Engineering a Safer World

Prof. Nancy Leveson
Massachusetts Institute of Technology

Why Our Efforts are Often Not Cost-Effective

- Efforts superficial, isolated, or misdirected
- Too much effort on assuring system safe vs. designing it to be safe
- Safety efforts start too late
- Inappropriate techniques for systems built today
- Focus efforts only on technical components of systems
- Systems assumed to be static through lifetime
- Limited learning from events

Traditional Approach to Safety

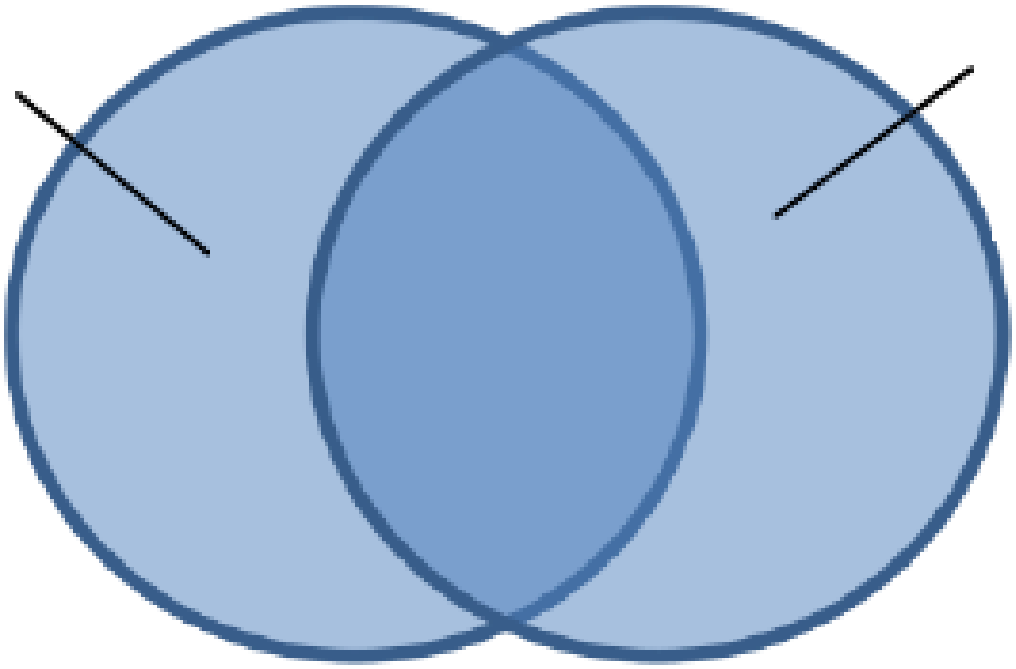
- Traditionally view safety as a failure problem
  - Chain of directly related failure events leads to loss
  - Establish barriers between events or try to prevent individual component failures

    e.g., redundancy, overdesign, safety margins, reward and punishment

Limitations of Traditional Approach

- Systems are becoming more complex
  - Accidents often result from interactions among components, not just component failures
  - Too complex to anticipate all potential interactions
    - By designers
    - By operators
  - Indirect and non-linear interactions
- Omits or oversimplifies important factors
  - Human error
  - New technology, particularly software
  - Culture and management
  - Evolution and adaptation

Confusing Safety and Reliability

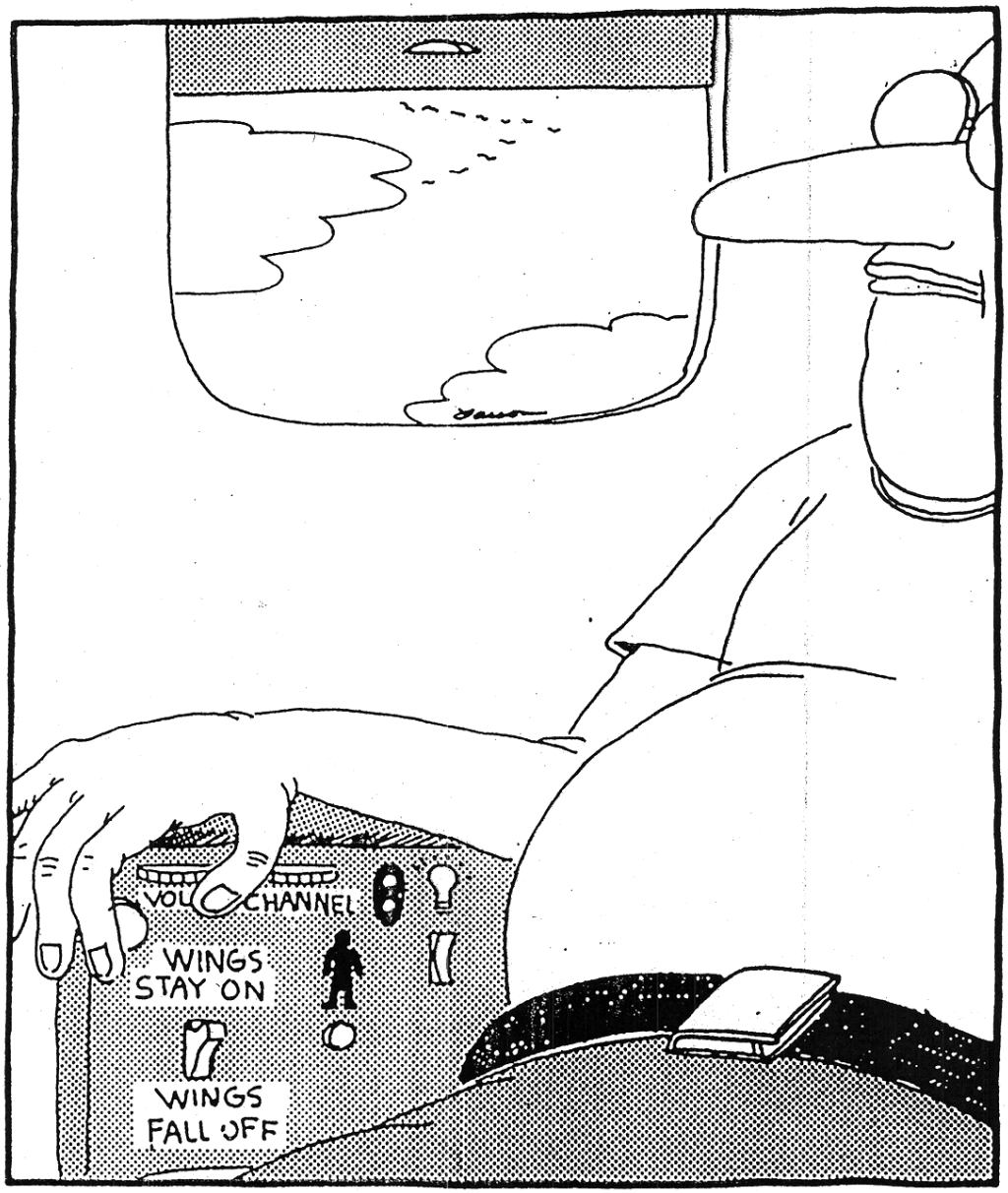


Software-Related Accidents

- Are usually caused by flawed requirements
  - Incomplete or wrong assumptions about operation of controlled system or required operation of computer
  - Unhandled controlled-system states and environmental conditions
- Merely trying to get the software “correct” or to make it reliable will not make it safer under these conditions.

Operator Error: Traditional View

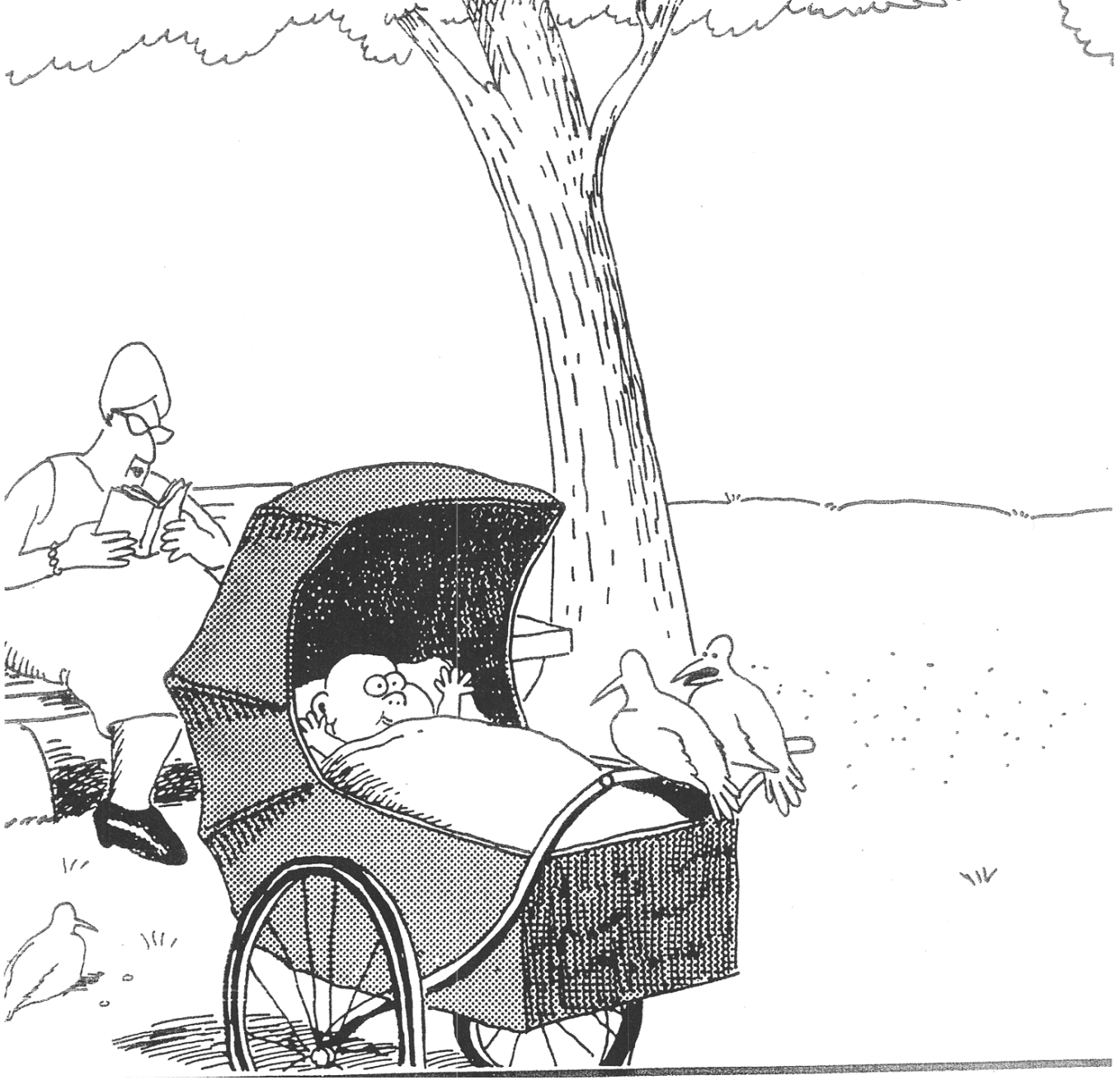
- Human error is cause of incidents and accidents
- So do something about human involved (suspend, retrain, admonish)
- Or do something about humans in general
  - Marginalize them by putting in more automation
  - Rigidify their work by creating more rules and procedures



