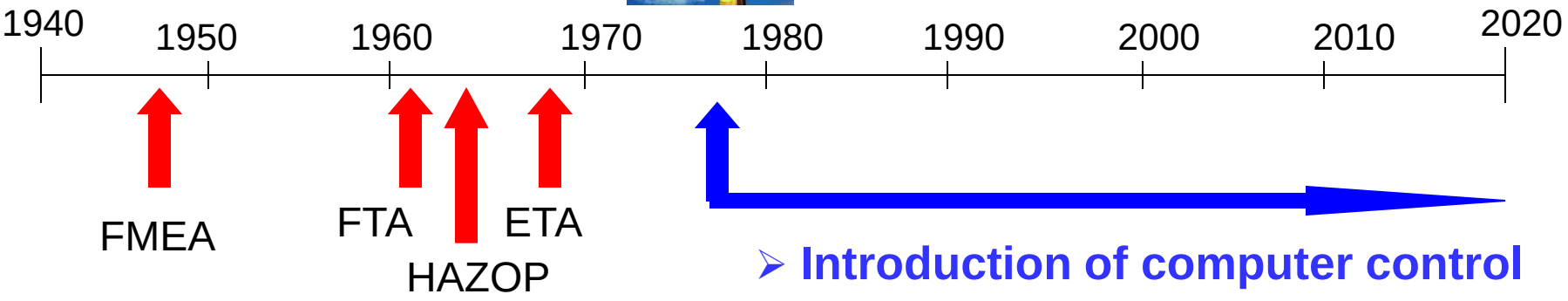
Some Factors in the Columbia Shuttle Loss



What do we need to get there?

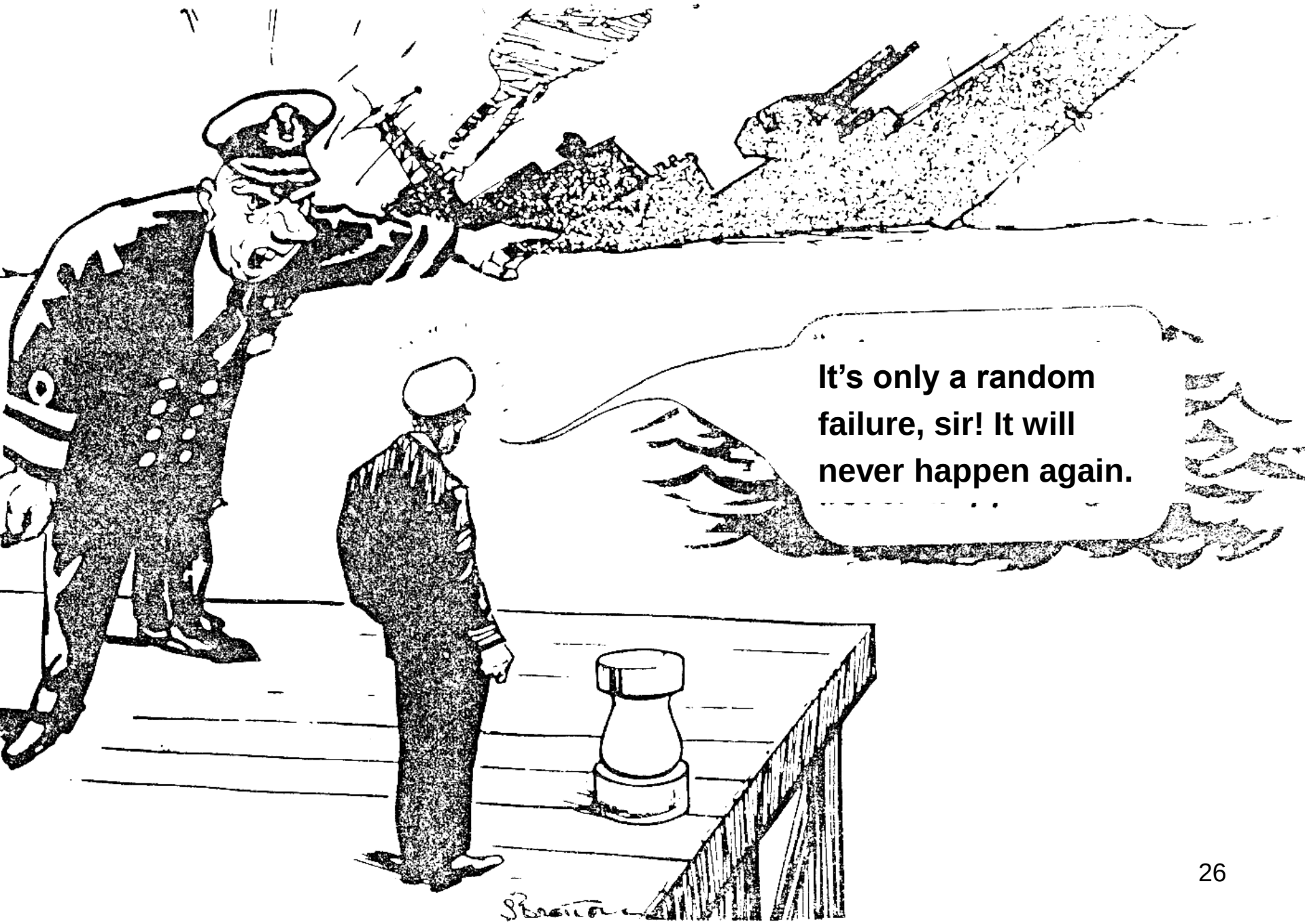
- An enhanced causality model
- **New collaborative tools that allow people with different backgrounds to analyze systems together**
- People willing to learn something new (perhaps the hardest)

Our current tools are all 50-75 years old but our technology is very different today



Assume accidents caused by component failures

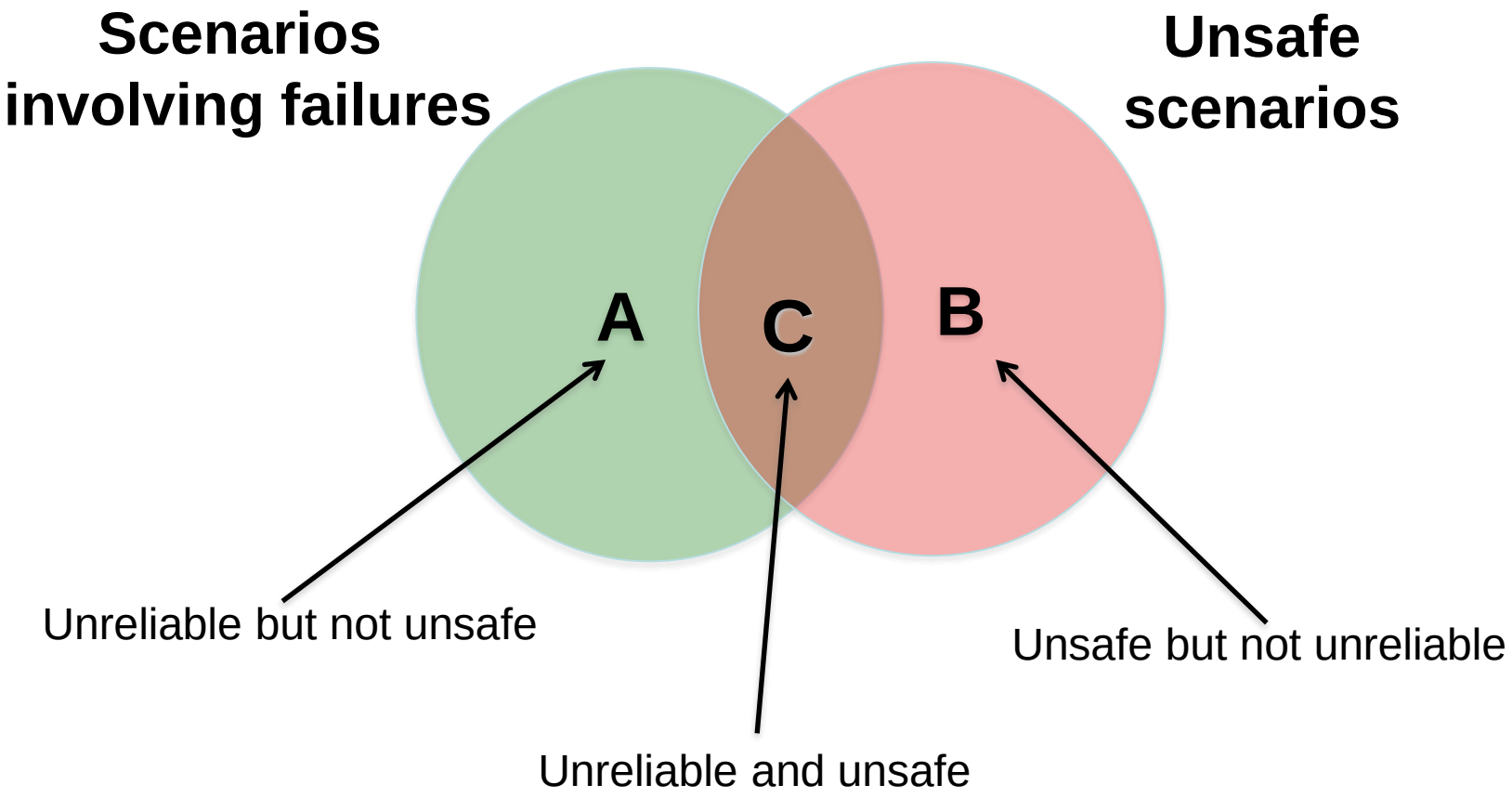
- **Introduction of computer control**
- **Exponential increases in complexity**
- **New technology**
- **Changes in human roles**


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

Warsaw A320 Accident

- Software protects against activating thrust reversers when airborne
- Hydroplaning and other factors made the software think the plane had not landed
- Pilots could not activate the thrust reversers and ran off end of runway into a small hill.

Confusing Safety and Reliability



Preventing Component or Functional Failures is Not Enough

Another Accident Involving Thrust Reversers

- Tu-204, Moscow, 2012
- Red Wings Airlines Flight 9268
- The soft 1.12g touchdown made runway contact a little later than usual.
- With the crosswind, this meant weight-on-wheels switches did not activate and the thrust-reverse system would not deploy.

Another Accident Involving Thrust Reversers

- Pilots believe the thrust reversers are deploying like they always do. With the limited runway space, they quickly engage high engine power to stop quicker. Instead this accelerated the Tu-204 forwards, eventually colliding with a highway embankment.

Another Accident Involving Thrust Reversers

- Pilots believe the thrust reversers are deploying like they always do. With the limited runway space, they quickly engage high engine power to stop quicker. Instead this accelerates the Tu-204 forwards, eventually colliding with a highway embankment.

In complex systems, human and technical considerations cannot be isolated

Human factors concentrates on the “screen out”

Hardware/Software engineering concentrates on the “screen in”

Not enough attention on integrated system as a whole

(e.g, mode confusion, situation awareness errors, inconsistent behavior, etc.

Role of humans in systems is changing

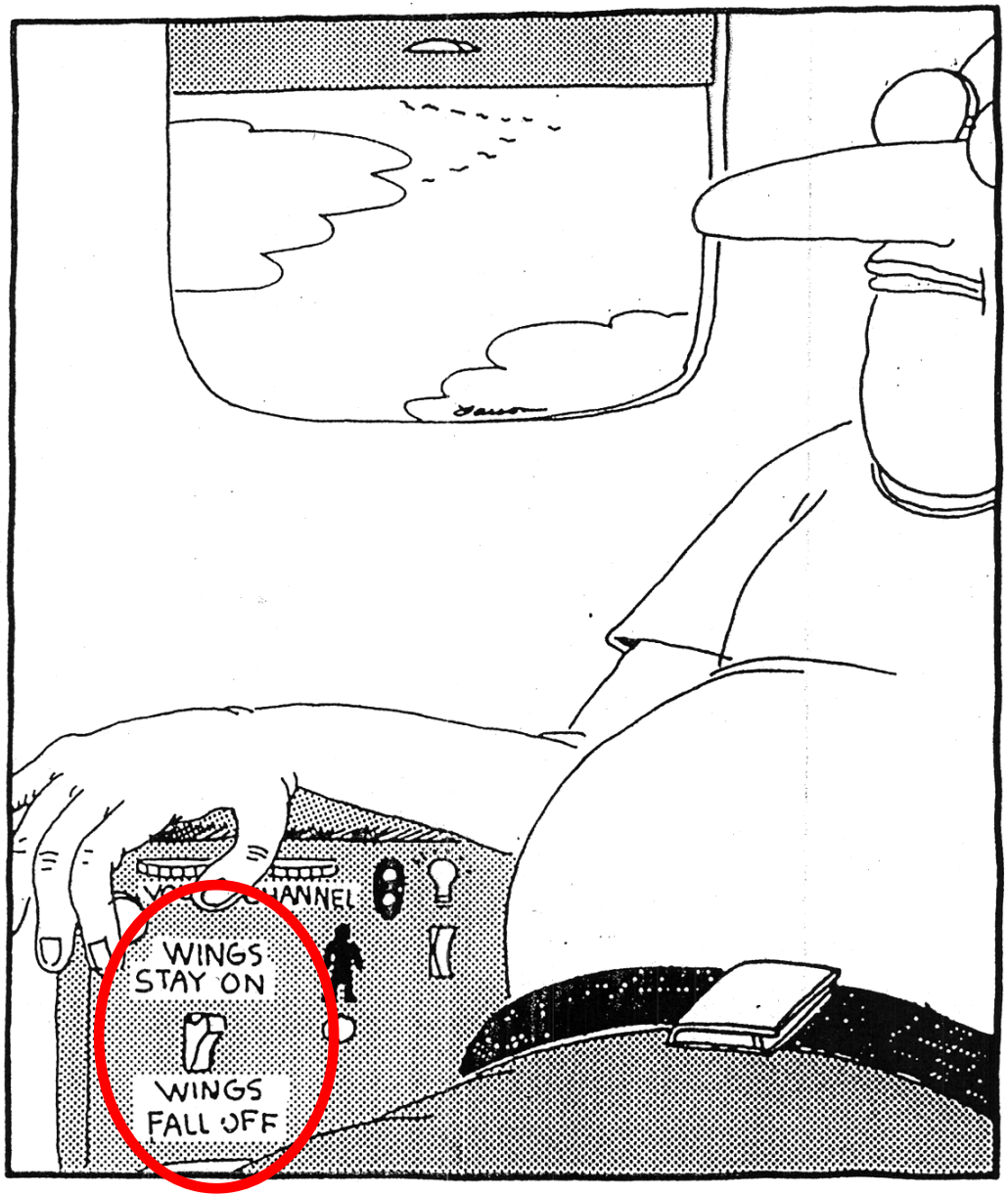
Traditional Approach

Typical assumption is that operator error is cause of most incidents and accidents

- So do something about operator involved (admonish, fire, retrain them)
- Or do something about operators in general
  - Marginalize them by putting in more automation
  - Rigidify their work by creating more rules and procedures

"Cause" from the American Airlines B-757 accident report (in Cali, Columbia):

> "Failure of the flight crew to revert to basic radio navigation at the time when the FMS-assisted navigation became confusing and demanded an excessive workload in a critical phase of flight."



Fumbling for his recline button Ted unwittingly instigates a disaster

A New Systems View of Operator Error

- Operator error is a symptom, not a cause
- All behavior affected by context (system) in which occurs
  - Role of operators is changing in software-intensive systems as is the errors they make
  - Designing systems in which operator error inevitable and then blame accidents on operators rather than designers
- To do something about operator error, must look at system in which people work:
  - Design of equipment
  - Usefulness of procedures
  - Existence of goal conflicts and production pressures
- **Human error is a symptom of a system that needs to be redesigned**

Lessons Learned from Past Accidents

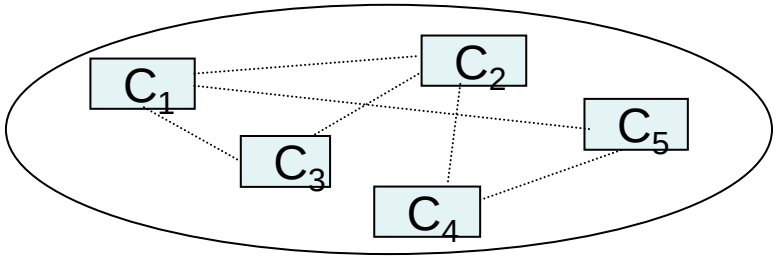
- Need to look beyond events to prevent accidents
  - Why did events occur?
  - To learn, we need to look at:
    - Conditions that lead to the events
    - Systemic factors that influence almost everything but not necessarily directly related (cannot just draw an arrow or assume a “failure”)
- Accidents today do not just result from component failures or operator errors. Need to consider design errors
- Cannot effectively tackle system safety without integrating human/software/hardware engineering.

