A systems-theoretic approach to analyze human-automation interactions

Dr. John Thomas
MIT

Human Factors in Control
April 2018
Halden, Norway

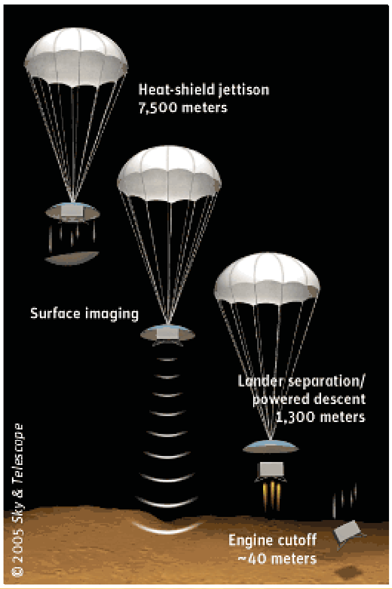
Outline

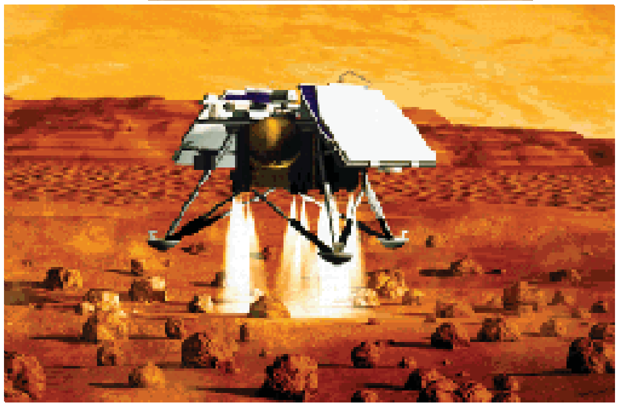
- Safety Engineering
  - Modern engineering challenges
  - Modern solutions
  - Application to human factors

Mars Polar Lander

- During the descent to Mars, the legs were deployed at an altitude of 40 meters.
- Touchdown sensors (on the legs) sent a momentary signal
- The software responded as it was designed to: by shutting down the descent engines.
- The vehicle free-fell and was destroyed upon hitting the surface at 50 mph (80 kph).

All components performed exactly as designed, all requirements met!