Fumbling for his recline button Ted unwittingly instigates a disaster

Operator Error: Systems View (Dekker, Rasmussen, etc.)

- Human error is a symptom, not a cause
- All behavior affected by context (system) in which occurs
- Role of operators in our systems is changing
  - Supervising rather than directly controlling
  - Systems are stretching limits of comprehensibility
  - Designing systems in which operator error inevitable and then blame accidents on operators rather than designers
- To do something about 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

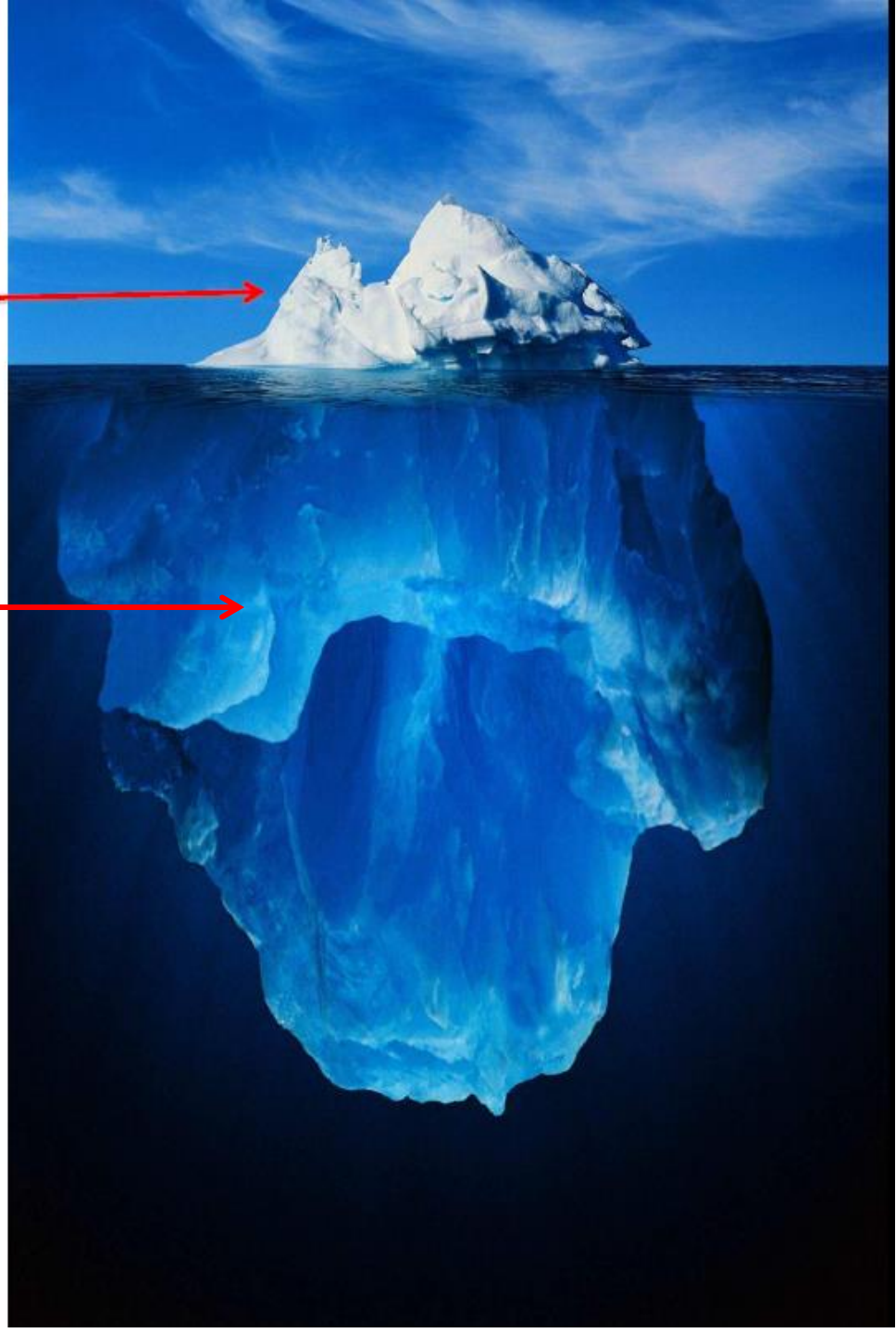


It's still hungry … and I've been stuffing worms into it all day.

LEVERAGE

Event-based thinking

Systems Thinking



WYOUNG@MIT.EDU

STAMP: System-Theoretic Accident Model and Processes

Based on Systems Theory
(vs. Reliability Theory)

Applying Systems Thinking to Safety

- 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

Safety as a Dynamic Control Problem

- Examples
  - O-ring did not control propellant gas release by sealing gap in field joint of Challenger Space Shuttle
  - Software did not adequately control descent speed of Mars Polar Lander
  - At Texas City, did not control the level of liquids in the ISOM tower;
  - In DWH, did not control the pressure in the well;
  - Financial system did not adequately control the use of financial instruments

Safety as a Dynamic Control Problem (2)

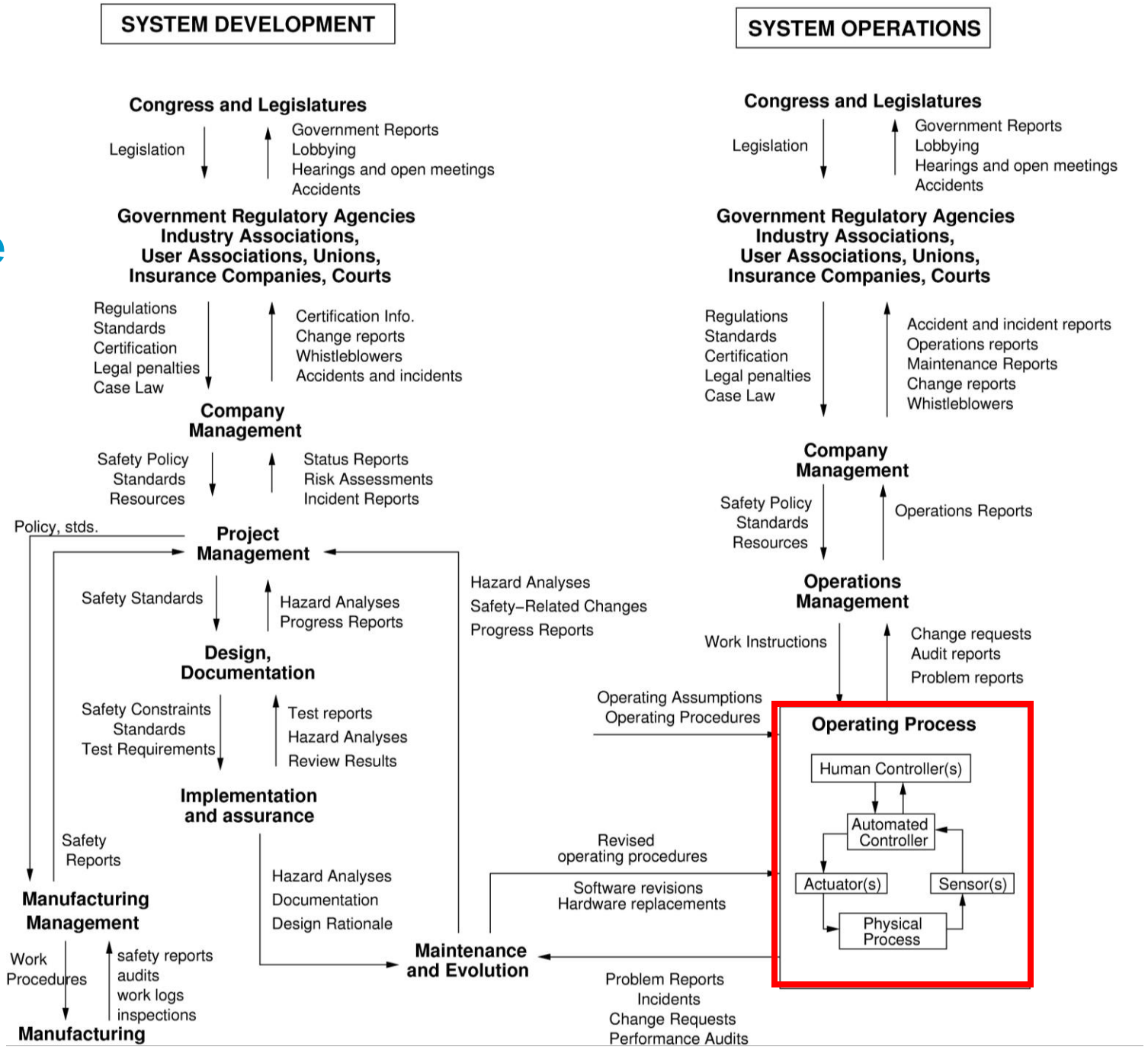
- Events are the result of the inadequate control
  - Result from lack of enforcement of safety constraints in system design and operations
- Most major accidents arise from a slow migration of the entire system toward a state of high-risk
  - Need to control and detect this migration
- A change in emphasis:

~~"prevent failures"~~

↓

"enforce safety constraints on system behavior"

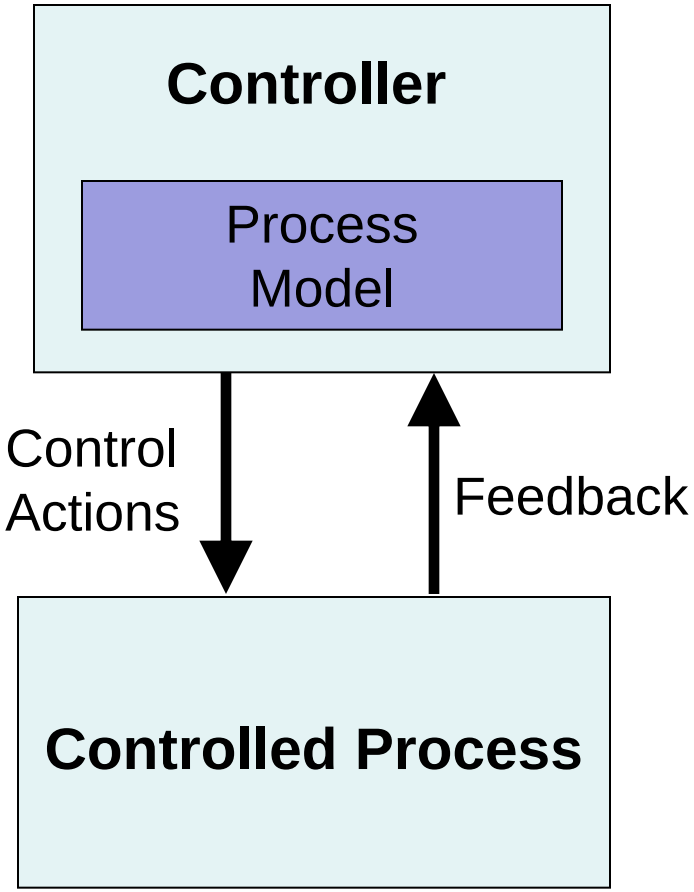
Example Safety Control Structure



Safety as a Control Problem (3)

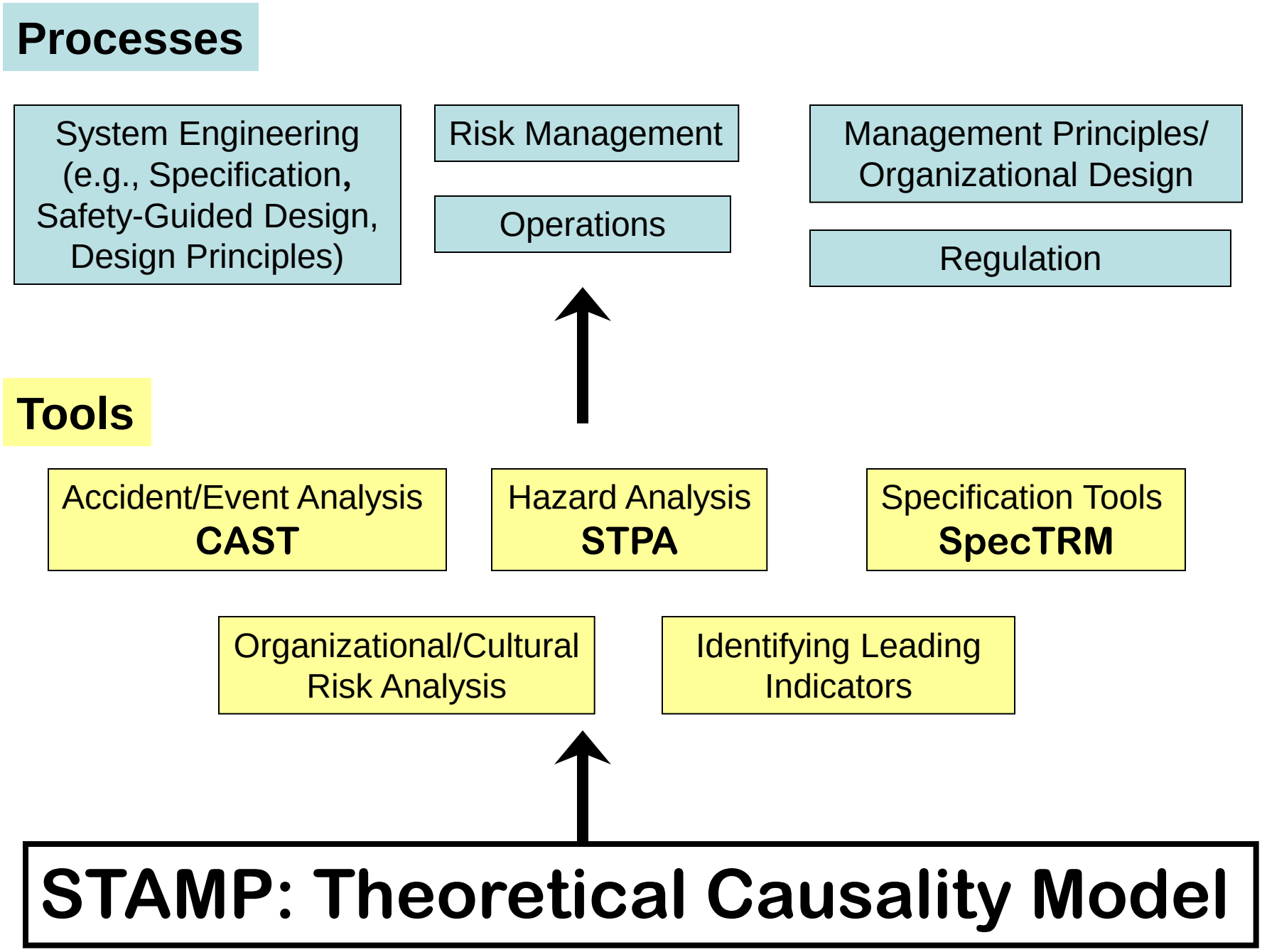
- **Goal: Design an effective control structure that eliminates or reduces adverse events**.
  - Need clear definition of expectations, responsibilities, authority, and accountability at all levels of safety control structure
  - Entire control structure must together enforce the system safety property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

Systems approach to safety engineering (STAMP)



- Controllers use a **process model** to determine control actions
- Accidents often occur when the process model is incorrect
- Four types of hazardous control actions:
  - Control commands required for safety are not given
  - Unsafe ones are given
  - Potentially safe commands given too early, too late
  - Control stops too soon or applied too long

(Leveson, 2003); (Leveson, 2011)


Processes
System Engineering (e.g., Specification, Safety-Guided Design, Design Principles)
Risk Management
Operations
Management Principles/ Organizational Design
Regulation
Tools
Accident/Event Analysis
CAST
Hazard Analysis
STPA
Specification Tools
SpecTRM
Organizational/Cultural Risk Analysis
Identifying Leading Indicators
STAMP: Theoretical Causality Model

STPA: System Theoretic Process Analysis

(A New Hazard Analysis Technique)

STPA (System-Theoretic Process Analysis)