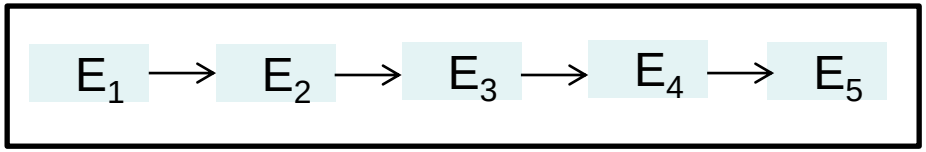
The Problem is Complexity

Ways to Cope with Complexity

- Analytic Decomposition
- Statistics
- Systems Theory

Analytic Decomposition ("Divide and Conquer")

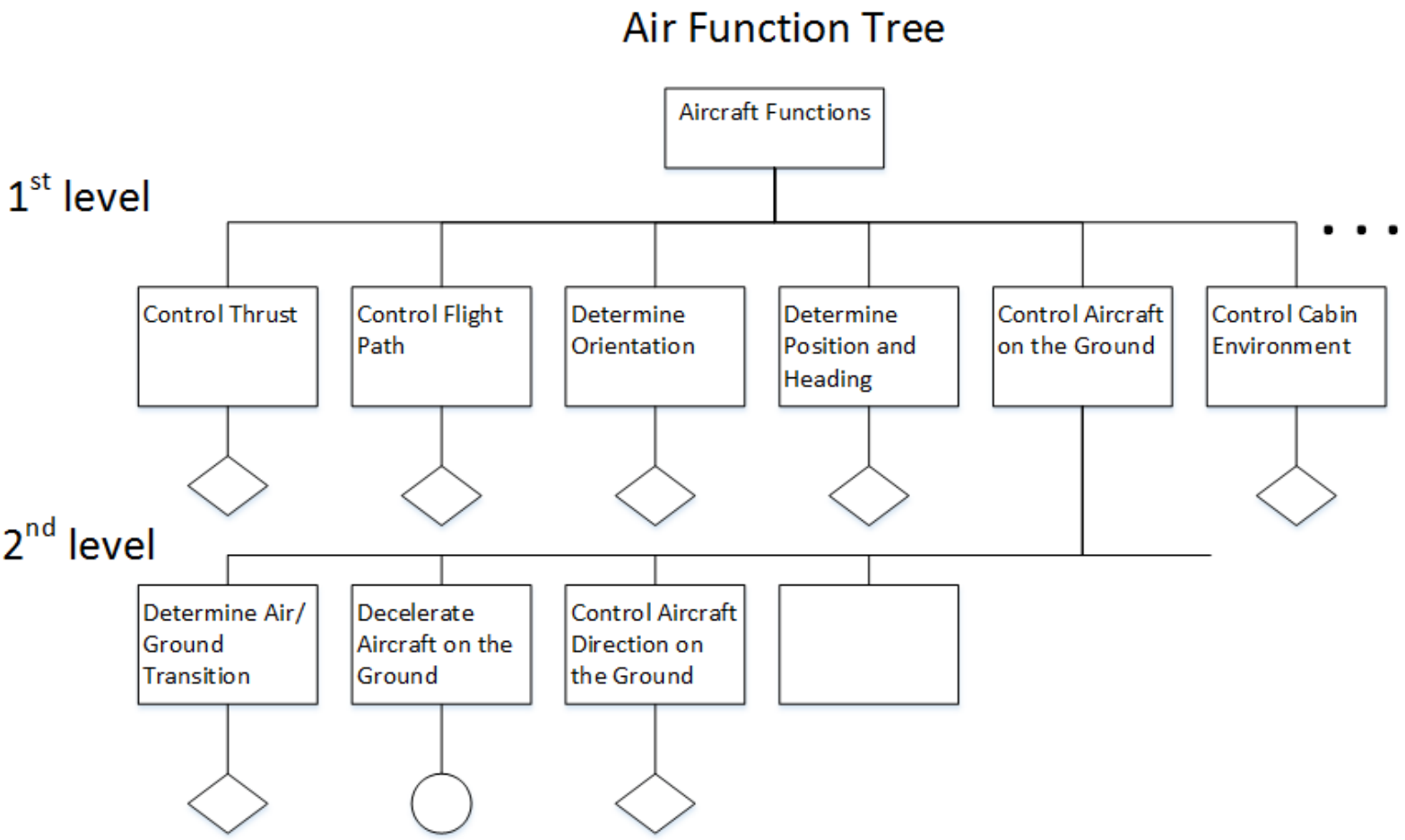




Analyze/examine pieces separately and combine results

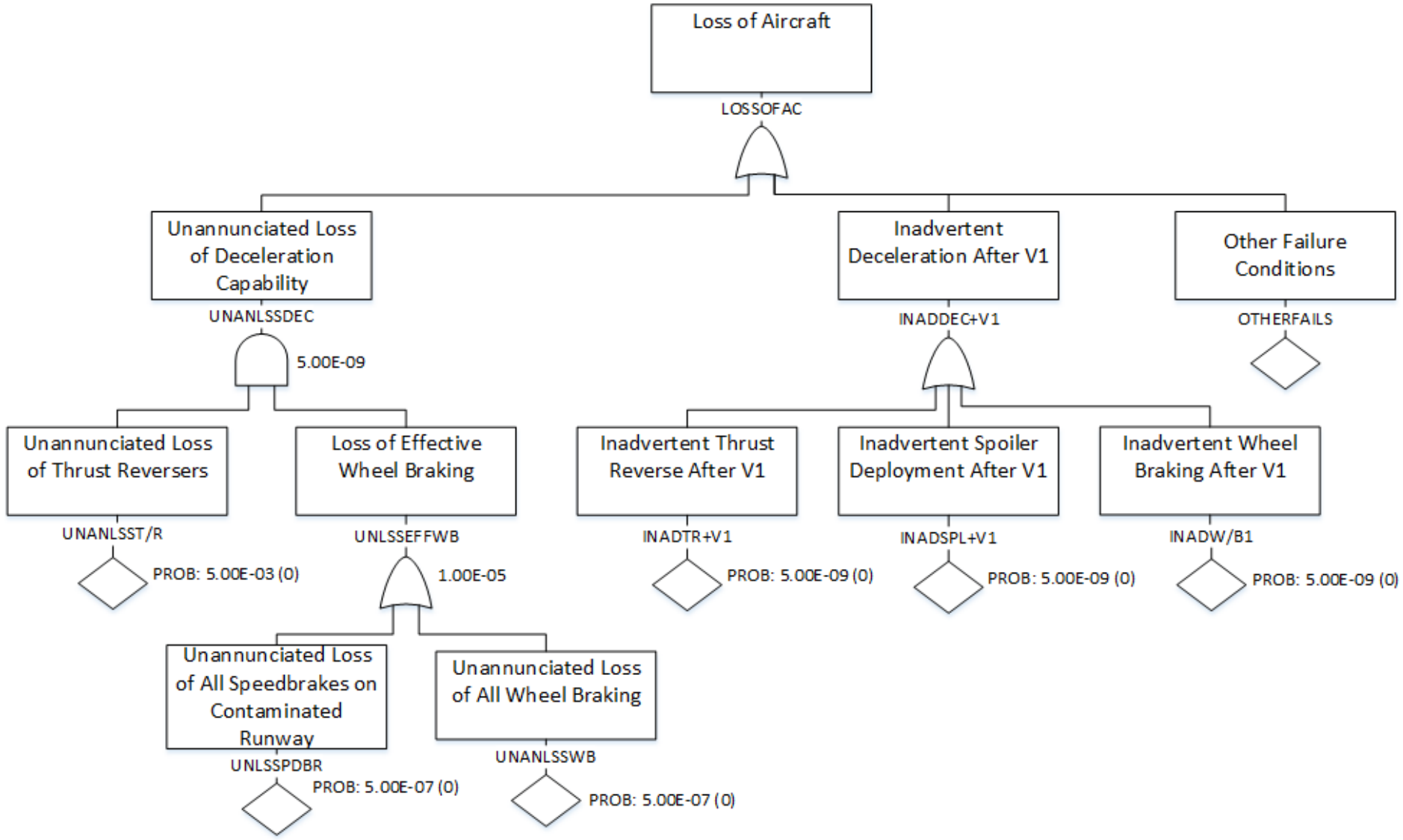
- Assumes such separation does not distort phenomenon
  - ✓ Each component or subsystem operates independently
  - ✓ Components act the same when examined singly as when playing their part in the whole
  - ✓ Components/events not subject to feedback loops and non-linear interactions
  - ✓ Interactions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

Typical Decomposition Approach (ARP 4761)



- ARP 4761A adding interactions among “failures” of functions but that is not the problem. Still bottom up.

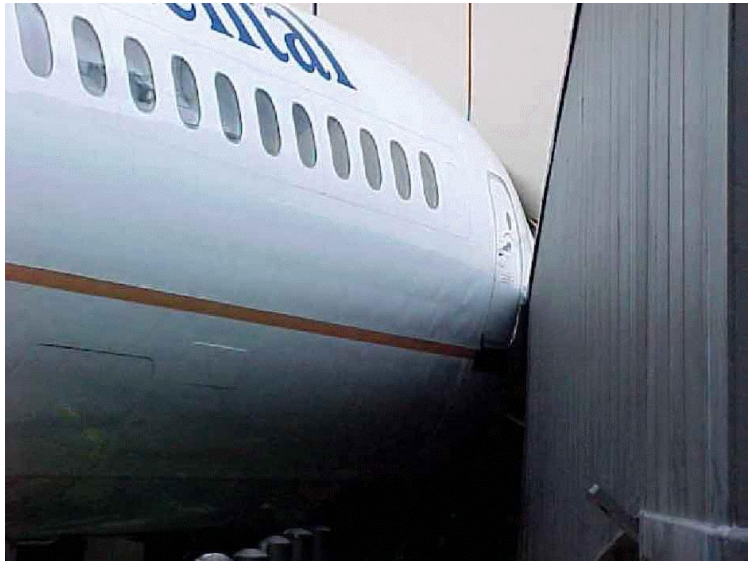
Combine individual component analyses bottom up



From SAE ARP 4761

Function	Failure Condition (Hazard Description)	Phase	Effect of Failure Condition on Aircraft/Crew	Classification
Decelerate Aircraft on the Ground	Loss of Deceleration Capability	Landing/ RTO/ Taxi		
	…	…	…	
	c. Unannunciated loss of deceleration capability	Taxi	Crew is unable to stop the aircraft on the taxi way or gate resulting In low speed contact with terminal, aircraft, or vehicles	Major
	d. Annunciated loss of deceleration capability	Taxi	Crew steers the aircraft clear of any obstacles and calls for a tug or portable stairs	No Safety Effect

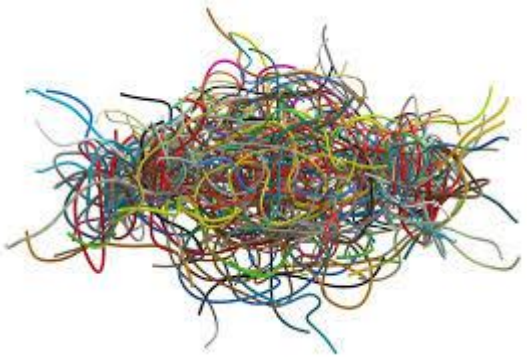
Continental Airlines Introduces the Improved Disembarkation Method



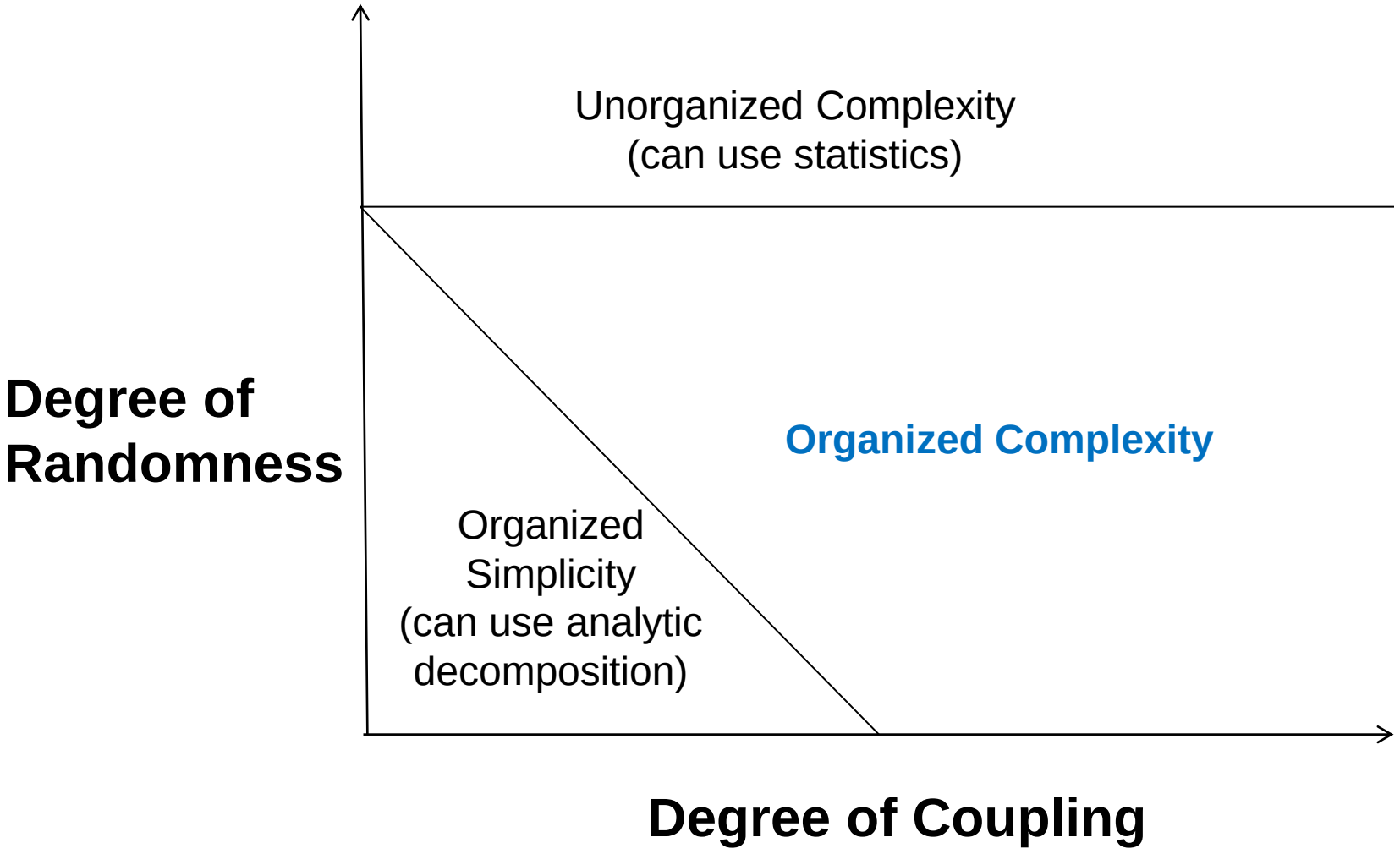
The Problem

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

  engineered and social systems



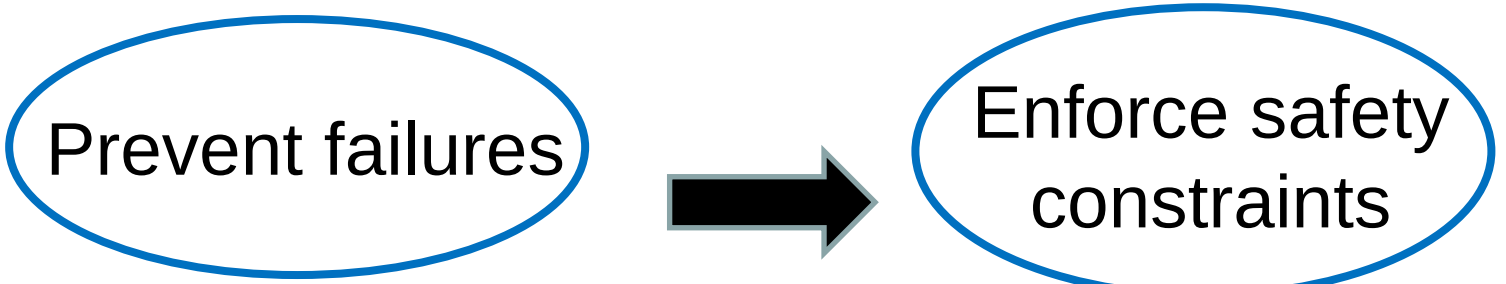
- Need a new theoretical basis
  - *System theory* can provide it



[Credit to Gerald Weinberg]

Here comes the paradigm change!