Bottom-up approach

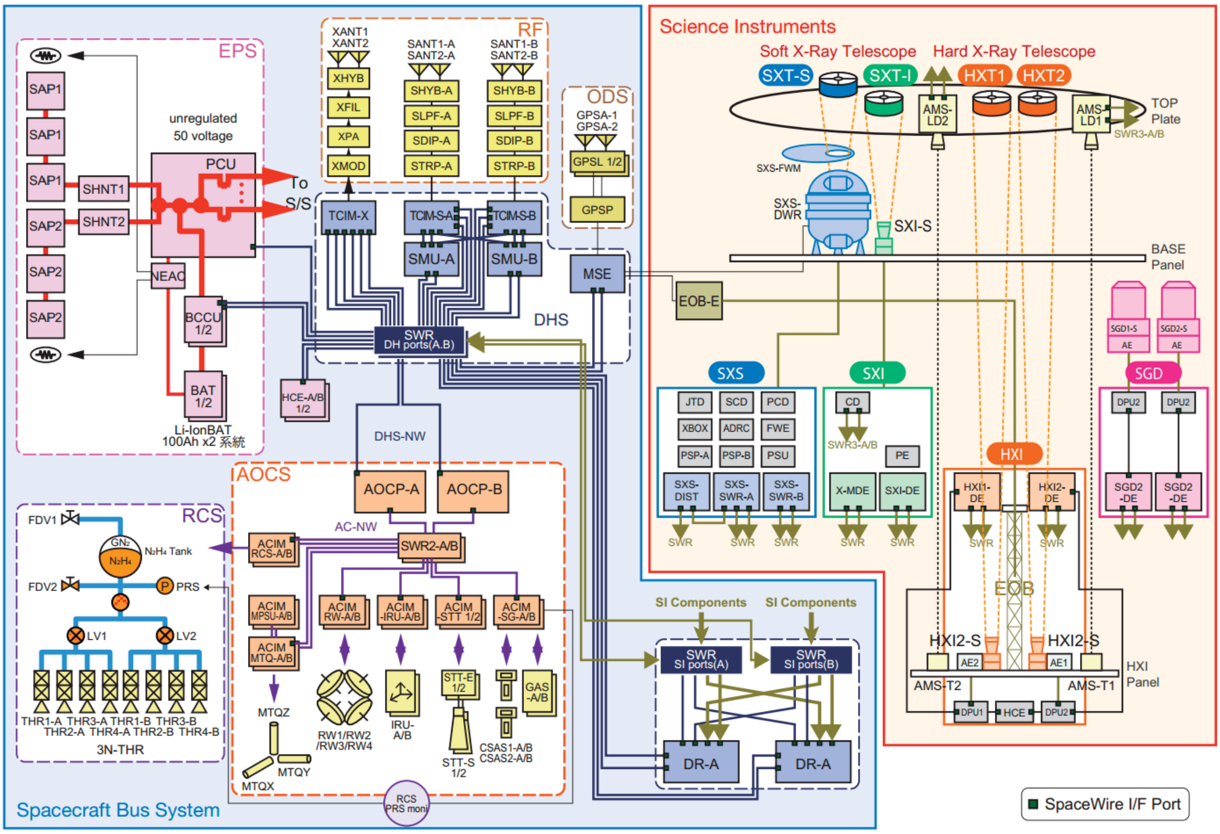


Figure 3.9: System block diagram. A is the primary and B is the redundant system.

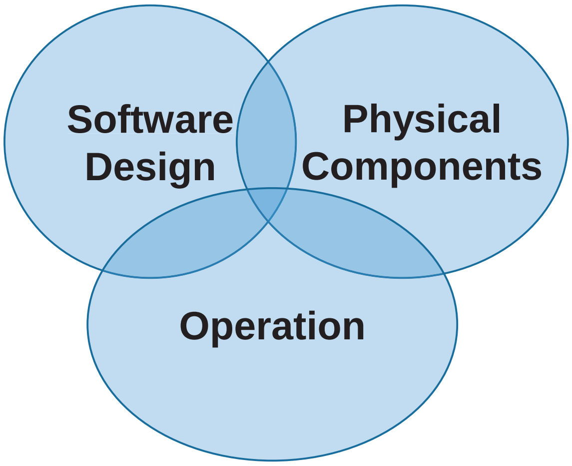
Tactics

What do we do before an accident?

- HW requirements: Sensor sensitivity
- SW requirements: React within X ms
- Processor loading
  - Initial plan: software runs after legs deployed
  - New plan: start software early to reduce processor load
- HW Testing: Verify HW sensitivity
- SW Testing: Verify SW reaction time
- Etc.

Systems View

What we missed

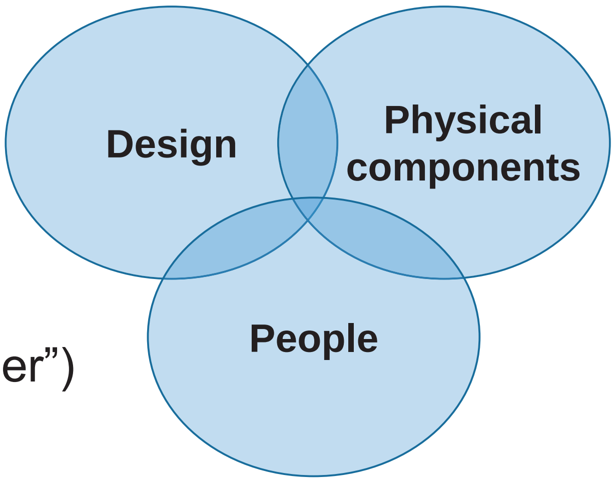


Hard to see problem by looking at any one part

Systems View

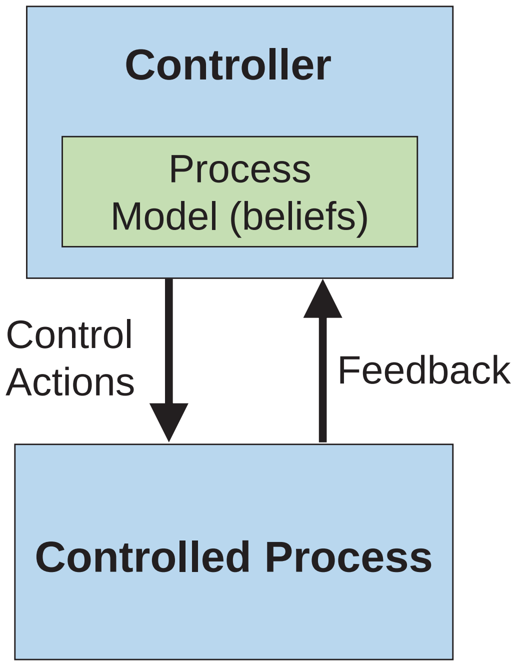
Many different factors were involved:

- Touchdown sensors
- Software implementation
- Software requirements
- Testing
- Engineering reviews
- Communication
- Time pressure
- Culture (“Faster, Better, Cheaper”)
- Etc.