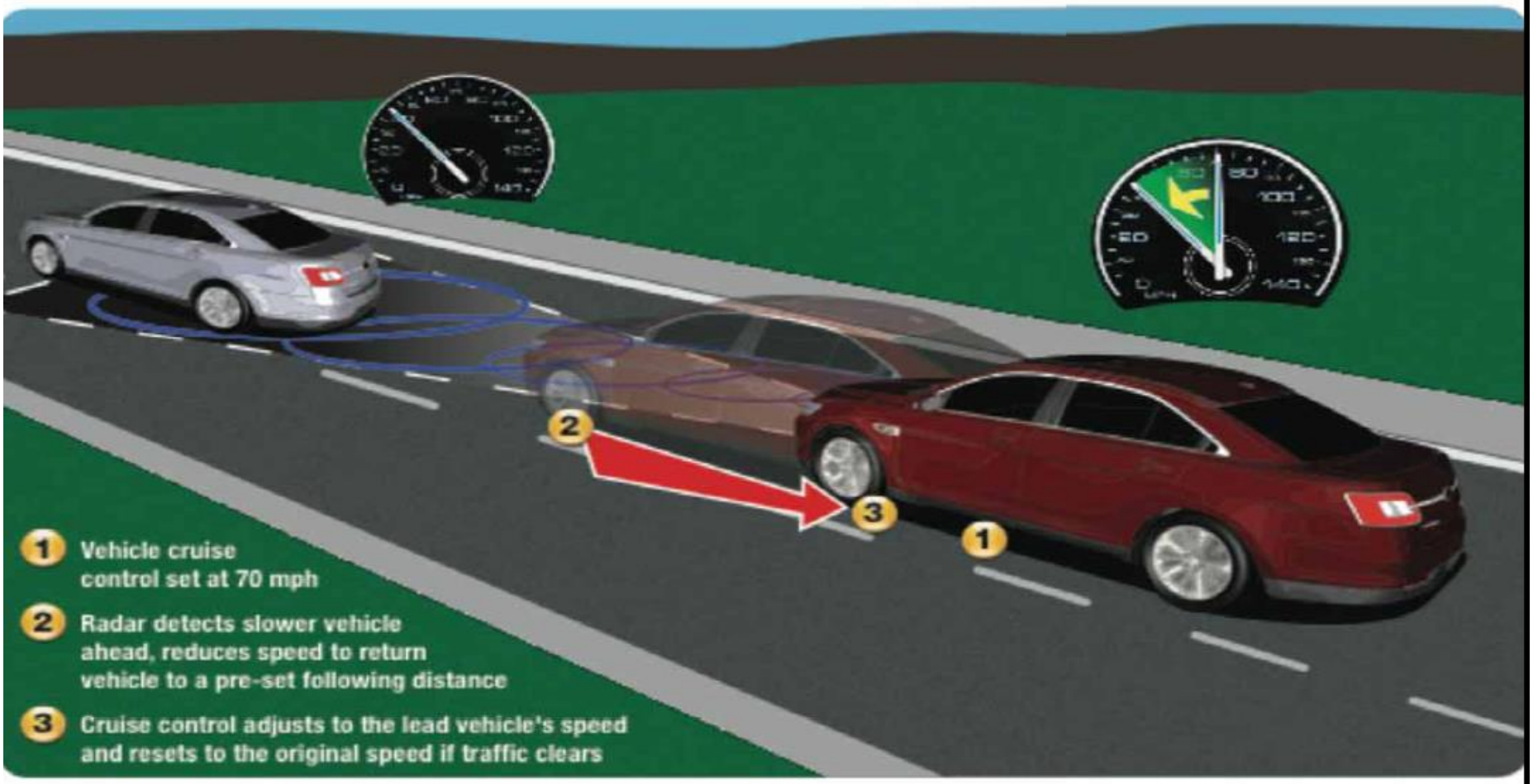
- Starts from hazards
- Identifies safety constraints (system and component safety requirements)
- Identifies scenarios leading to violation of safety constraints
- Can be used on technical design and organizational design
- Supports a safety-driven design process where
  - Hazard analysis influences and shapes early design decisions
  - Hazard analysis iterated and refined as design evolves

Unsafe Control Actions

Four Ways Unsafe Control Can Occur

- A control action required for safety is not provided or is not followed
- An unsafe control action is provided that leads to a hazard
- A potentially safe control action provided too late, too early, or out of sequence
- A safe control action is stopped too soon or applied too long (for a continuous or non-discrete control action)

ADAPTIVE CRUISE CONTROL



Source: media.ford.com

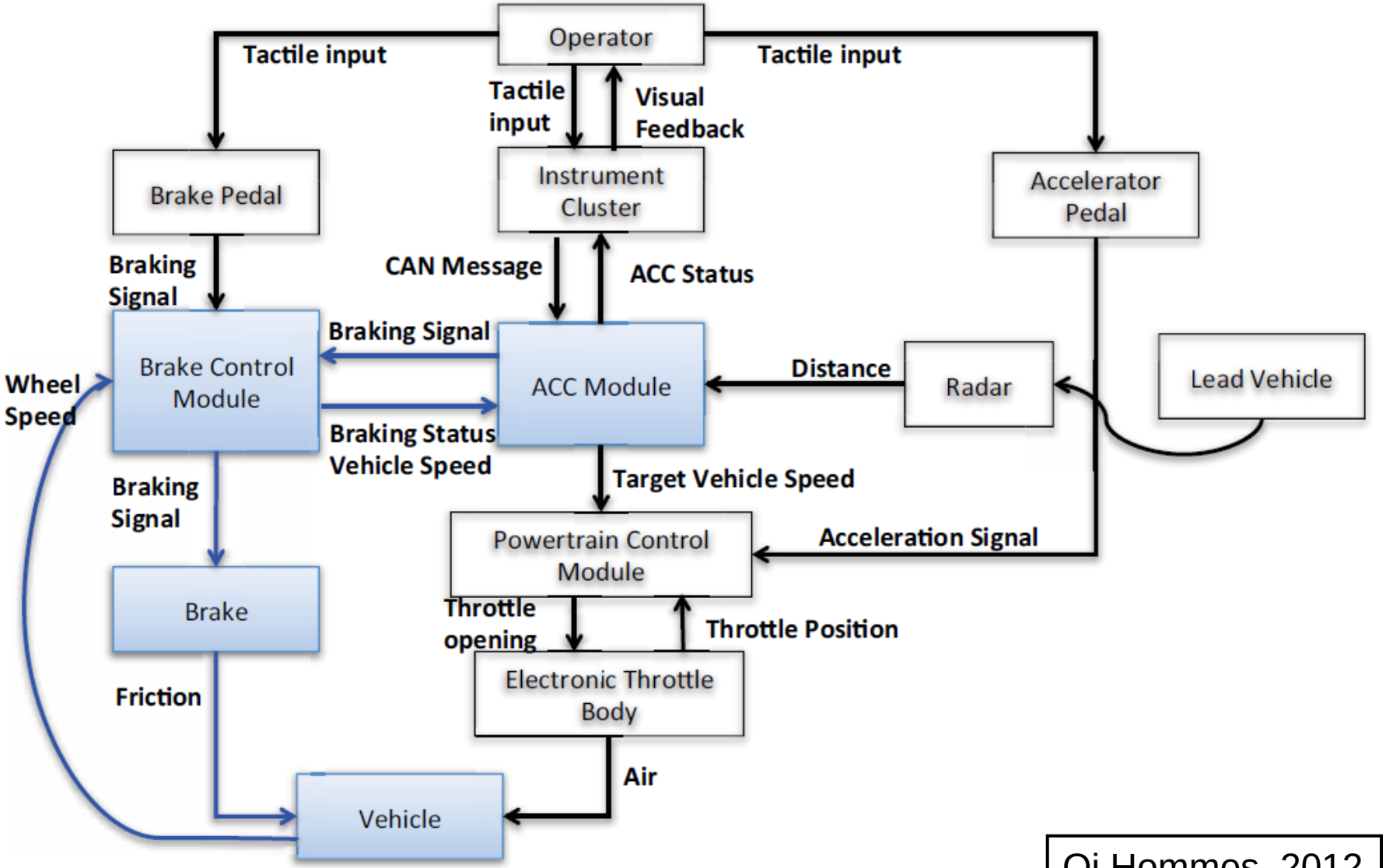
Qi Hommes, 2012

Accidents and Hazards

- **Accident**: Vehicle occupants are injured while ACC is engaged
- **Hazards**:
  - H1: ACC does not maintain a safe distance from the object in the front (resulting in a collision)
  - H2: ACC slows down the vehicle too abruptly (and vehicle is rear-ended).
- **Safety Requirements/Constraints**
  - ACC must not violate separation requirements with object ahead
  - ACC must not brake too abruptly