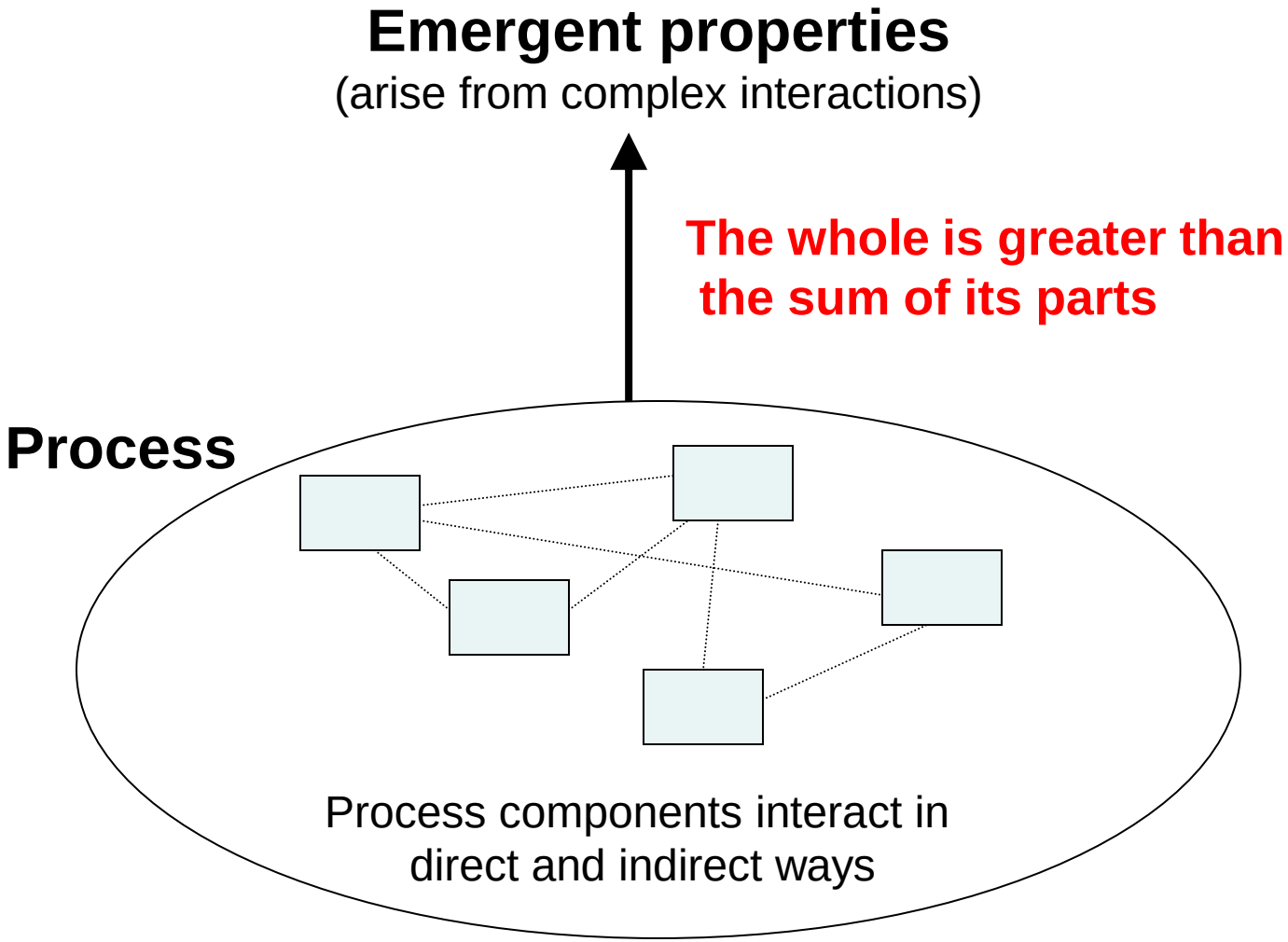
The paradigm change for effective safety and security engineering!



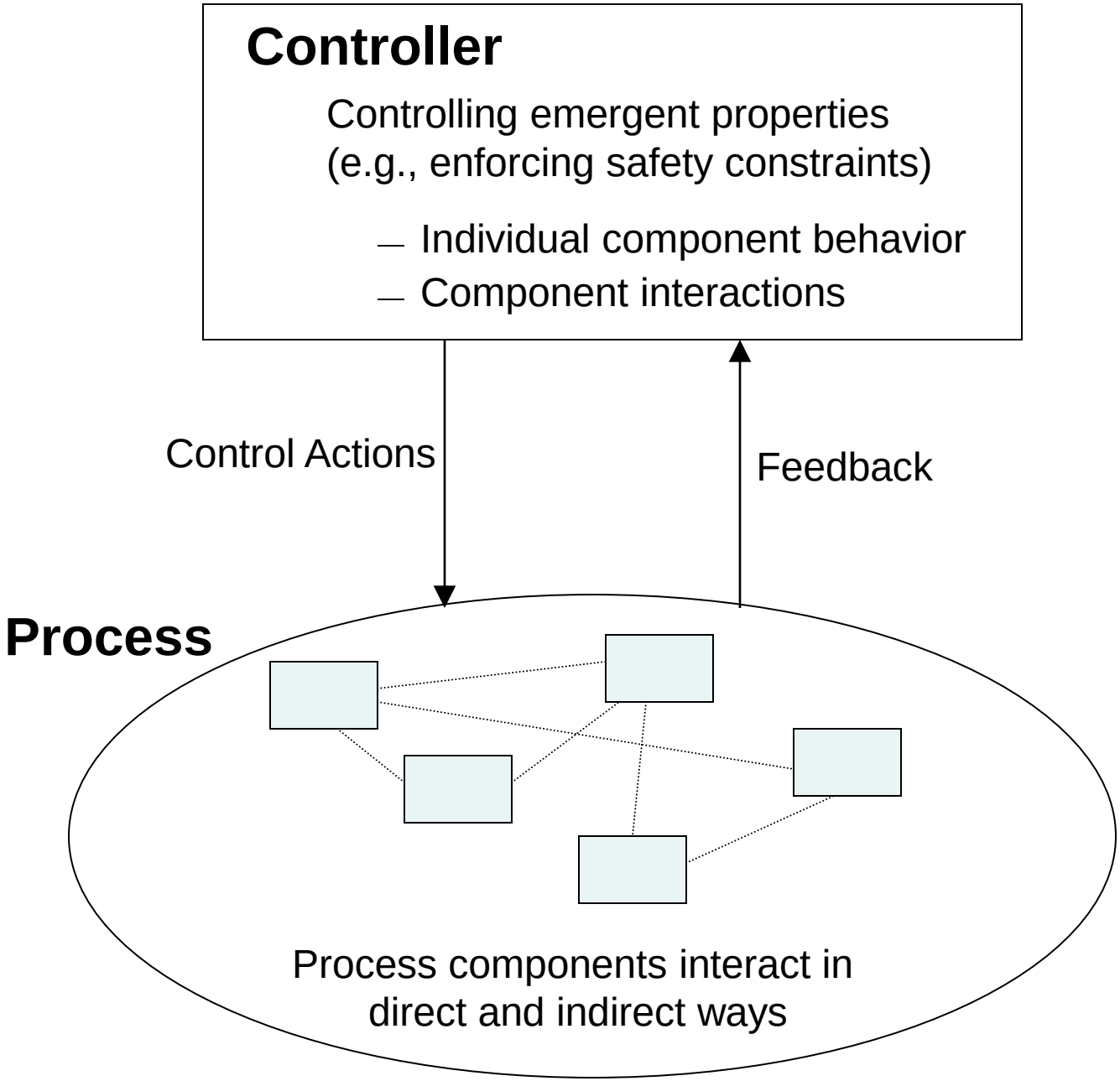
Treat Safety as a **Reliability** Problem

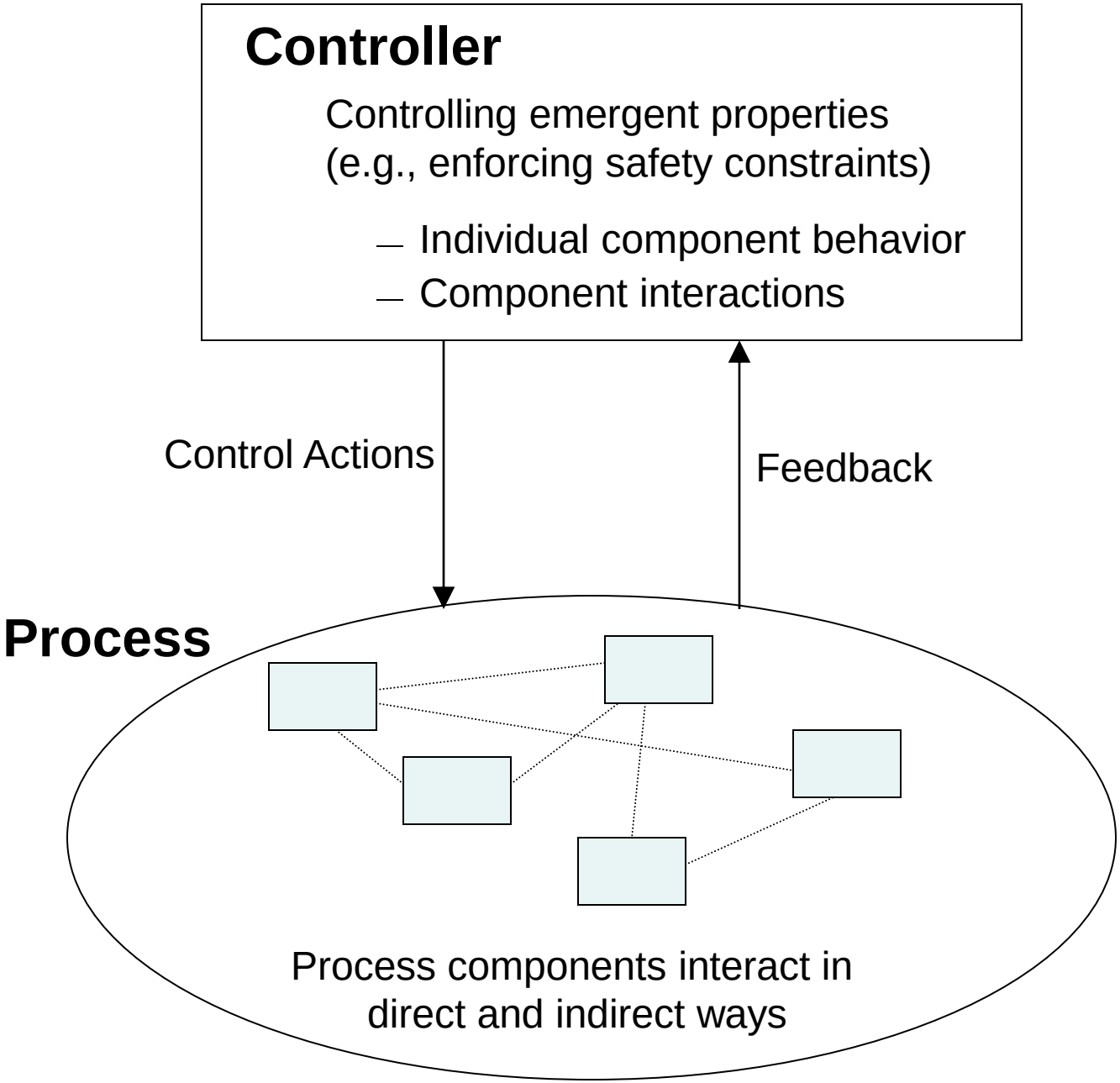
Treat Safety as a **Control** Problem

System Theory



Safety is an emergent property


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



ATC:
Safety
Throughput

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
- Operational processes

or through social controls

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

Controls/Controllers Enforce Safety Constraints

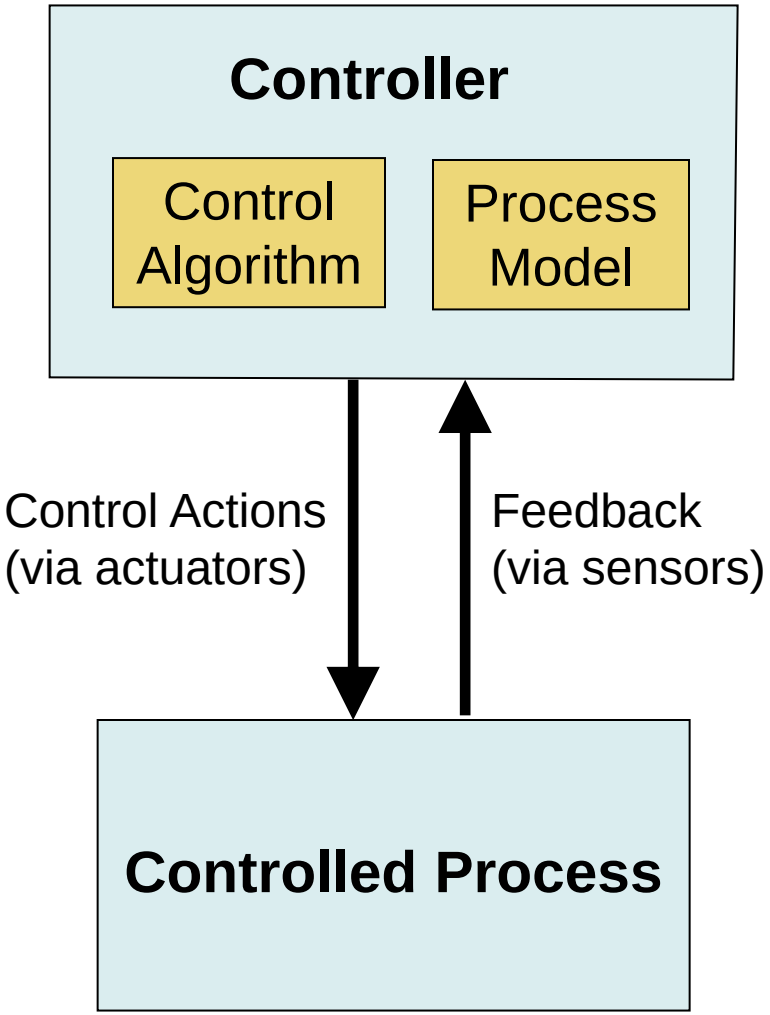
- Power must never be on when access door open (e.g, lockout/tagout)
- Public health system must prevent exposure of public to contaminated water, food products, and viruses
- Pressure in a offshore well must be controlled
- Two aircraft/automobiles must not violate minimum separation
- Aircraft must maintain sufficient lift to remain airborne
- Integrity of hull must be maintained on a submarine
- Toxic chemicals/radiation must not be released from plant
- Workers must not be exposed to workplace hazards

These are the High-Level Functional Safety/Security Requirements to Address During Design

STAMP
(System-Theoretic Accident Model and Processes)