Qi Hommes, 2012

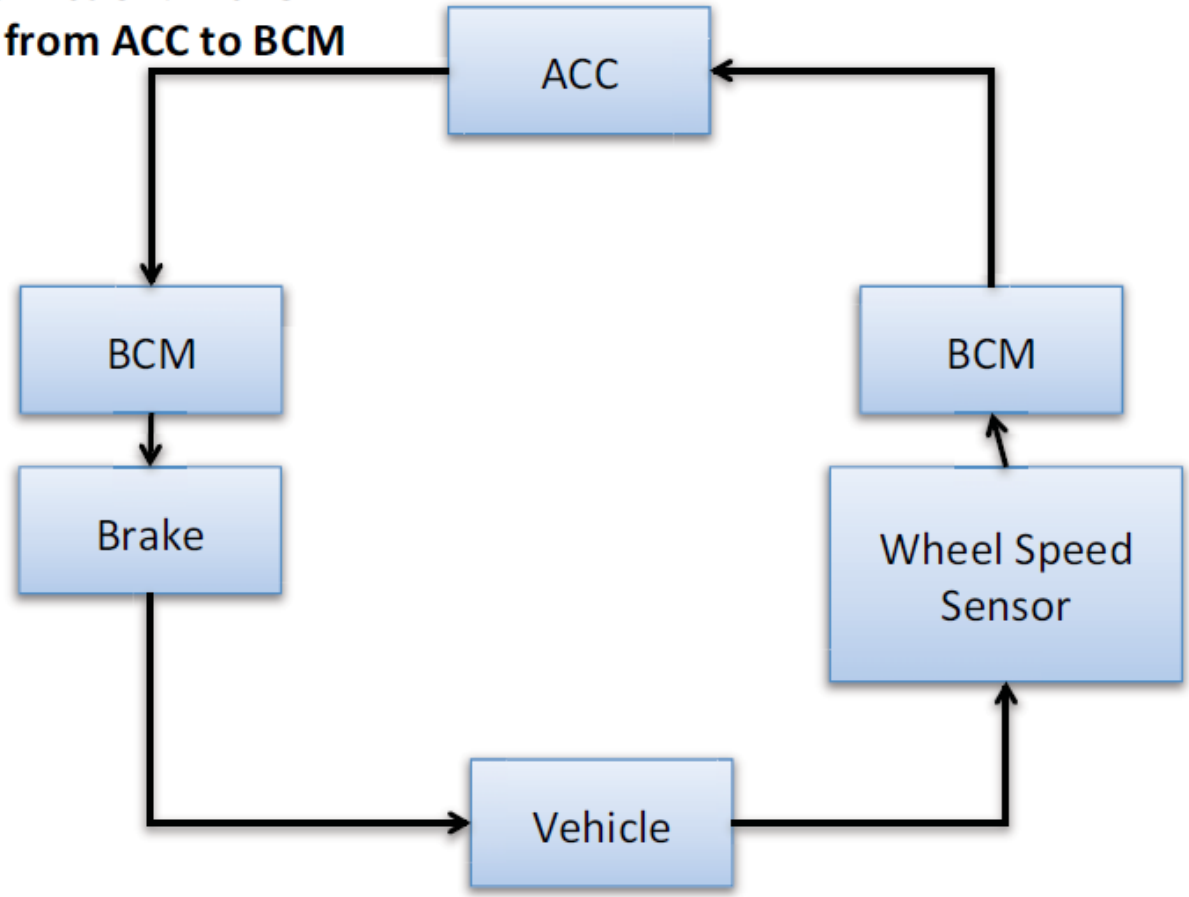
Example: ACC – BCM Control Loop



Qi Hommes, 2012

Reformatted Control Loop

Control Action: Brake
Signal from ACC to BCM



Qi Hommes, 2012

STPA Step 1: Unsafe Control Actions

Control Action	Not Providing Causes Hazard	Providing Causes Hazard	Wrong Timing or Order Causes Hazard	Stopped too Soon or Applied Too Long
Brake Signal from ACC to BCM	Vehicle does not brake when the distance to the lead vehicle is less than the value set by the operator. (H1)	Commanded decceleration amount is too small when the vehicle is too close to the object in the front. (H1)	Braking is commanded too late when the distance to the lead vehicle is too close. (H1)	Braking stops before the safety distance between the vehicles are reached. (H1)
		Braking is commented when the distance to the lead vehicle is larger than the set value. (H2)		
		Braking is too fast/harsh when the didstance to the lead vehicle is less than the set value. (H2)		

Qi Hommes, 2012

Generating Refined Safety Requirements

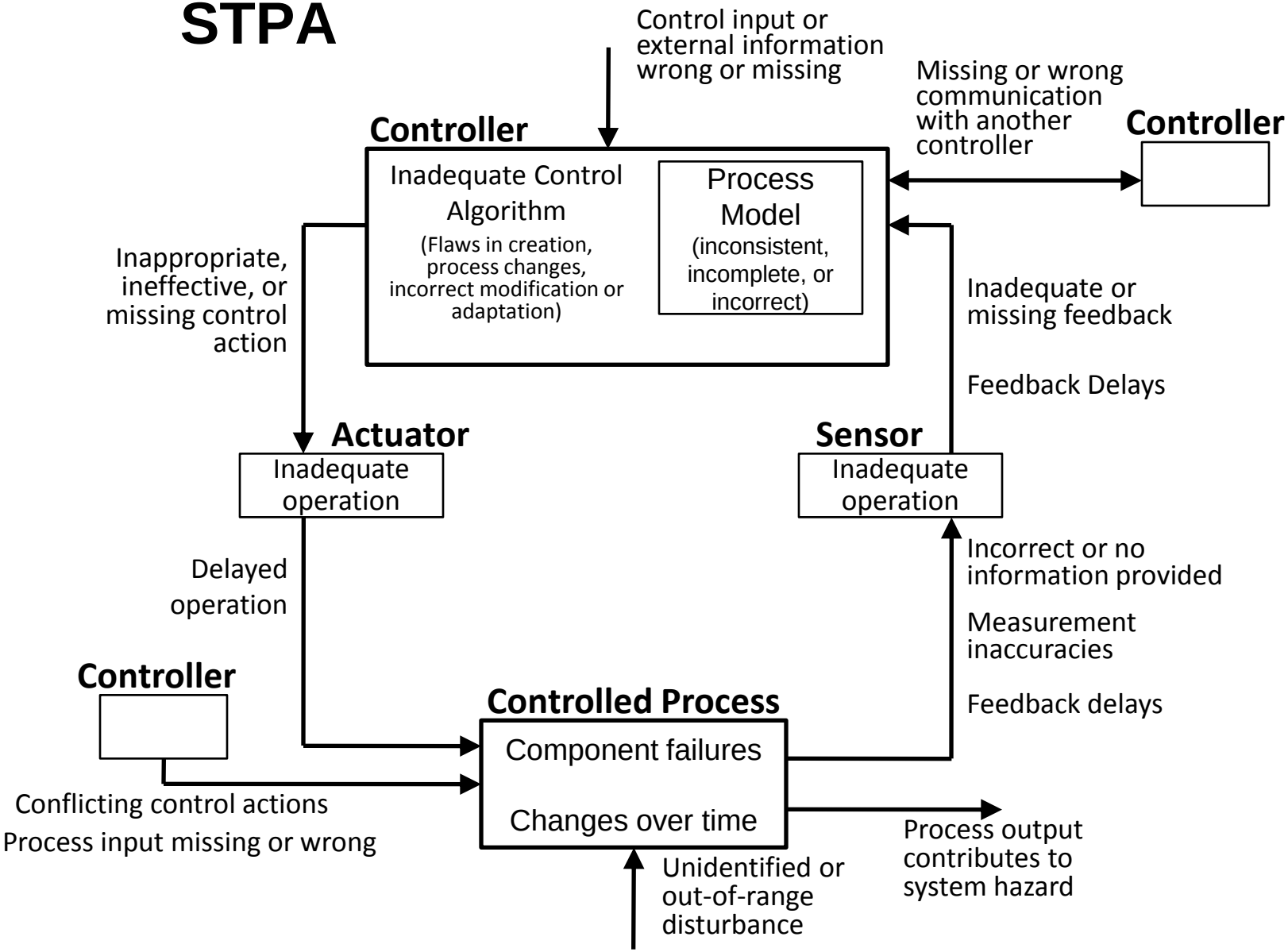
- Use the unsafe control actions in the table to refine the high-level system and component functional requirements
  - ACC shall maintain a TBD amount of distance between the vehicle and the object in front when engaged
  - ACC shall limit vehicle acceleration to no more than TBC m/s^2
  - …
- But not enough …

Qi Hommes, 2012

STPA Step 2

- Identify detailed causal scenarios leading to violation of safety requirements (constraints)
- Will identify more detailed (refined) safety-related requirements
- Again, use to improve design

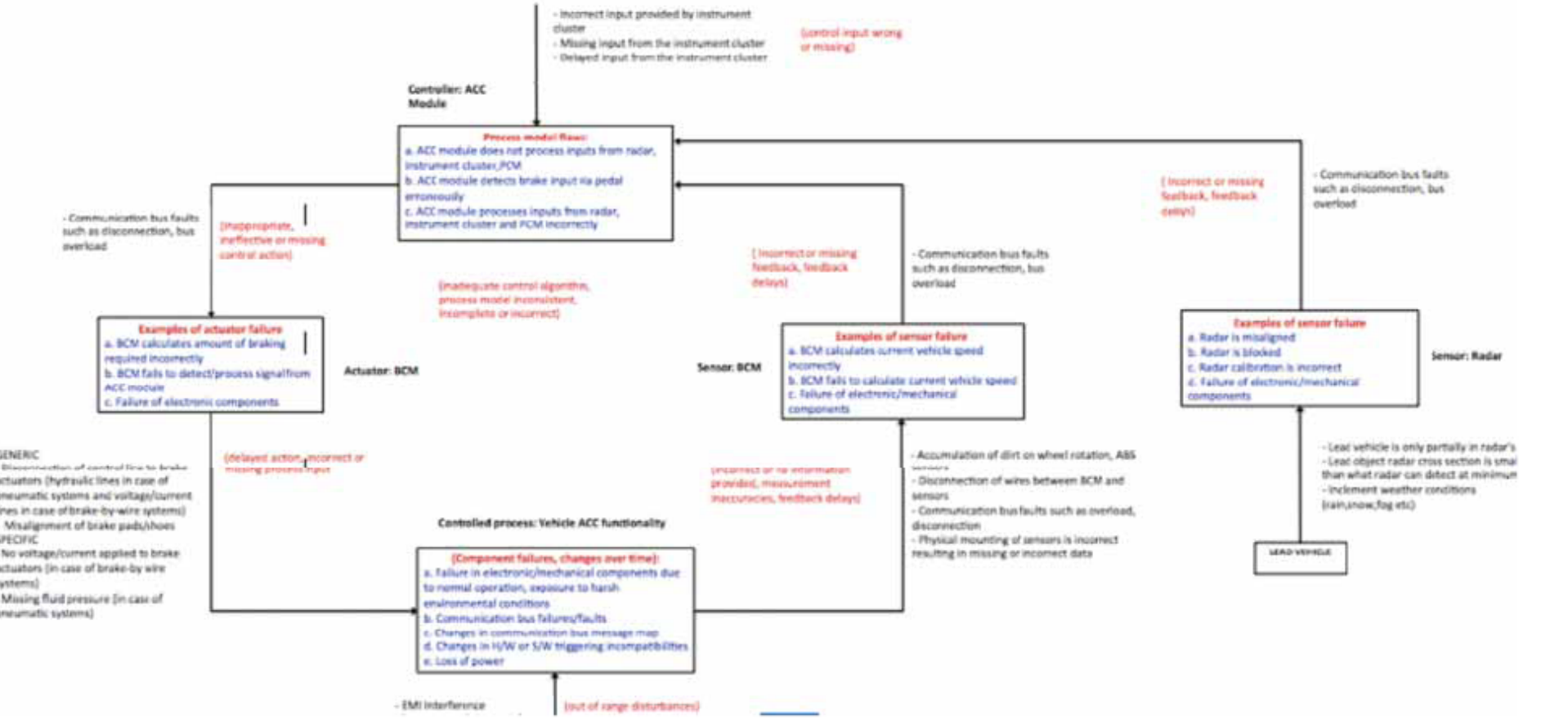
STPA


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
Sensor
Inadequate operation
Inadequate operation
Delayed operation
Incorrect or no information provided
Measurement inaccuracies
Feedback delays
Controller
Controlled Process
Component failures
Changes over time
Conflicting control actions
Process input missing or wrong
Process output contributes to system hazard
Unidentified or out-of-range disturbance