- A new, more powerful accident/loss causality model
- Based on systems theory, not reliability theory
- Defines accidents/losses as a dynamic control problem (vs. a failure problem)
- Applies to VERY complex systems
- Includes
  - Scenarios from traditional hazard analysis methods (failure events)
  - Component interaction accidents
  - Software and system design errors
  - Human errors
  - Entire socio-technical system (not just technical part)

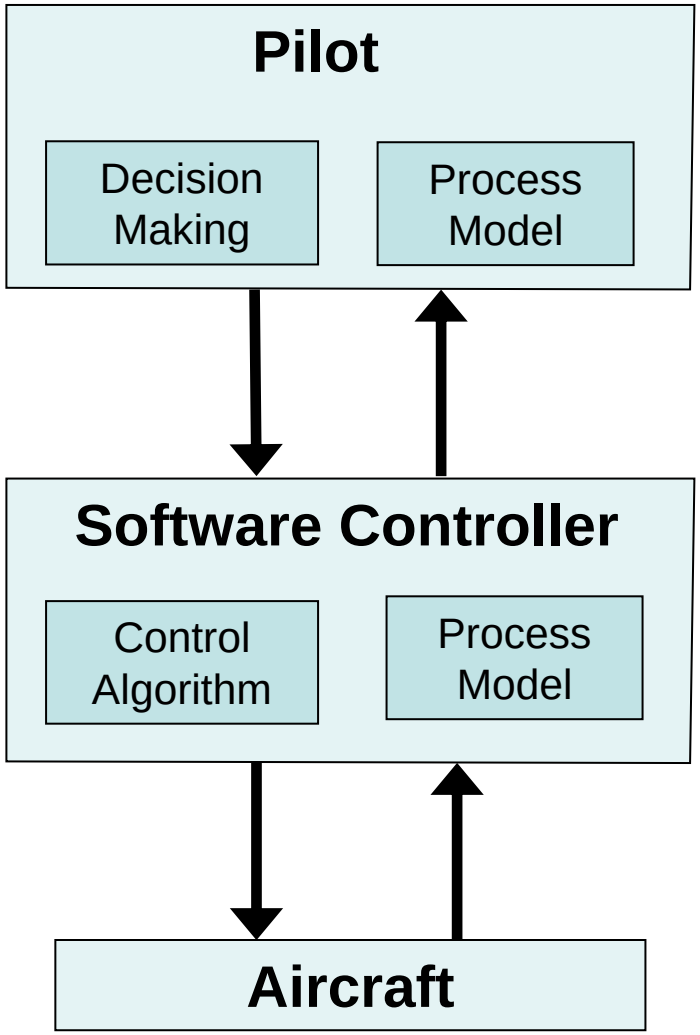
Treating Safety as a Control Problem



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

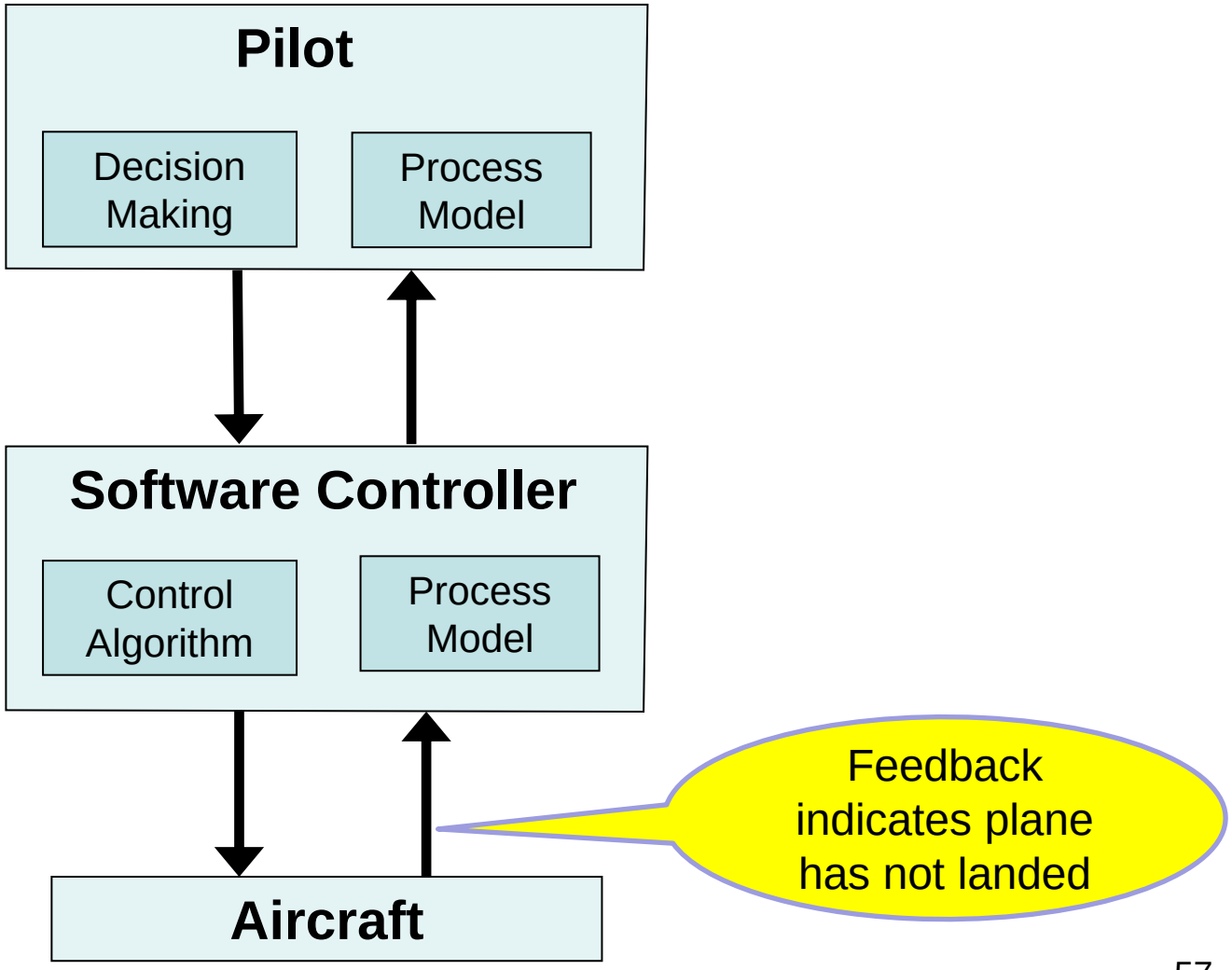
Warsaw (Reverse Thrusters)

Hazard: Inadequate aircraft deceleration after landing



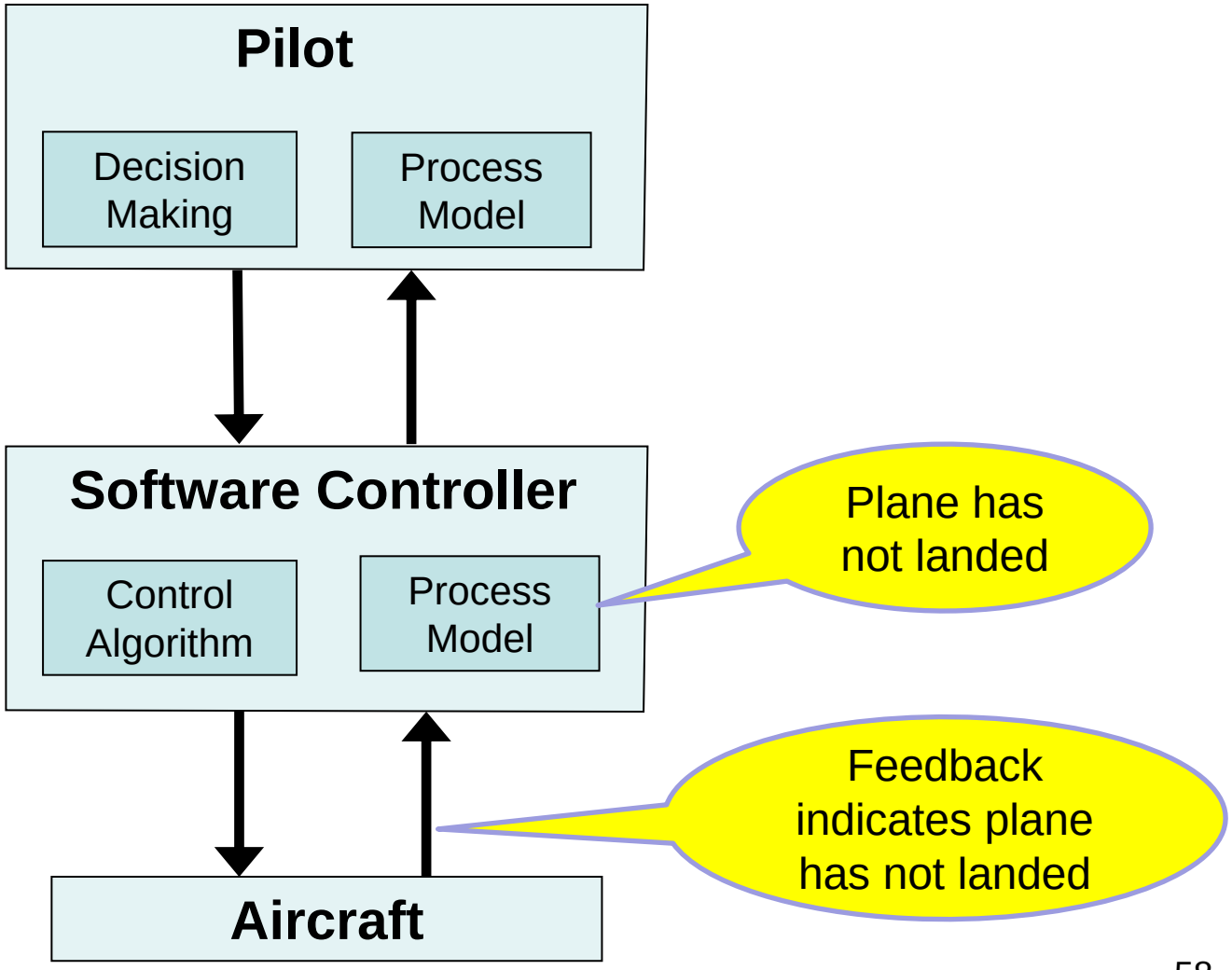
Warsaw (Reverse Thrusters)

Hazard: Inadequate aircraft deceleration after landing



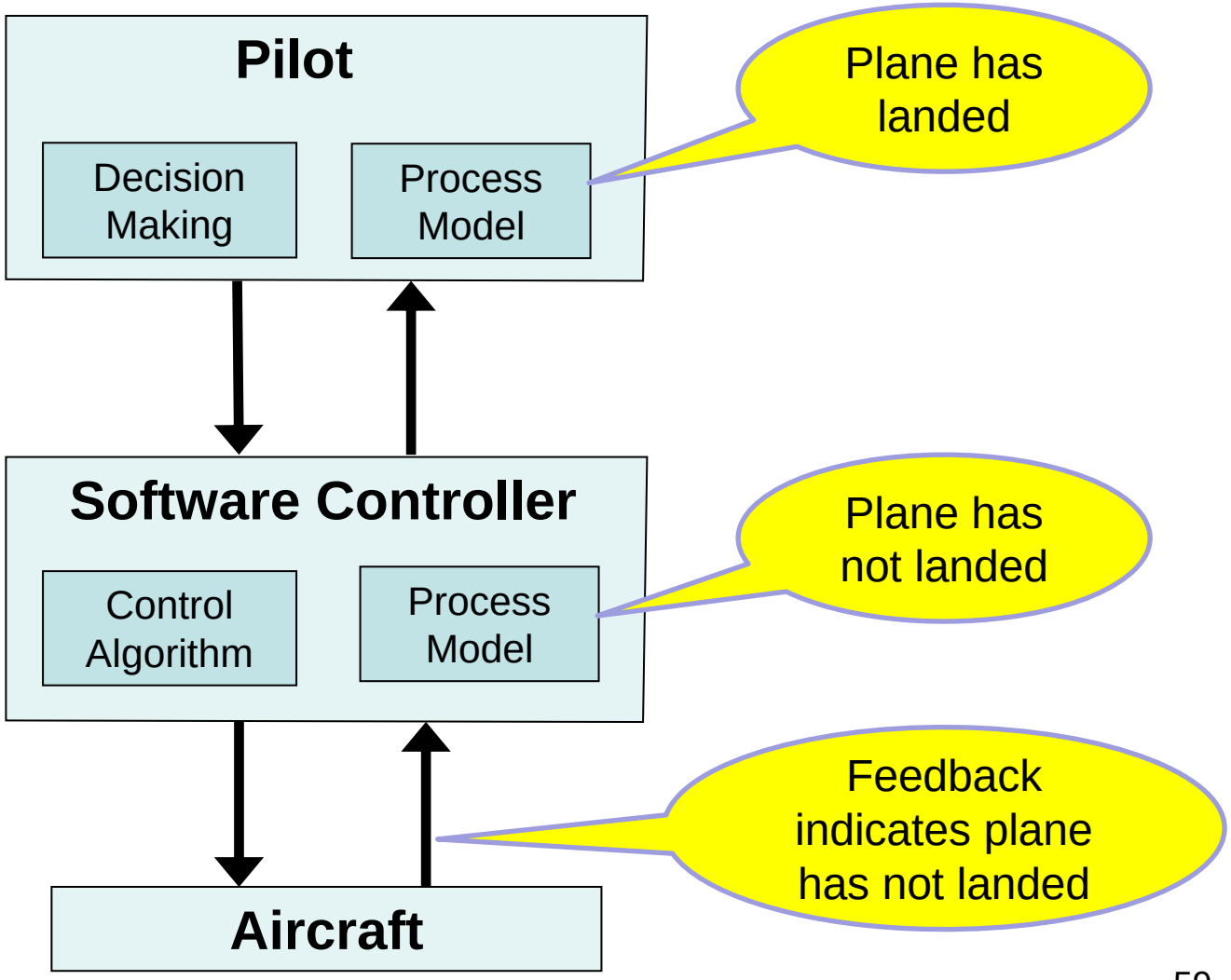
Warsaw (Reverse Thrusters)

Hazard: Inadequate aircraft deceleration after landing



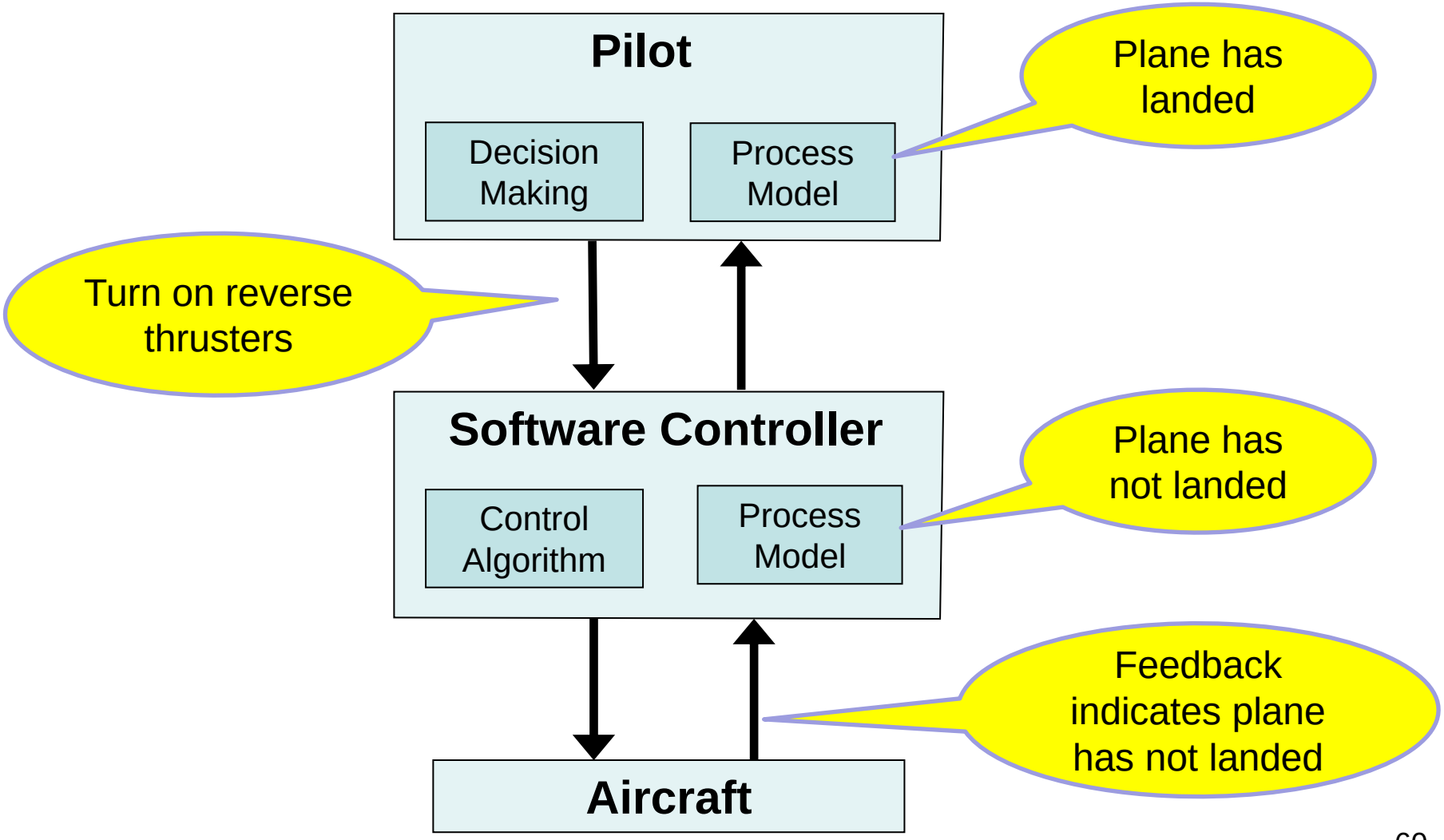
Warsaw (Reverse Thrusters)

Hazard: Inadequate aircraft deceleration after landing



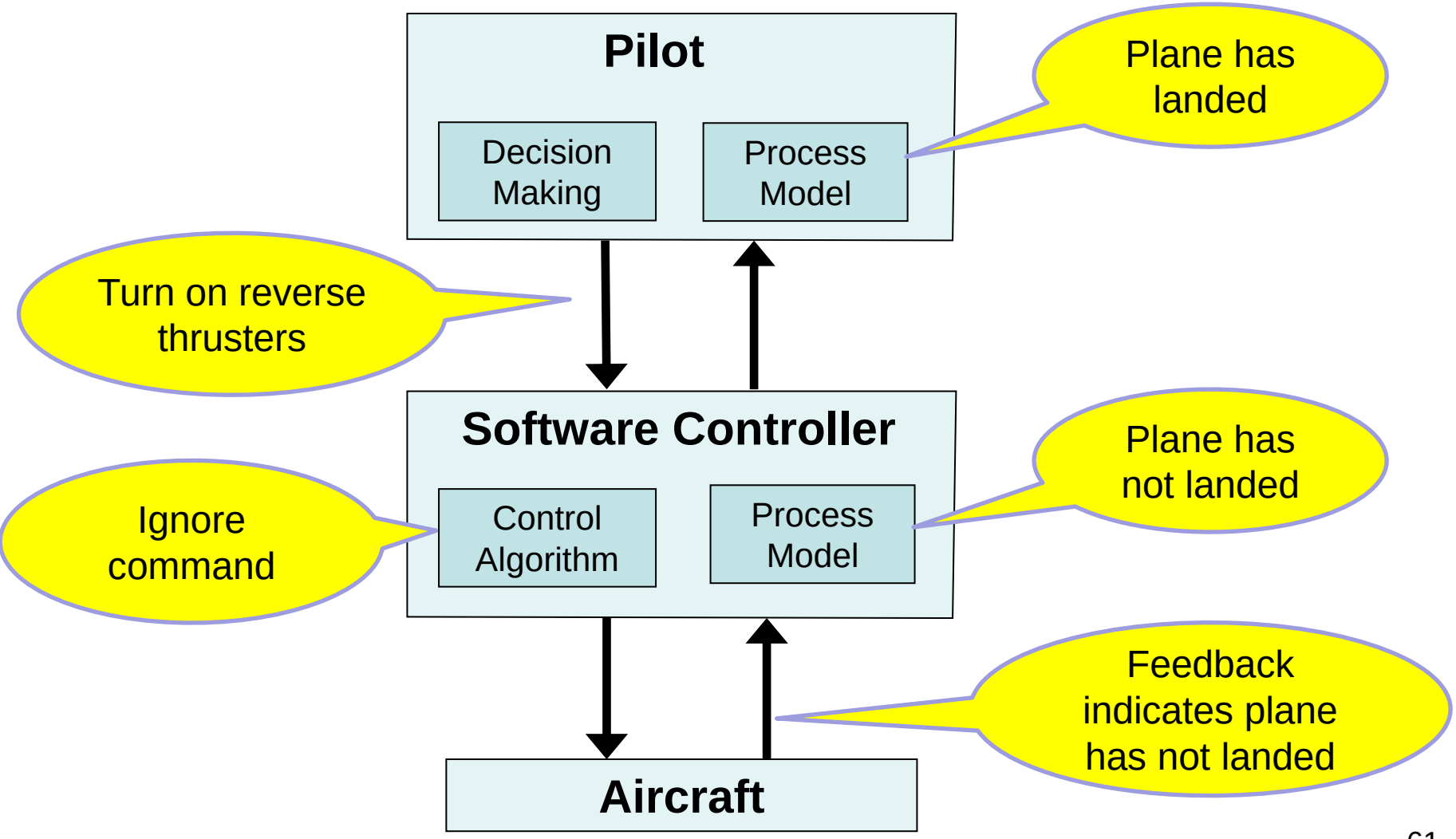
Warsaw (Reverse Thrusters)

Hazard: Inadequate aircraft deceleration after landing



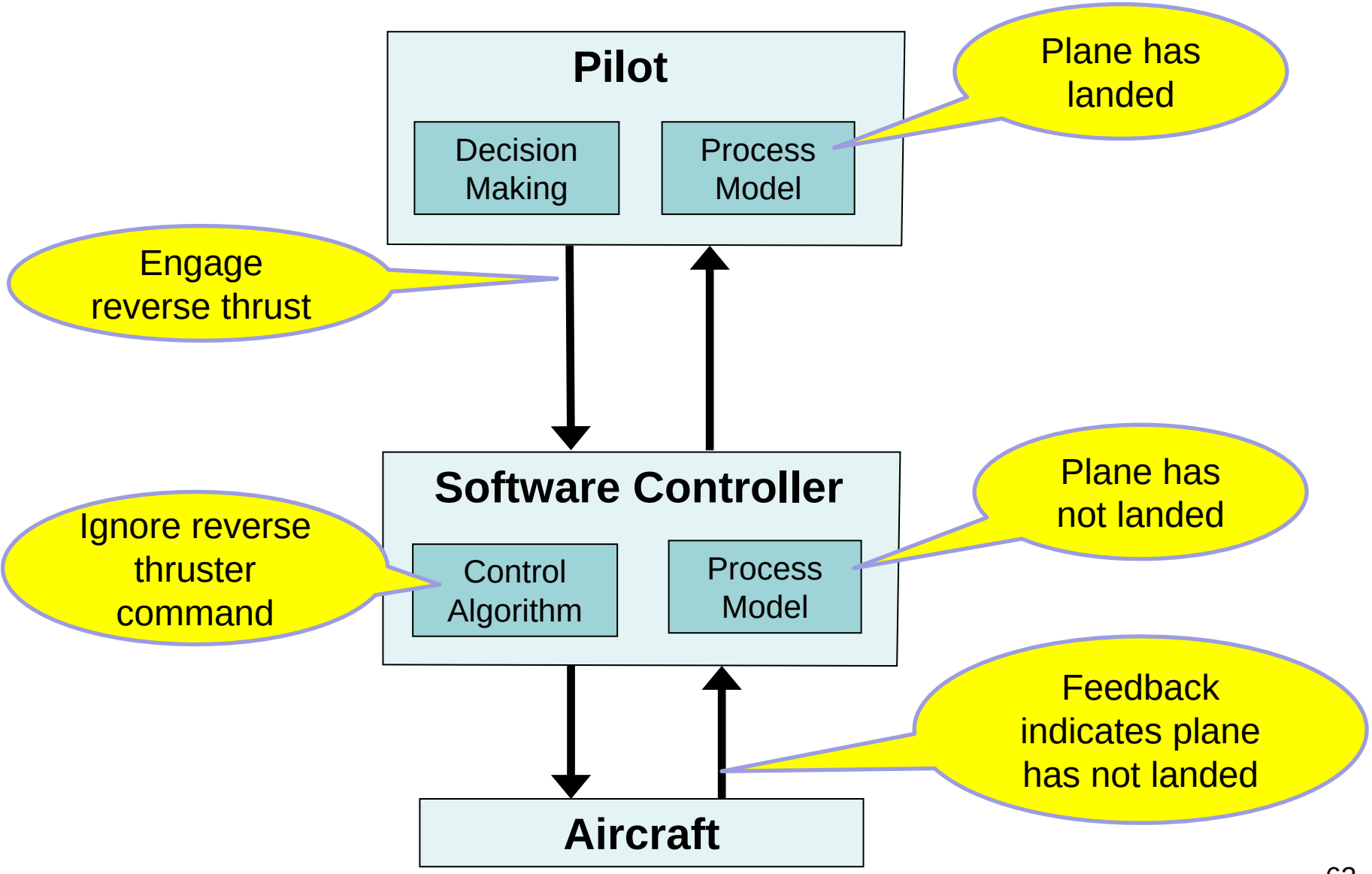
Warsaw (Reverse Thrusters)

Hazard: Inadequate aircraft deceleration after landing



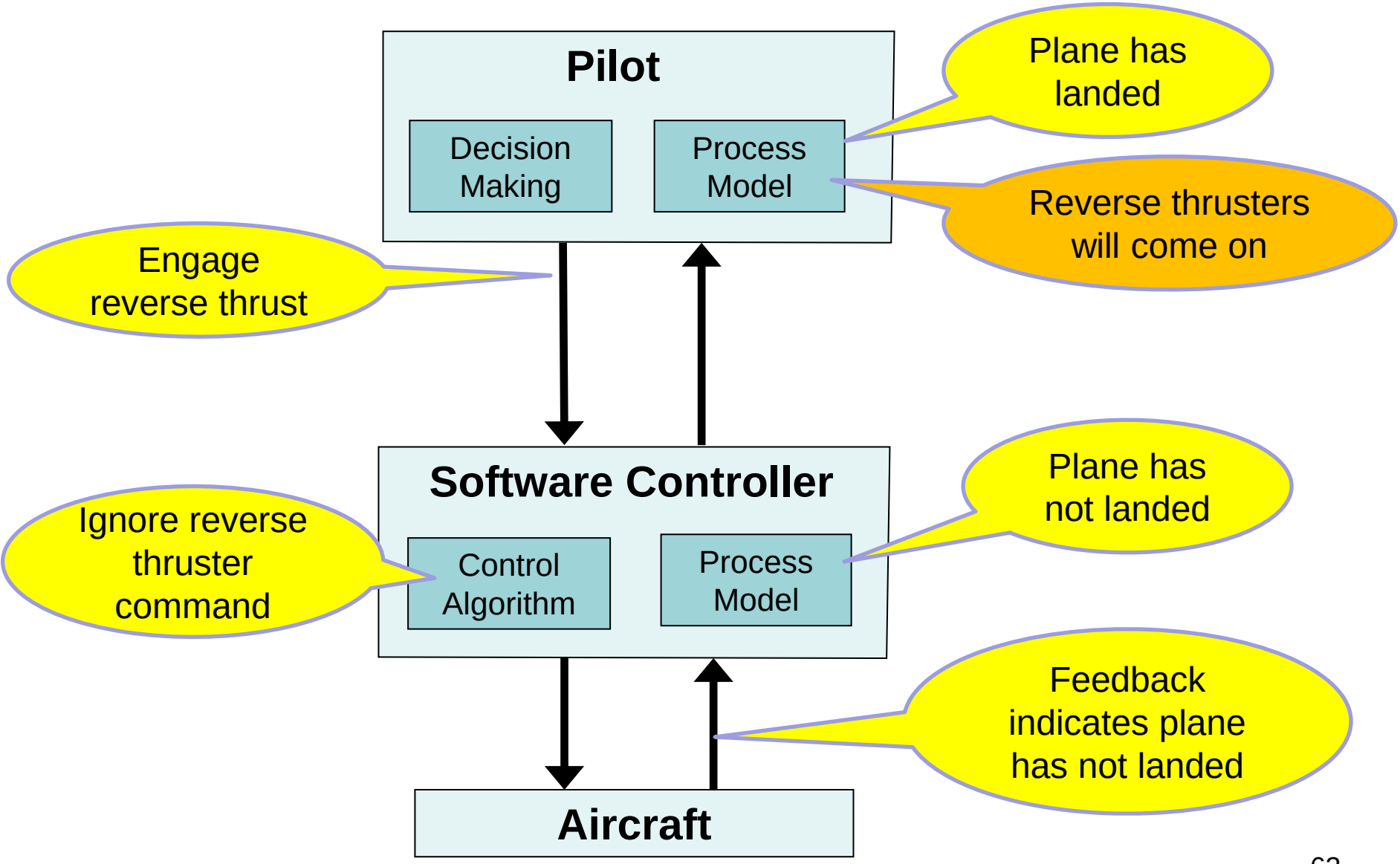
Moscow (Reverse Thrusters)

Hazard: Inadequate Deceleration after Landing



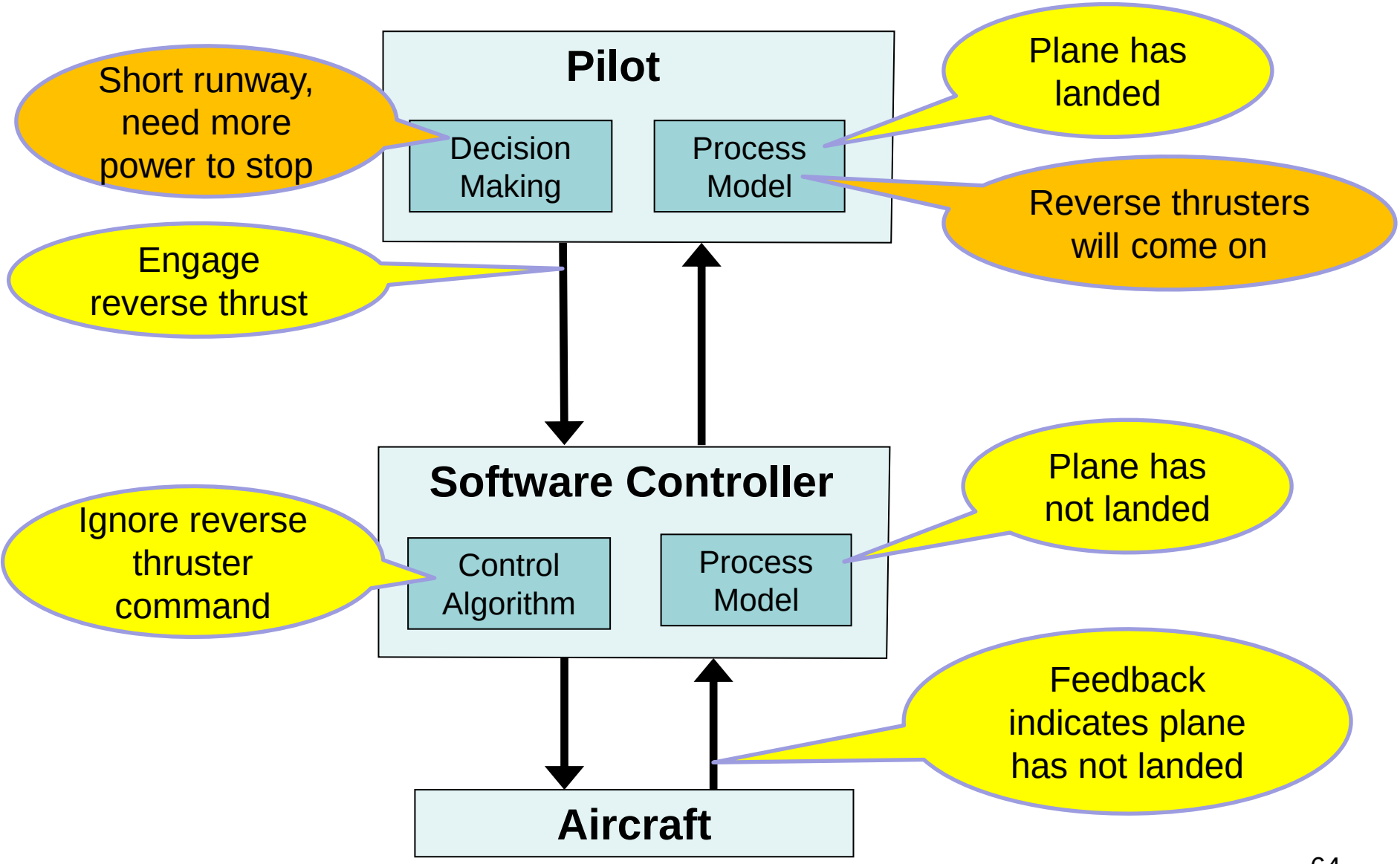
Moscow (Reverse Thrusters)