Causal Analysis Results

nsafe Control Action:

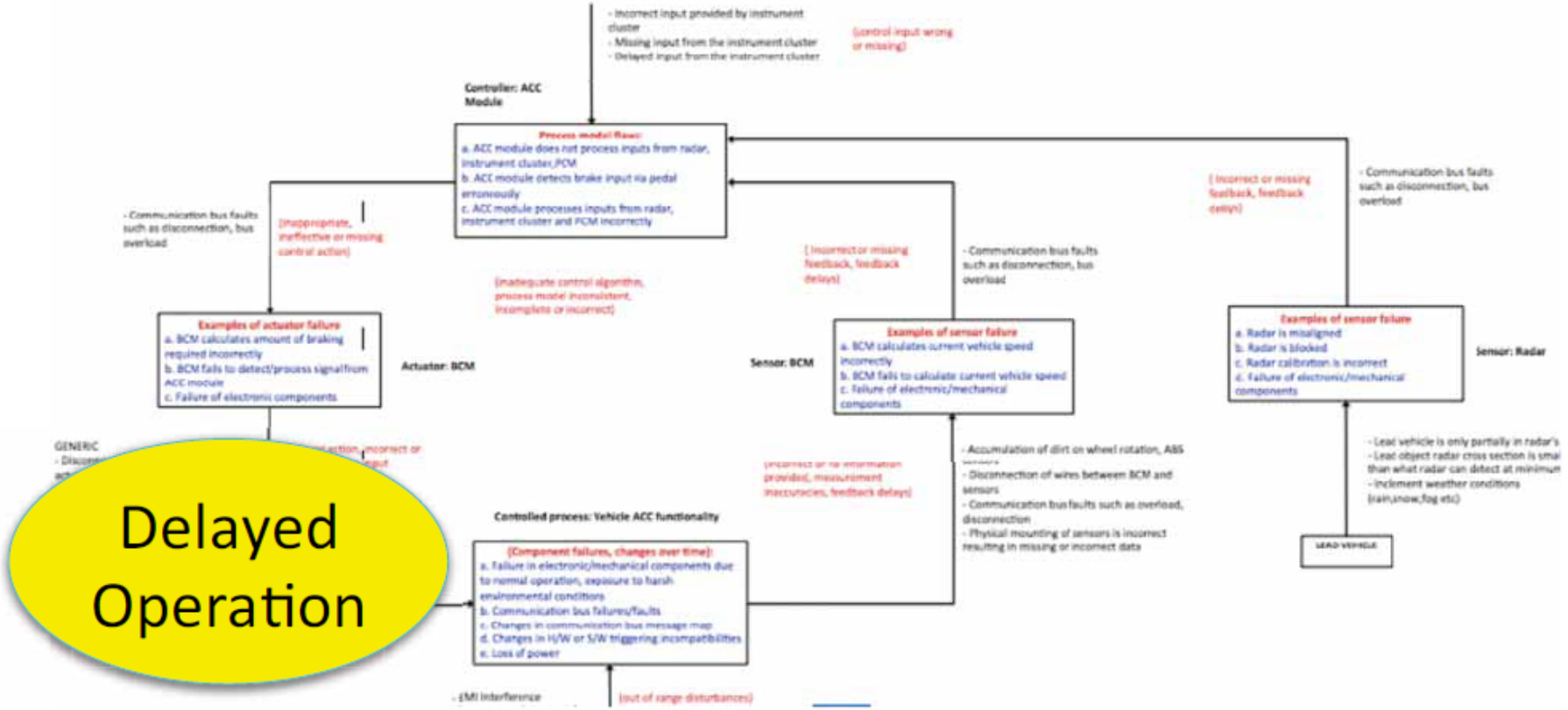
ehicle does not brake when the distance to the object in front is less than preset value.



Qi Hommes, 2012

Causal Analysis Results (2)

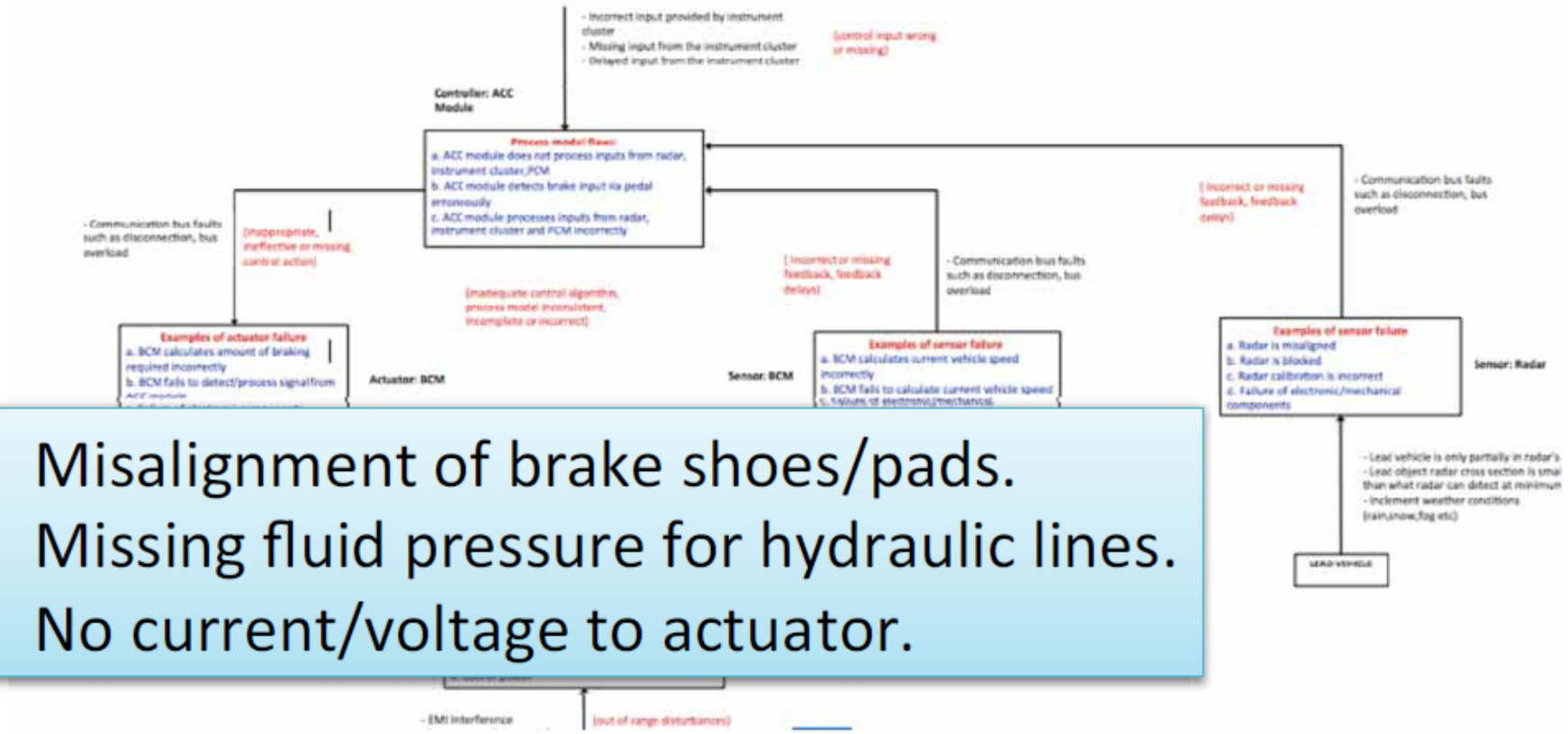
Unsafe Control Action:
Vehicle does not brake when the distance to the object in front is less than preset value.



Qi Hommes, 2012

Causal Analysis Results (3)

Unsafe Control Action:
Vehicle does not brake when the distance to the object in front is less than preset value.