Hazard: Inadequate Deceleration after Landing



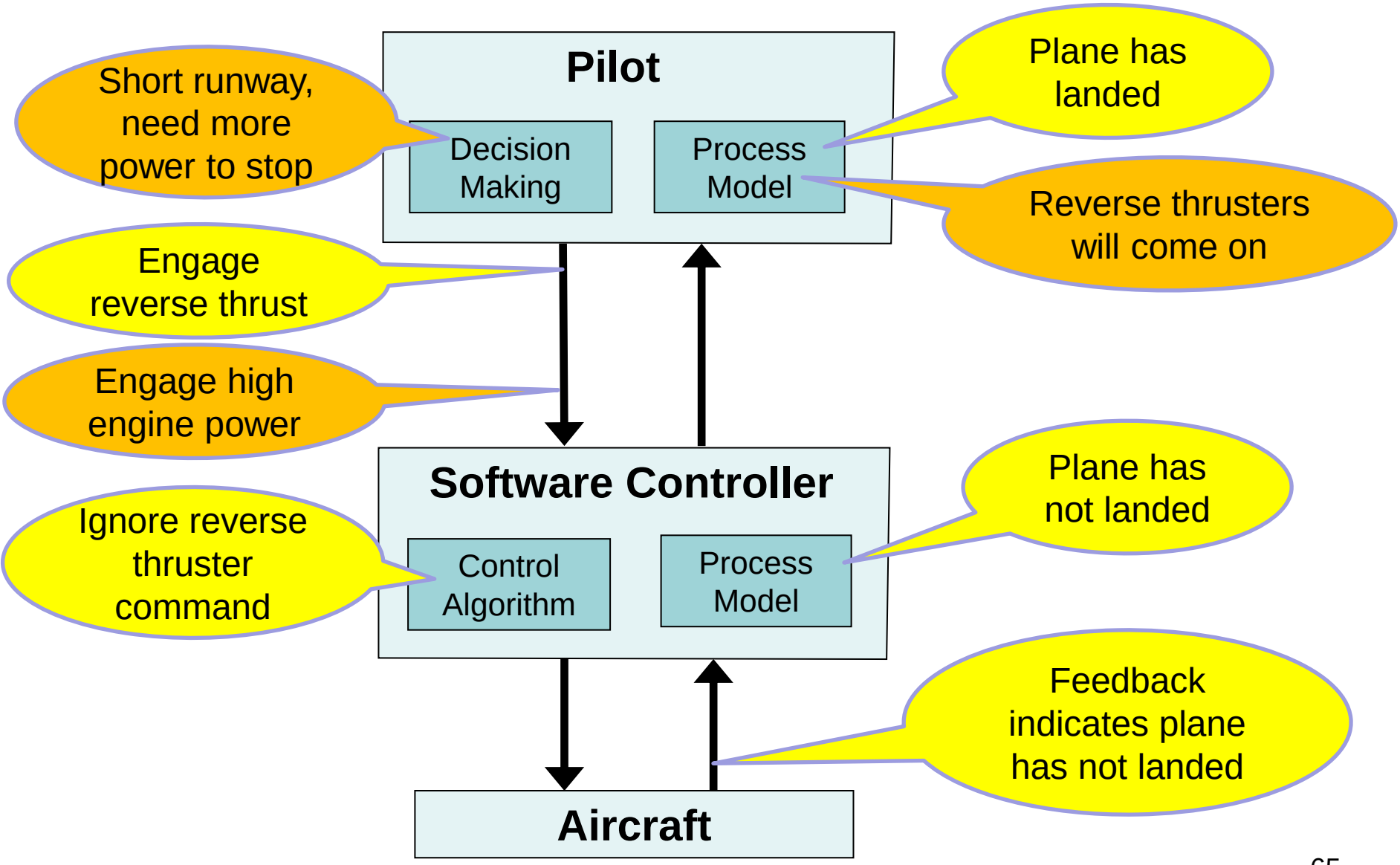
Moscow (Reverse Thrusters)

Hazard: Inadequate Deceleration after Landing



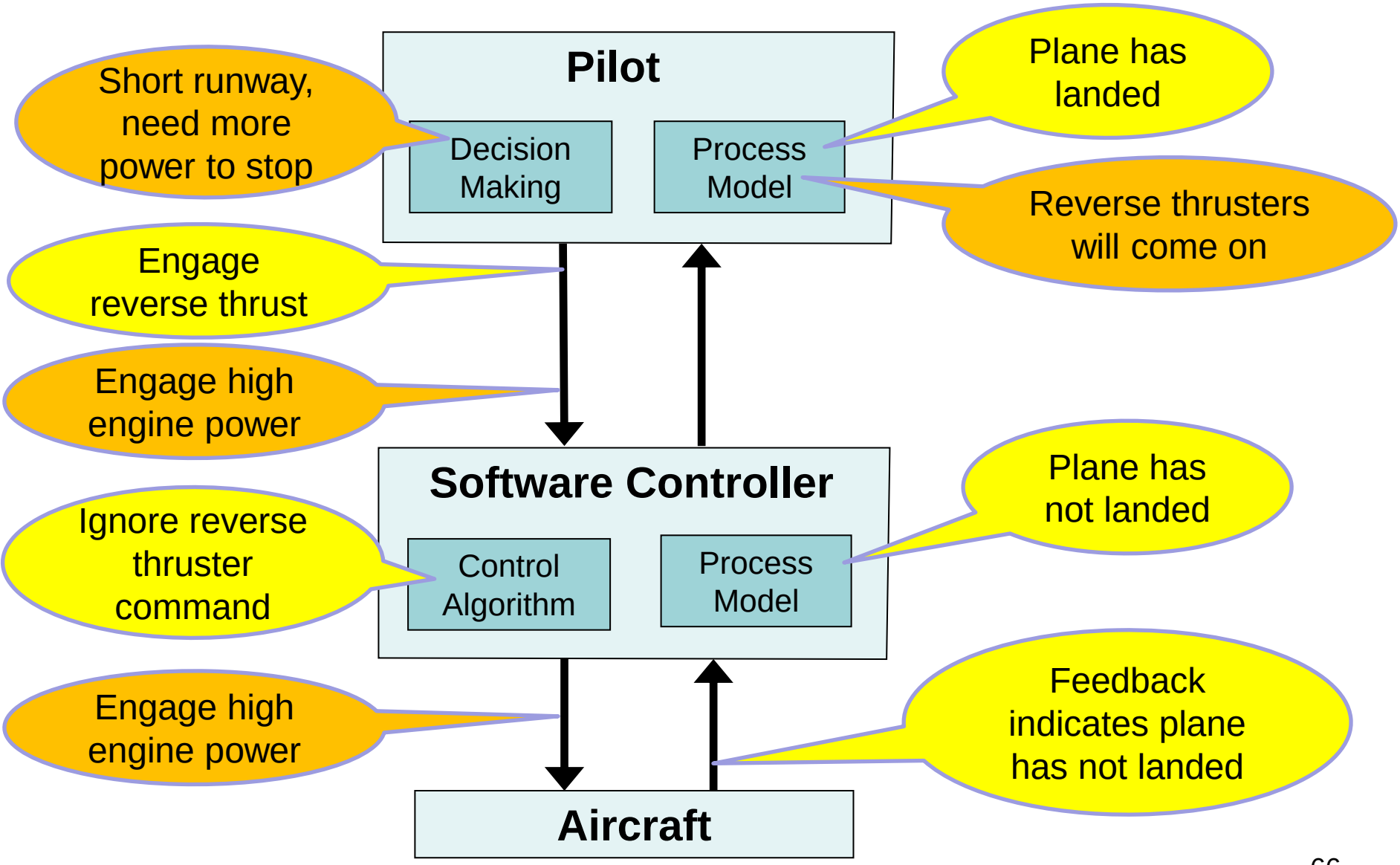
Moscow (Reverse Thrusters)

Hazard: Inadequate Deceleration after Landing



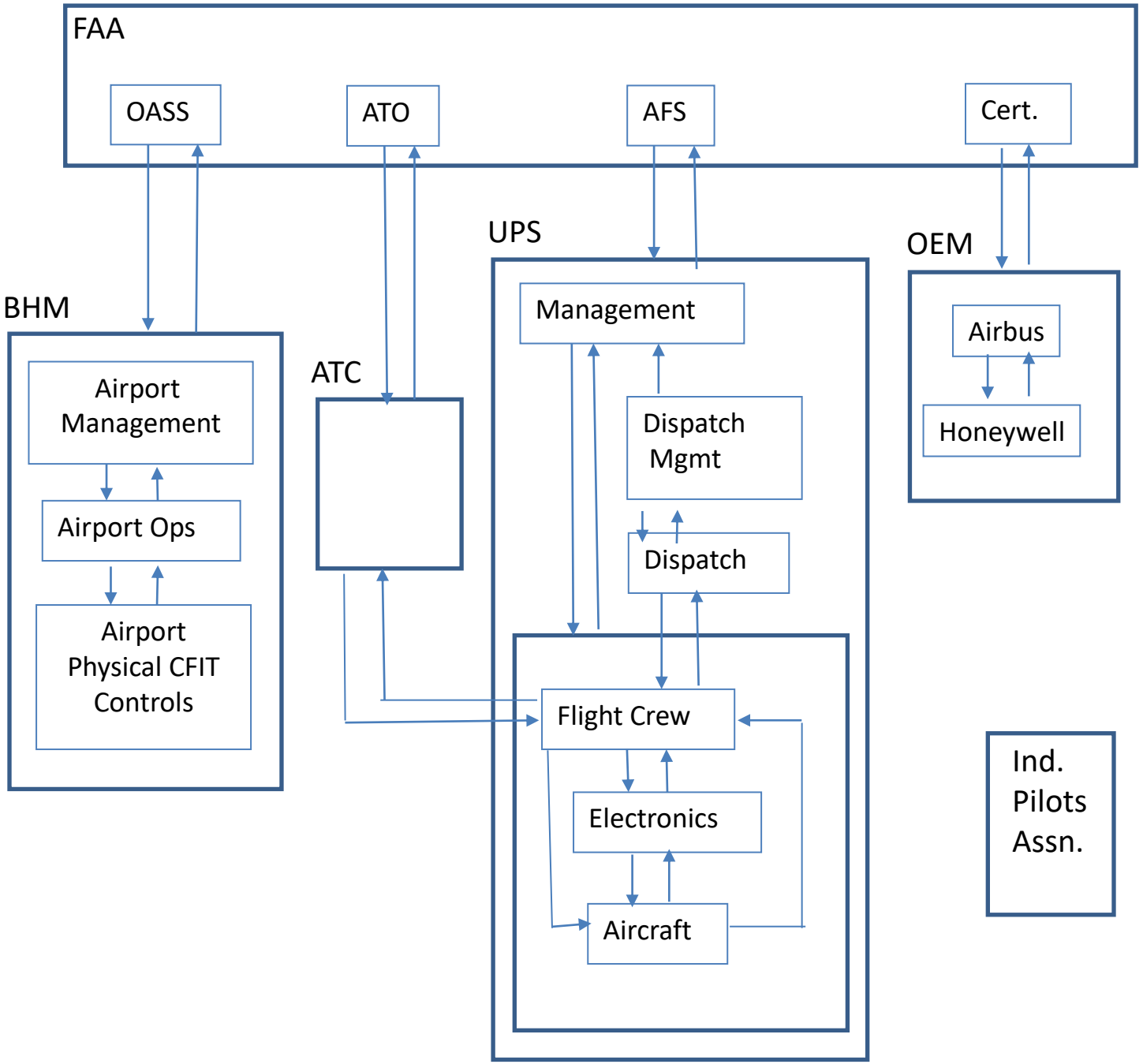
Moscow (Reverse Thrusters)

Hazard: Inadequate Deceleration after Landing



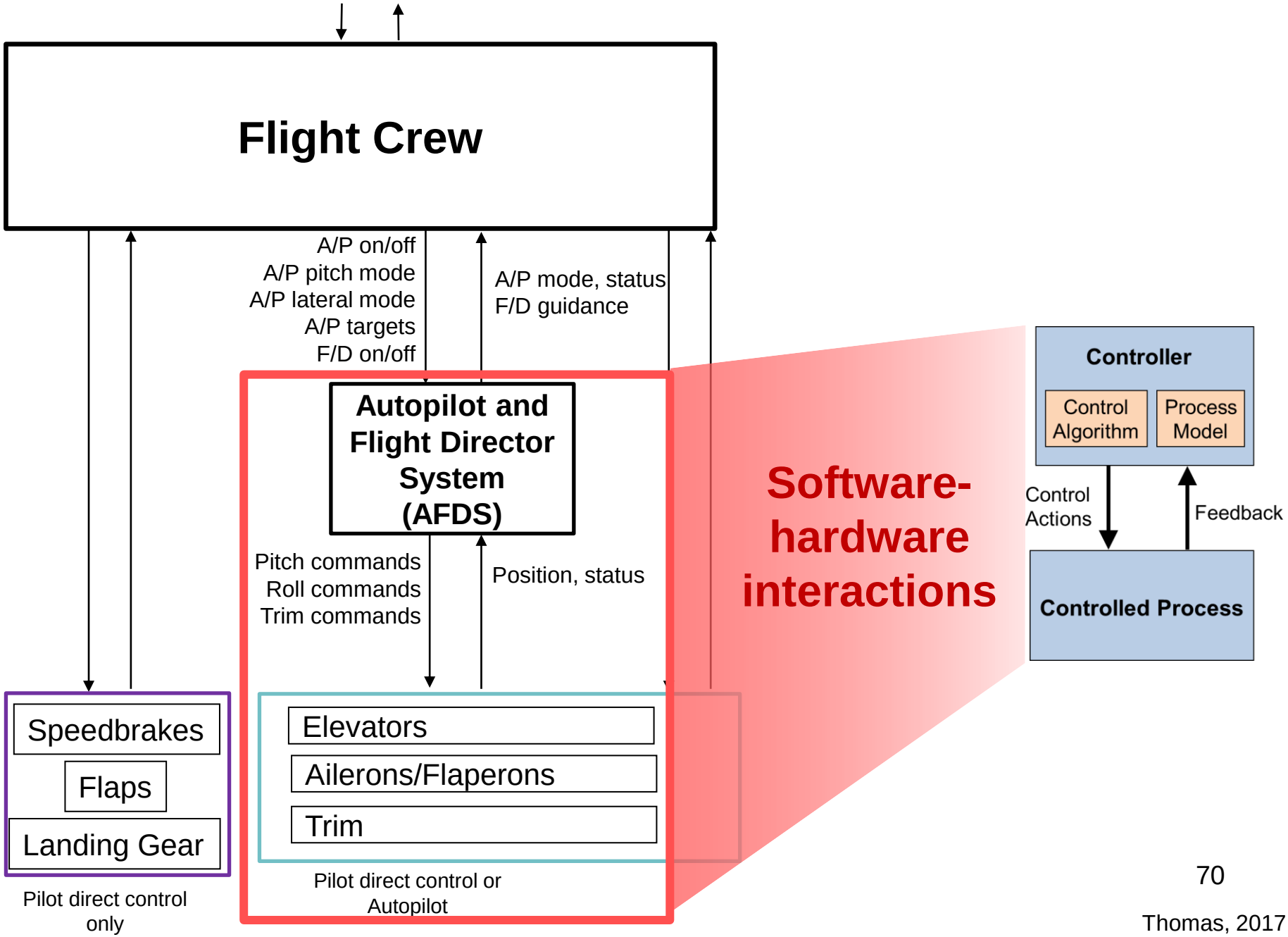
CAST: A System-Theoretic Accident Analysis Tool

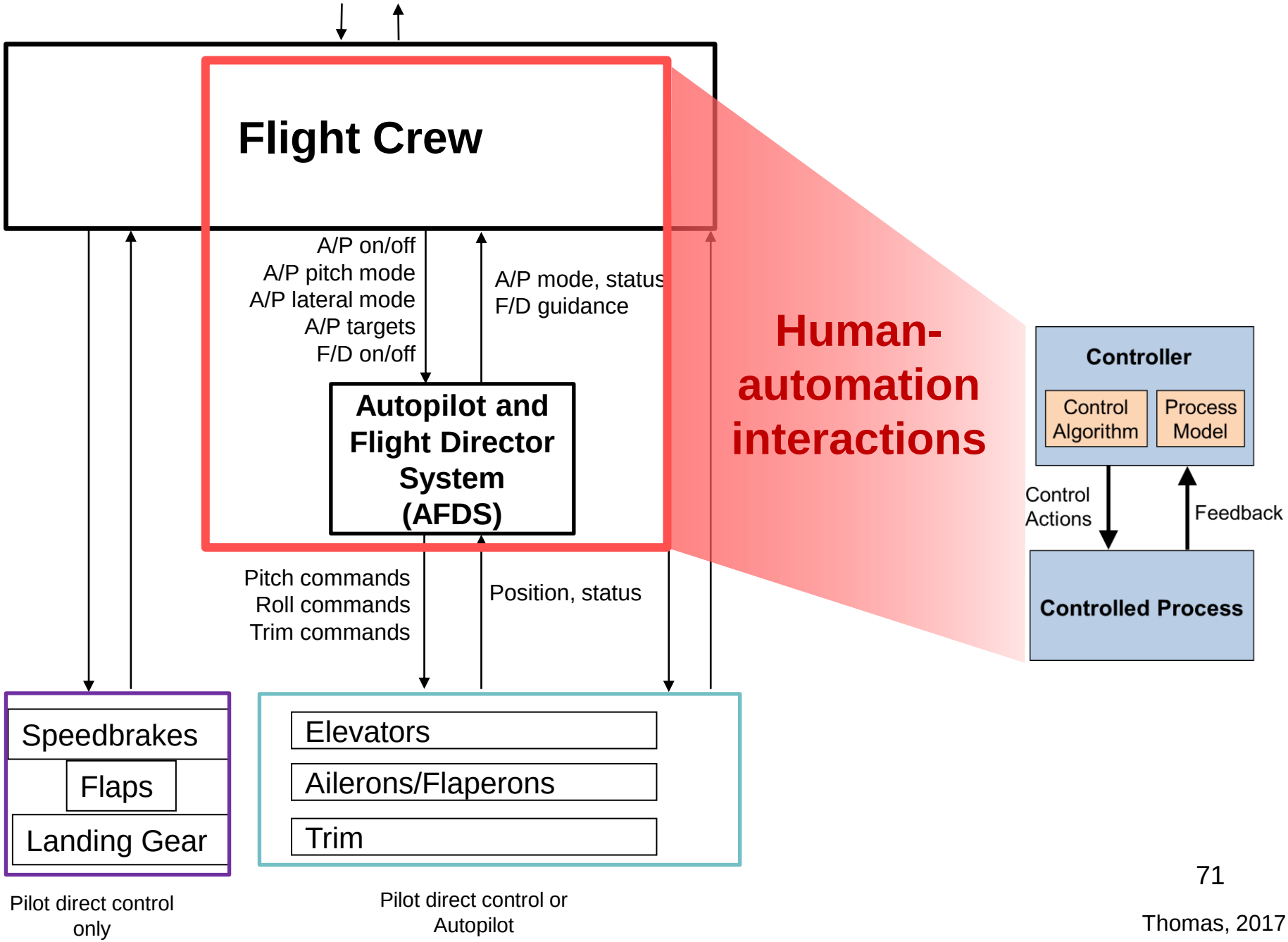
- Guides the identification of systemic factors, not just component failures
- Assists in generating the questions that need to be asked during the investigation
- Structured, step-by-step process


FAA
OASS
ATO
AFS
Cert.
BHM
Airport Management
Airport Ops
Airport Physical CFIT Controls
ATC
UPS
Management
Dispatch Mgmt
Dispatch
Flight Crew
Electronics
Aircraft
OEM
Airbus
Honeywell
Ind. Pilots Assn.

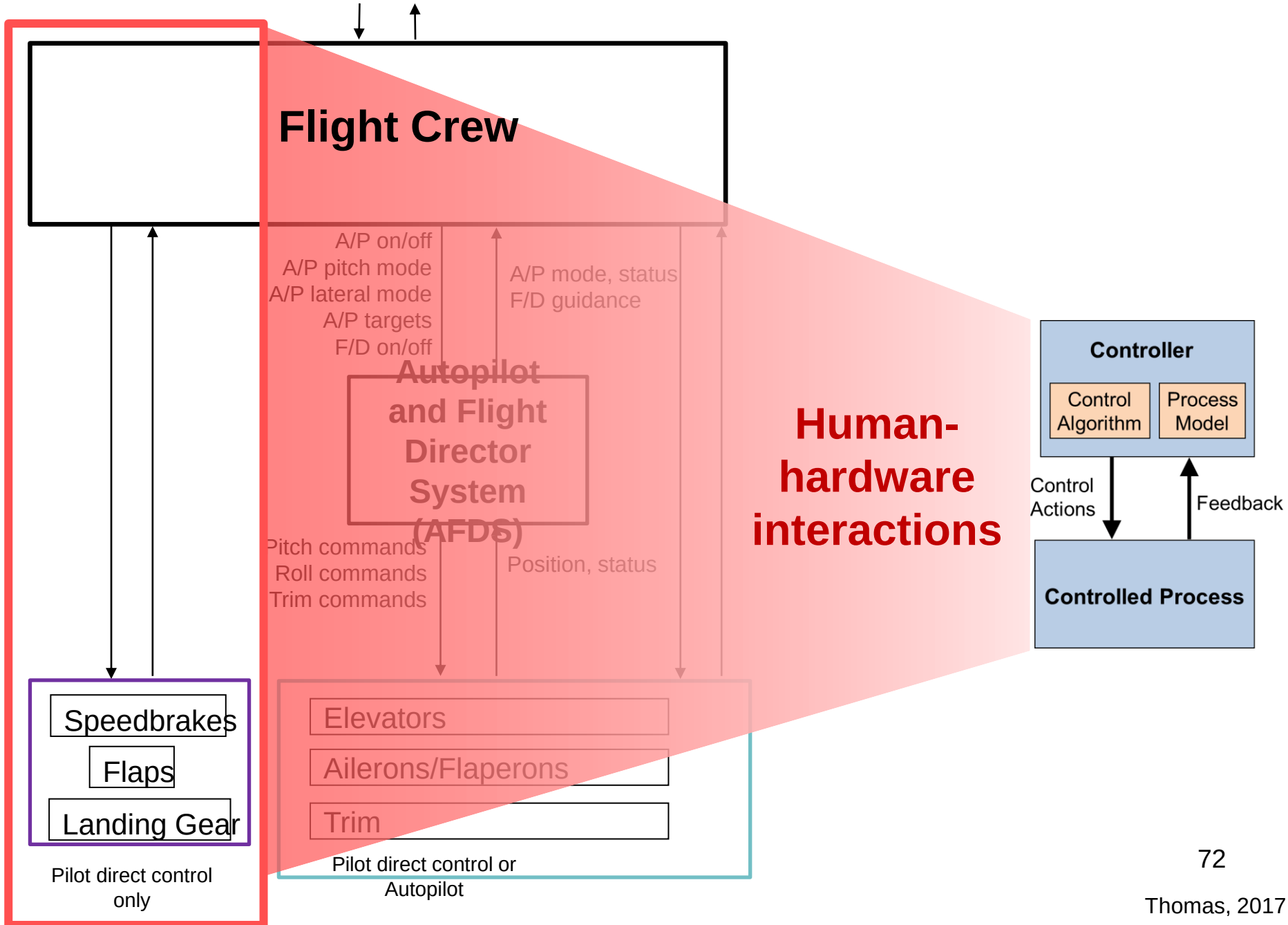
Additional Systemic Factors

- Industry and organizational safety culture
- Safety information system
- Communication and coordination among controllers
- Dynamics and changes over time

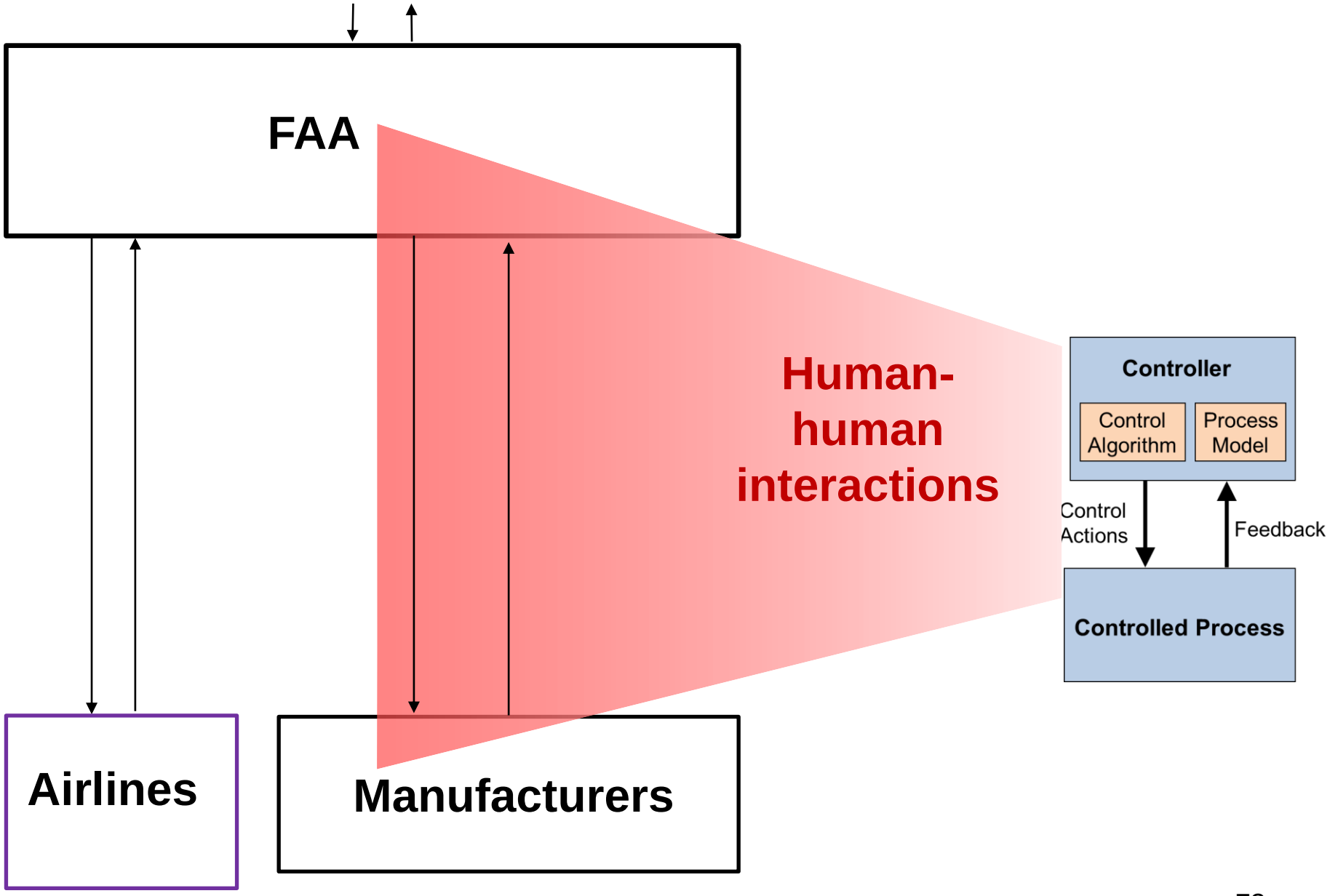

Flight Crew
A/P on/off
A/P pitch mode
A/P lateral mode
A/P targets
F/D on/off
A/P mode, status
F/D guidance
Autopilot and Flight Director System (AFDS)
Pitch commands
Roll commands
Trim commands
Position, status
Speedbrakes
Flaps
Landing Gear
Elevators
Ailerons/Flaperons
Trim
Pilot direct control only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


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
Human-hardware interactions
Controller
Control Algorithm
Process Model
Control Actions
Feedback
Controlled Process
Speedbrakes
Flaps
Landing Gear
Pilot direct control only
Elevators
Ailerons/Flaperons
Trim
Pilot direct control or Autopilot

Thomas, 2017


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

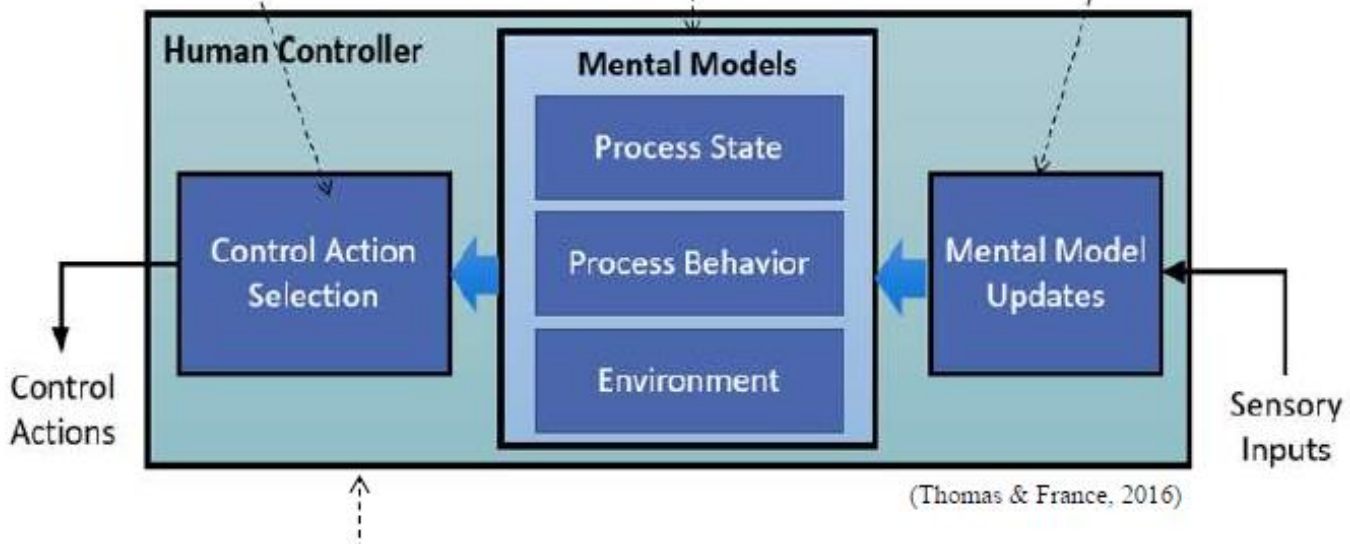
Thomas, 2017

A NEW MODEL FOR HUMAN CONTROLLERS

Captures the controller's goals and how decisions are made based on the mental models

Captures specific types of flaws in the way the human controller conceptualizes the system and environment

Captures the influence of human experiences, and expectations on the processing of sensory input



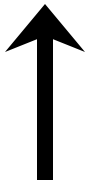
(Thomas & France, 2016)

What kinds of tools are available?