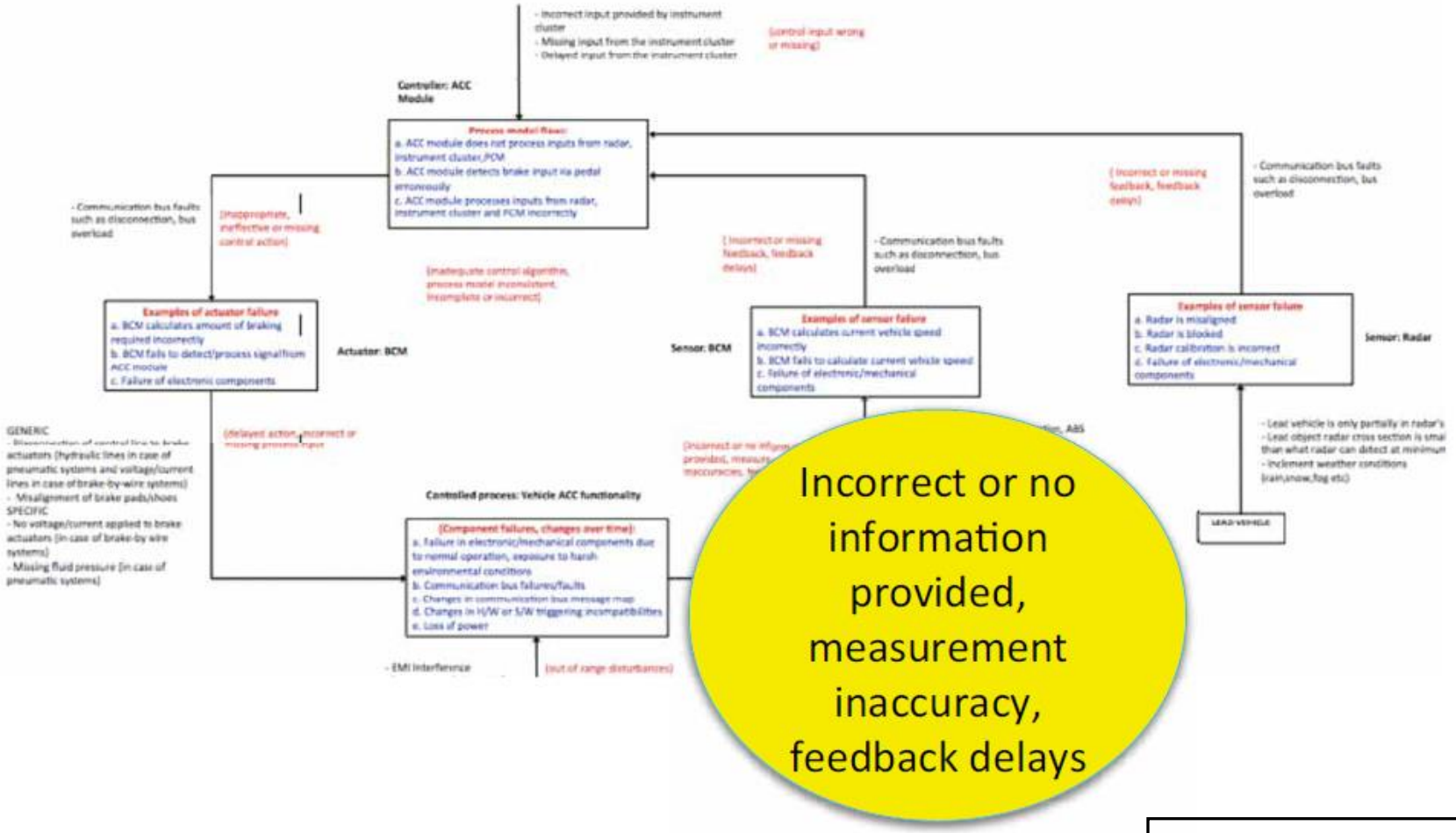


- Misalignment of brake shoes/pads.
- Missing fluid pressure for hydraulic lines.
- No current/voltage to actuator.

Qi Hommes, 2012

Causal Analysis Results (4)

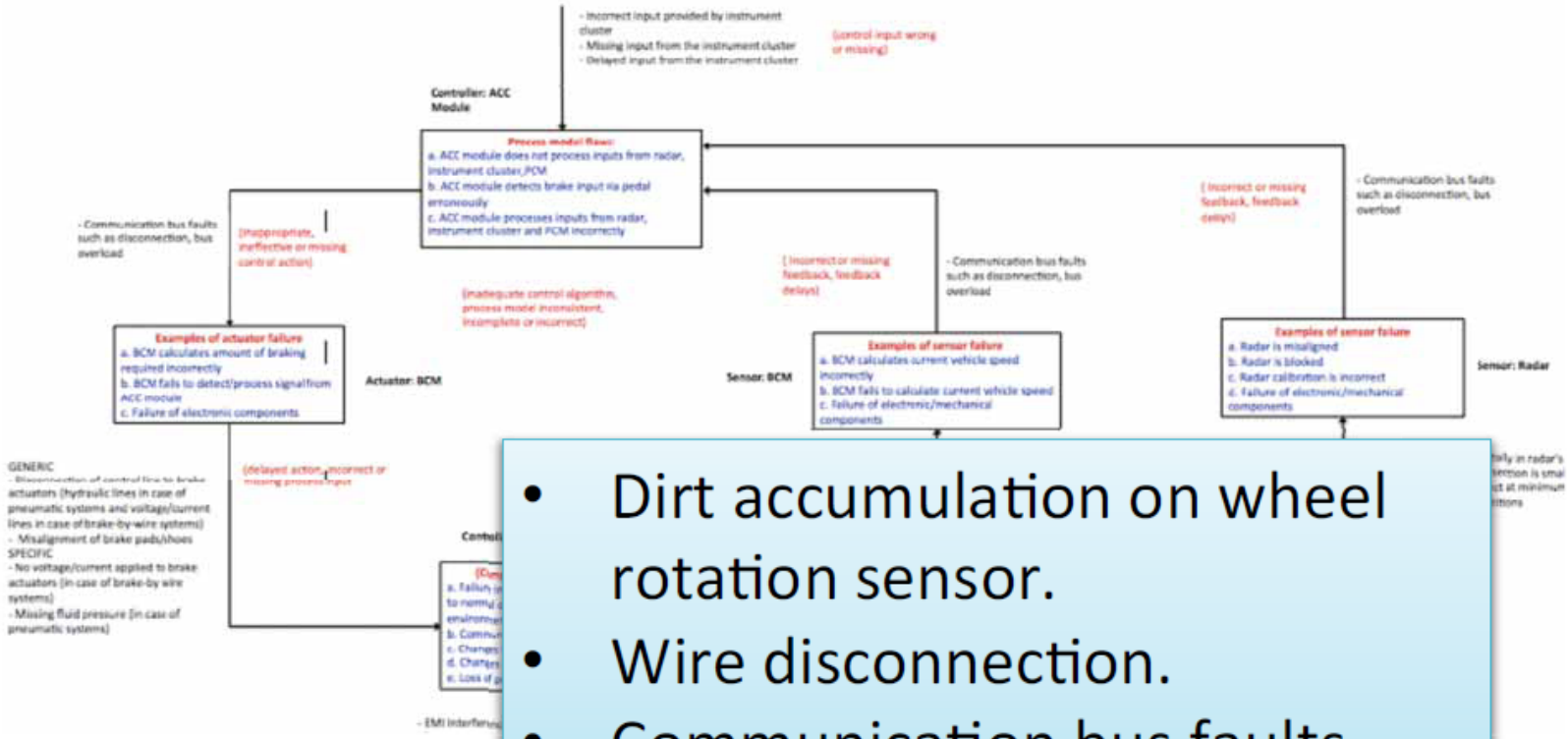
Unsafe Control Action:
Vehicle does not brake when the distance to the object in front is less than preset value.



Qi Hommes, 2012

Causal Analysis Results (5)

Unsafe Control Action:
Vehicle does not brake when the distance to the object in front is less than preset value.



- Dirt accumulation on wheel rotation sensor.
- Wire disconnection.
- Communication bus faults, overload, message priority.

Qi Hommes, 2012

Example Causal Scenarios for Radiation Treatment

- **Scenario 1** - Operator was expecting patient to have been positioned, but table positioning was delayed compared to plan (e.g. because of delays in patient preparation or patient transfer to treatment area; because of unexpected delays in beam availability or technical issues being processed by other personnel without proper communication with the operator).

- **Controls:**
  - Provide operator with direct visual feedback to the gantry coupling point, and require check that patient has been positioned before starting treatment (M1).
  - Provide a physical interlock that prevents beam-on unless table positioned according to plan

Example Causal Scenarios (2)

- **Scenario 2** - Operator is asked to turn the beam on outside of a treatment sequence (e.g. because the design team wants to troubleshoot a problem) but inadvertently starts treatment and does not realize that the facility proceeds with reading the treatment plan.

- **Controls**:
  - Reduce the likelihood that non-treatment activities have access to treatment related input by creating a non-treatment mode to be used for QA and experiments, during which facility does not read treatment plans that may have been previously been loaded (M2);
  - Make procedures (including button design if pushing a button is what starts treatment) to start treatment sufficiently different from non-treatment beam on procedures that the confusion is unlikely.