Processes

System Engineering | Risk Management | Organizational Design (SMS)

Operations | Certification and Acquisition | Regulation



Tools

Accident Analysis
CAST

Hazard Analysis
STPA

Security Analysis
STPA-Sec

Organizational/Cultural Risk Analysis

Identifying Leading Indicators

MBSE
SpecTRM & …

STAMP: Theoretical Causality Model

What do we need to get there?

- An enhanced causality model
- New collaborative tools that allow people with different backgrounds to design systems together
- **People willing to learn something new (perhaps the hardest)**

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"

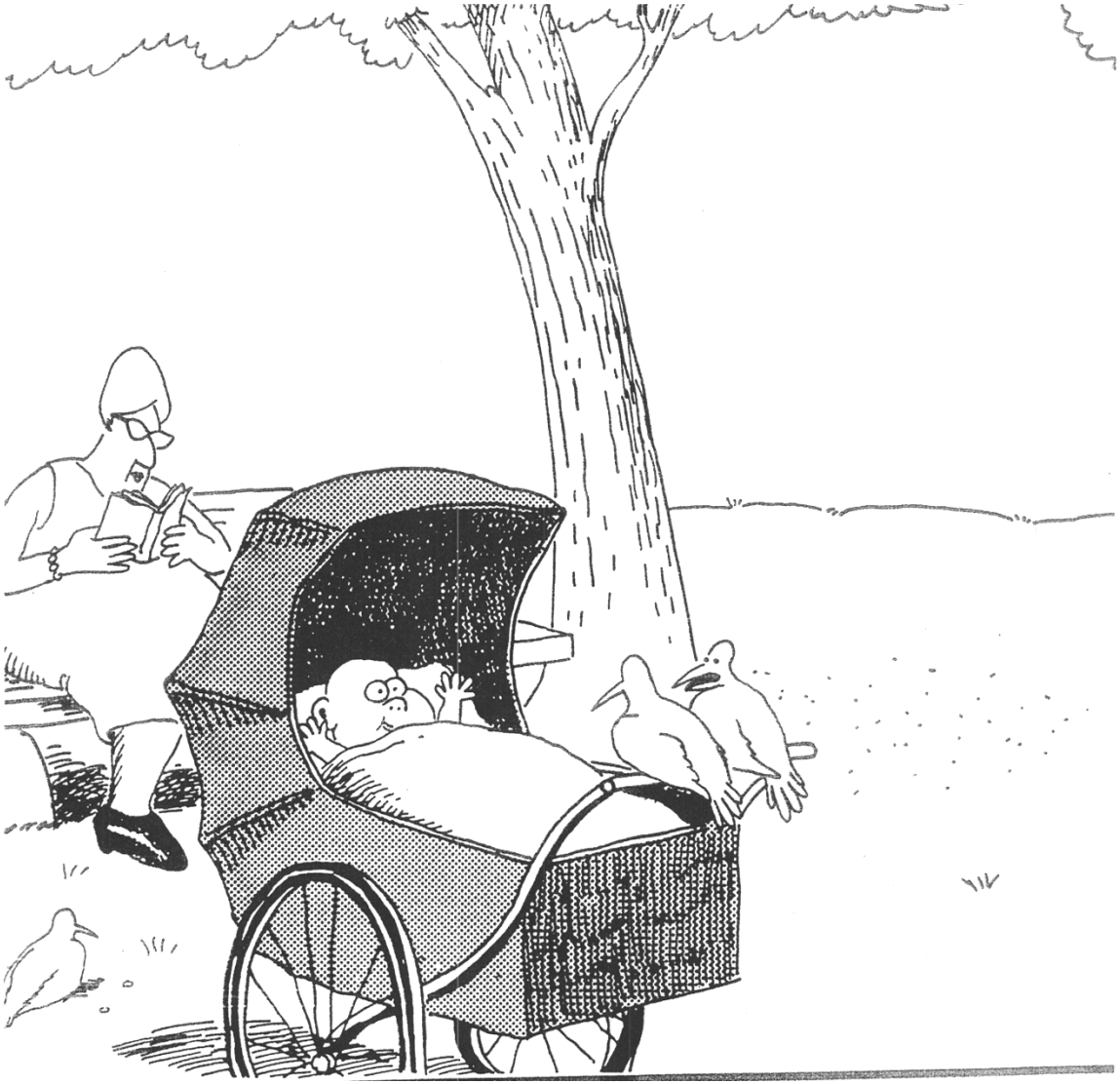
As move further up, can see farther than on lower points

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, richer perspective for interpreting previous answers.



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

We Need New Tools for the New Problems

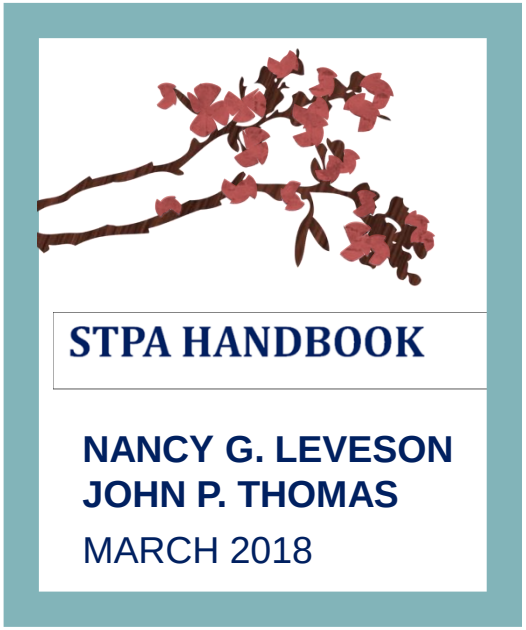
Is STAMP Practical?

- Tools have been or are being used in a large variety of industries
  - Automobiles (>80% use)
  - Aircraft and Spacecraft (extensive use and growing)
  - Defense systems (UAVs, AF GBSD, Army FVL, etc.)
  - Ships/Marine
  - Air Traffic Control
  - Medical Devices and Hospital Safety
  - Chemical plants
  - Oil and Gas
  - Nuclear and Electric Power
  - Robotic Manufacturing / Workplace Safety
- 2,316 registrants (73 countries) for STAMP Workshop this year
- New international standards (autos, aircraft, defense) created or in development or already satisfied (MIL-STD-882)

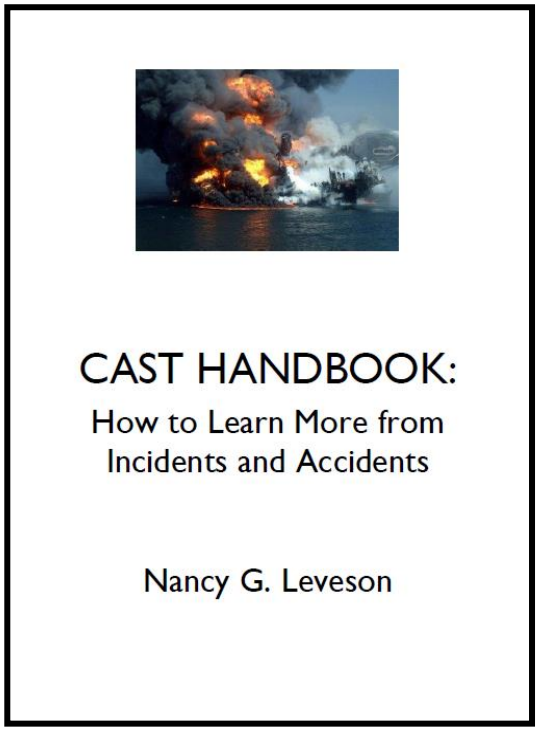
More Information

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

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



Free download:
http://psas.scripts.mit.edu
(**80,000+** downloads in 30 mos. Japanese, Chinese, and Korean versions)



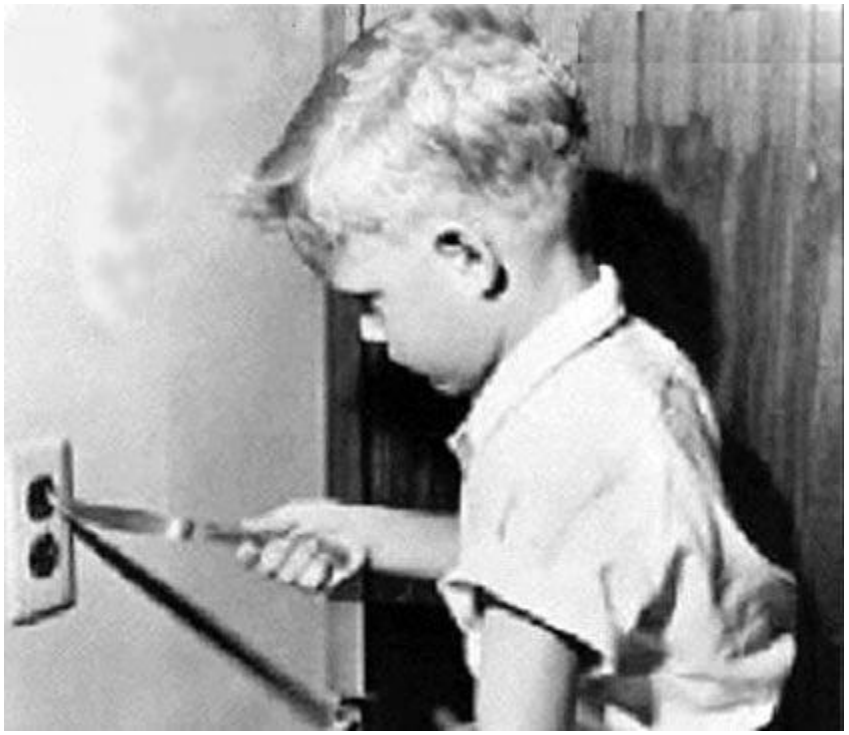
Free download:
http://sunnyday.mit.edu/CAST-Handbook.pdf

Safety-II claims and why it is dangerous

- Safety-I doesn't exist and never has except perhaps long ago in workplace safety, not in product/system safety
  - Strawman argument
  - Nobody relies on accident investigation: 90% of effort on prevention
- Not a sociotechnical approach
- Not a systems approach
- All terms are mis-defined, even the math ones
- Safety-II suggests:
  - Do proactive analysis: already do exactly what he suggests only better (he proposes non-rigorous methods to do what we already have much better, more rigorous methods to achieve)
  - Learn from success, not failure
    - Learn very little from success and usually learn the wrong thing
    - Engineering learns from failures
- FRAM problems (a specification language, and an old one at that, not a hazard analysis or analysis of any kind. Nothing to do with safety)

Safe or Unsafe?

Safety Depends on Context



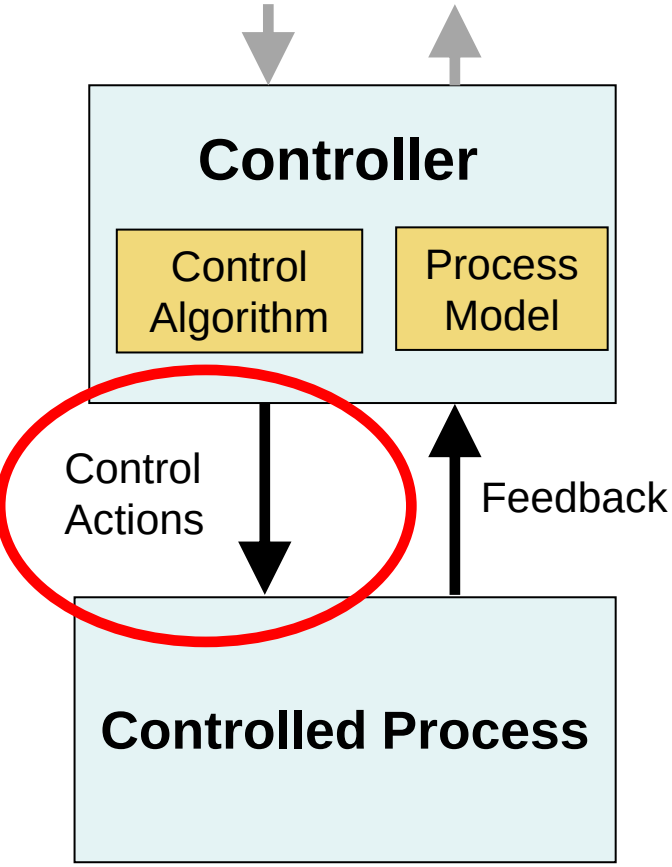
Summary

- Accidents today do not just result from component failures. Need to consider design errors
- Software
  - Contributes differently to accidents than hardware
    - Does not “fail” but can contribute to unsafe system behavior (including unsafe human behavior)
  - Adds almost unlimited complexity but
    - Cannot exhaustively test
    - Is not by itself safe or unsafe
- Safety depends on context

Two Types of Accidents

- **Component Failure Accidents**
  - Single or multiple component failures
  - Usually assume random failure
- **Component Interaction Accidents**
  - Arise in interactions among components
  - Related to complexity (coupling) in our system designs, which leads to design and system engineering errors
  - No components may have “failed”
  - Exacerbated by introduction of computers and software but the problem is system design errors
    - Software allows almost unlimited complexity in our designs

Hazard and Accident Analysis with STAMP



Analysis:

1. Identify potential unsafe control actions
2. Identify why they might be (or were) given (eliminate or mitigate)
3. If safe ones provided, then why not followed?

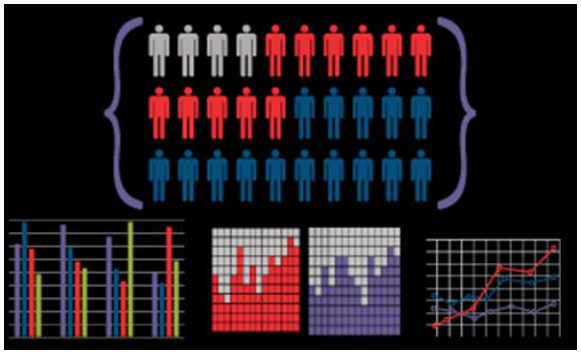
Epidemiological Model (1)

- John Gordon 1949: A Harvard University epidemiologist
- Stresses multifactorial nature of accidents
- Accidents conceptualizes in terms of
  1. Agent (physical energy)
  2. Environment
  3. Host (victim)

- Result from complex interactions between these three things. Cannot be explained by
  - Considering only one of these three factors
  - By simple linear interactions between events

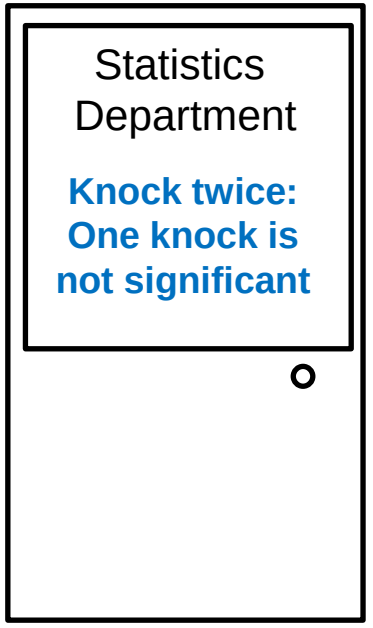
Epidemiological Model (2)

- Two types:
  - Descriptive epidemiology: determine incidence, prevalence, and mortality rates for accidents in large population groups according to characteristics (age, sex, geographical area)
  - Investigative epidemiology: collect specific data on causes of injuries to devise feasible countermeasures.
- Assumes common factors present and these can be determined by statistical evaluation of accident data.
- No assumption about specific relationships between factors, previously unrecognized relationships can be discovered.



Epidemiological Model (3)

- Not widely used: dependent on
  - Quality of database used
  - Statistical significance of anomalies found in sample
- Limitations:
  - Data reported by accident investigators may be limited or filtered
  - Important relationships (linear and nonlinear) may not be captured by a purely statistical approach.



Safety as a 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