Tools to Help with STPA

- Thomas has defined a procedure and is prototyping automation to help perform STPA
  - Uses a model-based requirements development toolset called SpecTRM
  - Generates model-based requirements from hazard analysis
- Additional tools being developed by Qi Hommes at Volpe
- Antoine: Ways to organize the causal scenarios generated in Step 2
- Visualization tools

Evaluation on Real Systems

- Non-advocate safety assessment of U.S. Ballistic Missile Defense System
  - 2 people for 3 months
  - Deployment and testing held up for 6 months because so many scenarios identified for inadvertent launch.
  - In many of these scenarios:
    - All components were operating exactly as intended but complexity of component interactions led to unanticipated system behavior
    - Examples: missing case in software requirements, timing problem in sending and receiving messages, etc.
  - STPA also identified component failures that could cause inadequate control (most analysis techniques consider only these failure events)

Evaluating STPA on Real Systems (2)

- JAXA HTV
  - Found everything found in fault tree analysis and more (mostly related to system design and software)
- NextGen In-Trail Procedure (Air Traffic Control)
  - Hard to compare but we found more scenarios than their fault tree and event tree mix
- Nuclear Power Plants
  - Experimental comparison performed by EPRI and experts on each technique
  - Results not available yet but informally STPA was only one that found a real accident scenario that had occurred (and none of analysts knew about)

Evaluating STPA on Real Systems (3)

- Blood Gas Analyzer (Vincent Balgos)
  - 75 scenarios found by FMEA
  - 175 identified by STPA
  - Took much less time and resources (mostly human)
  - Only STPA found scenario that had led to a Class 1 recall by FDA (actually found nine scenarios leading to it)
- Proton Radiation Therapy (Gantry 2): Blandine Antoine, Martin Rejzak, Christian Hilbes
- Lots more in all kinds of industries
- Biggest surprise (to me) was required much less resources

Use Without Evaluation

- Medtronic Artificial Pancreas
- Nuclear Power Plant for U.S. NRC
- CO_2 Capture, Transport, and Storage
- Automotive problems
- JAXA new manned spacecraft (Safety-Guided Design)
- Large Oil & Gas Engineering Consulting Firm
- NextGen TBO (PHA, Safety-Guided Design)
- Integrated Modular Avionics
  - Interoperability (Consistency Analysis)
  - Change analysis

Learning from Events

- CAST: Causal Analysis based on System Theory
- Goal: more complete causal analysis of accidents, incidents, and adverse events

Learning from Events

- Non-serious events and incidents are a precious opportunity – we too often waste them.
- “Operator error” is a useless finding
  - Focus on “why” not “who” or “what”
  - Blame is the enemy of safety
- Root cause seduction

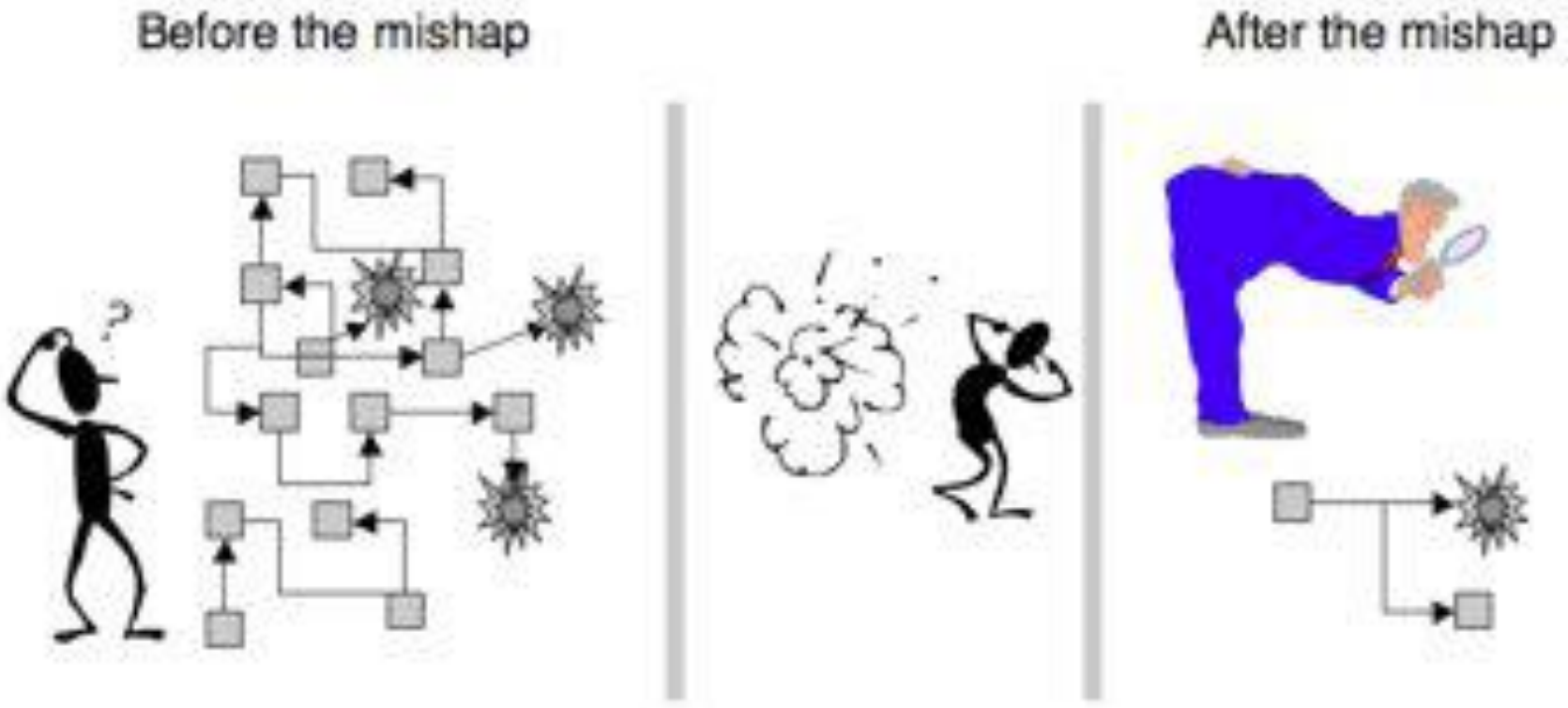
Root Cause Seduction

- Assuming there is a root cause gives us an illusion of control.
  - Usually focus on operator error or technical failures
  - Ignore systemic and management factors
  - 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

Three Levels of Analysis

- What (events)
  - e.g., explosion
- Who and how (conditions)
  - e.g., bad valve design, operator did not notice something
- Why (systemic factors)
  - e.g., production pressures, cost concerns, flaws in design process, flaws in reporting process, etc.
  - Why was safety control structure ineffective in preventing the loss?

Hindsight Bias



Sidney Dekker, 2009

Hindsight Bias

- After an incident
  - Easy to see where people went wrong, what they should have done or avoided
  - Easy to judge about missing a piece of information that turned out to be critical
  - Easy to see what people should have seen or avoided
- Almost impossible to go back and understand how world looked to somebody not having knowledge of outcome

Overcoming Hindsight Bias

- Assume nobody comes to work to do a bad job.
  - Assume were doing reasonable things given the complexities, dilemmas, tradeoffs, and uncertainty surrounding them.
  - Simply finding and highlighting people's mistakes explains nothing.
  - Saying what did not do or what should have done does not explain why they did what they did.
- Investigation reports should explain
  - Why it made sense for people to do what they did rather than judging them for what they allegedly did wrong and
  - What changes will reduce likelihood of happening again

CAST (Causal Analysis using STAMP)

- Identify system hazard violated and the system safety design constraints
- Construct the safety control structure as it was designed to work
  - Component responsibilities (requirements)
  - Control actions and feedback loops
- For each component, determine if it fulfilled its responsibilities or provided inadequate control.
  - If inadequate control, why? (including changes over time)
    - Context
    - Process Model Flaws
- For humans, why did it make sense for them to do what they did (to reduce hindsight bias)

CAST (2)

- Examine coordination and communication
- Consider dynamics and migration to higher risk
- Determine the changes that could eliminate the inadequate control (lack of enforcement of system safety constraints) in the future.
- Generate recommendations
- Continuous Improvement
  - Assigning responsibility for implementing recommendations
  - Follow-up to ensure implemented
  - Feedback channels to determine whether changes effective
    - If not, why not?

Evaluating CAST on Real Accidents

- Used on many types of accidents
  - Aviation
  - Trains (Chinese high-speed train accident)
  - Chemical plants and off-shore oil drilling
  - Road Tunnels
  - Medical devices
  - Etc.
- All CAST analyses so far have found more factors than NTSB and other accident reports

Evaluations (2)

- Jon Hickey, US Coast Guard applied to aviation training accidents
  - US Coast Guard currently uses HFACS (based on Swiss Cheese Model)
  - Spate of recent accidents but couldn't find any common factors
  - Using CAST, found common systemic factors not identified by HFACS
  - USCG now deciding whether to adopt CAST
- Dutch Safety Agency using it on a large variety of accidents (aircraft, railroads, traffic accidents, child abuse, medicine, airport runway incursions, etc.)

Organizational Aspects of Risk

- Examples so far focus on physical level
- Also requirements and control responsibilities at management level to satisfy system safety requirements
- Can identify unsafe control actions and causal scenarios at higher levels of the control structure (perform a risk analysis) and build in controls to prevent them
- Behavior and control structures change over time
  - Prevent migration to higher levels of risk
  - Detect when occurs

Organizational Aspects of Risk (2)

- Can look at non-safety risks, including project risks, budget risks, schedule risks and tradeoffs
- Goal may be to evaluate an existing control structure or to create a new one
- Creating leading indicators
- Current or past examples:
  - NASA safety management after Columbia
  - Radiation therapy at UCSD and UCLA hospitals (and maybe Boston Mass General)
  - CO_2 capture, transport, and storage (Samadi, Ecole des Mines)
  - Product Development Process (Goerges, Cummins Engine)

Other Topics Covered by STAMP

- Operations
- Managing safety-critical projects
- Integrating safety into system engineering
  - Designing safety into systems from the beginning
  - Specification to support maintenance and evolution

Current Projects

- Human factors engineering
  - Design to reduce human error
  - Integrating sophisticated human factors into hazard analysis
- Leading Indicators
- Cyber Warfare and other security applications
- Food safety
- More applications: high-speed rail, autos, medicine, NextGen (TBO)
- Financial system application
- Other emergent system properties
- Tools and formal assistance with analysis

Nancy Leveson, *Engineering a Safer World:*

MIT Press, January 2012