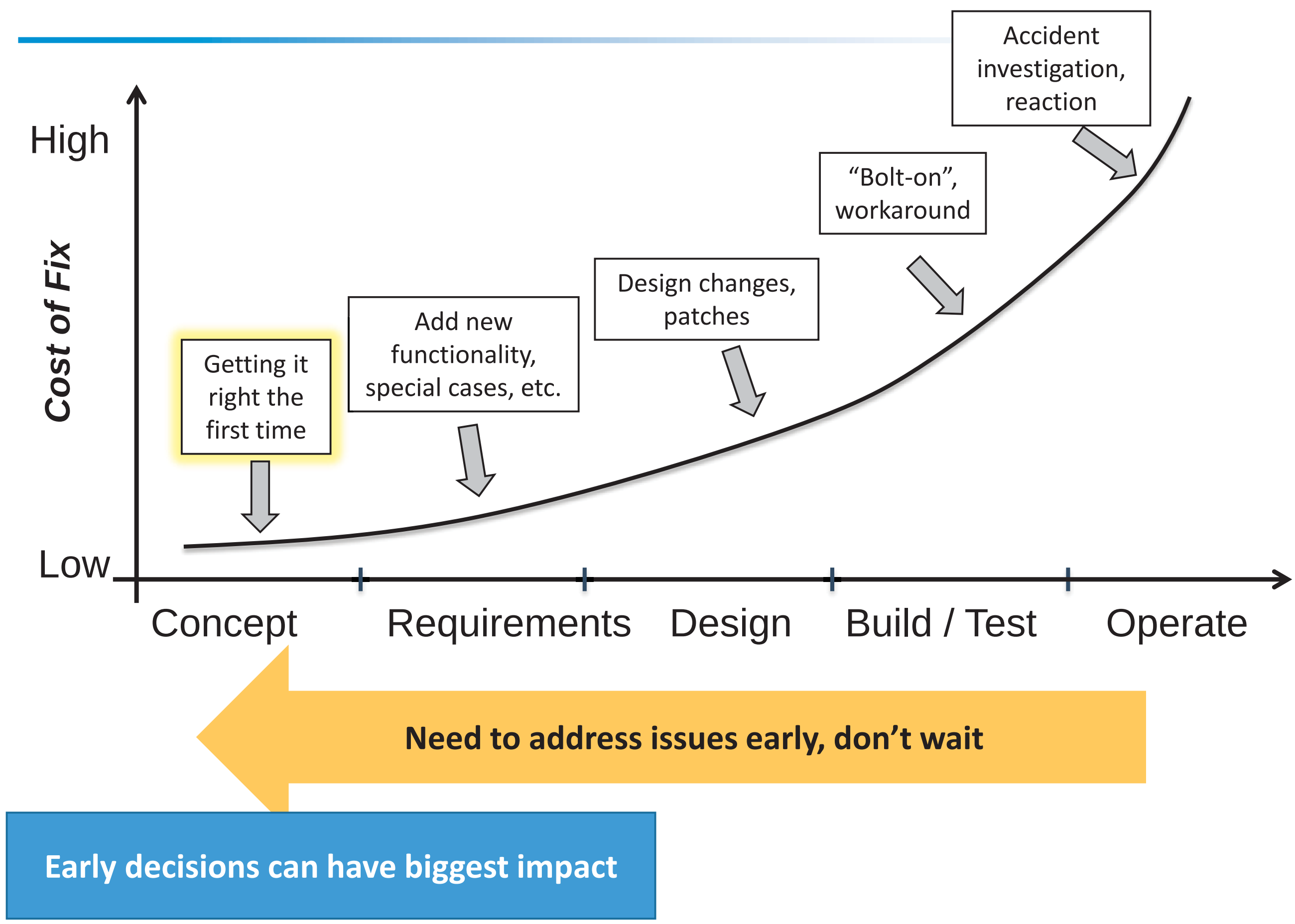
Hard to anticipate these problems by looking at any single component!

A different view



- Provides another way to think about accidents
- Emphasis on interactions
- Forms foundation for STAMP/STPA

Fixing problems

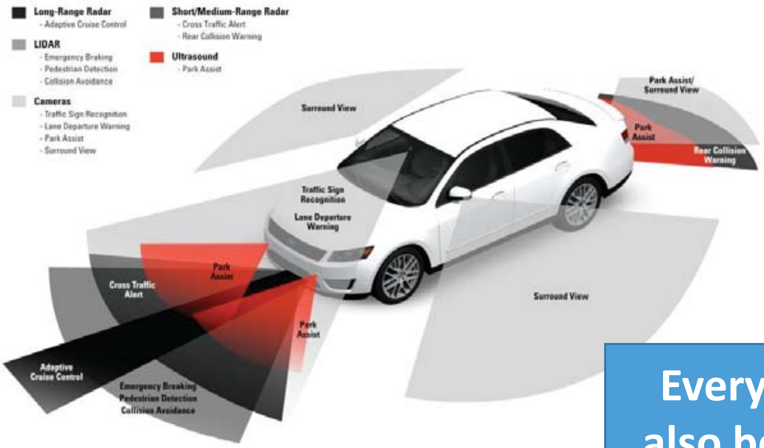


Adapted from Young, 2014

This presentation: automotive

Challenging problem:

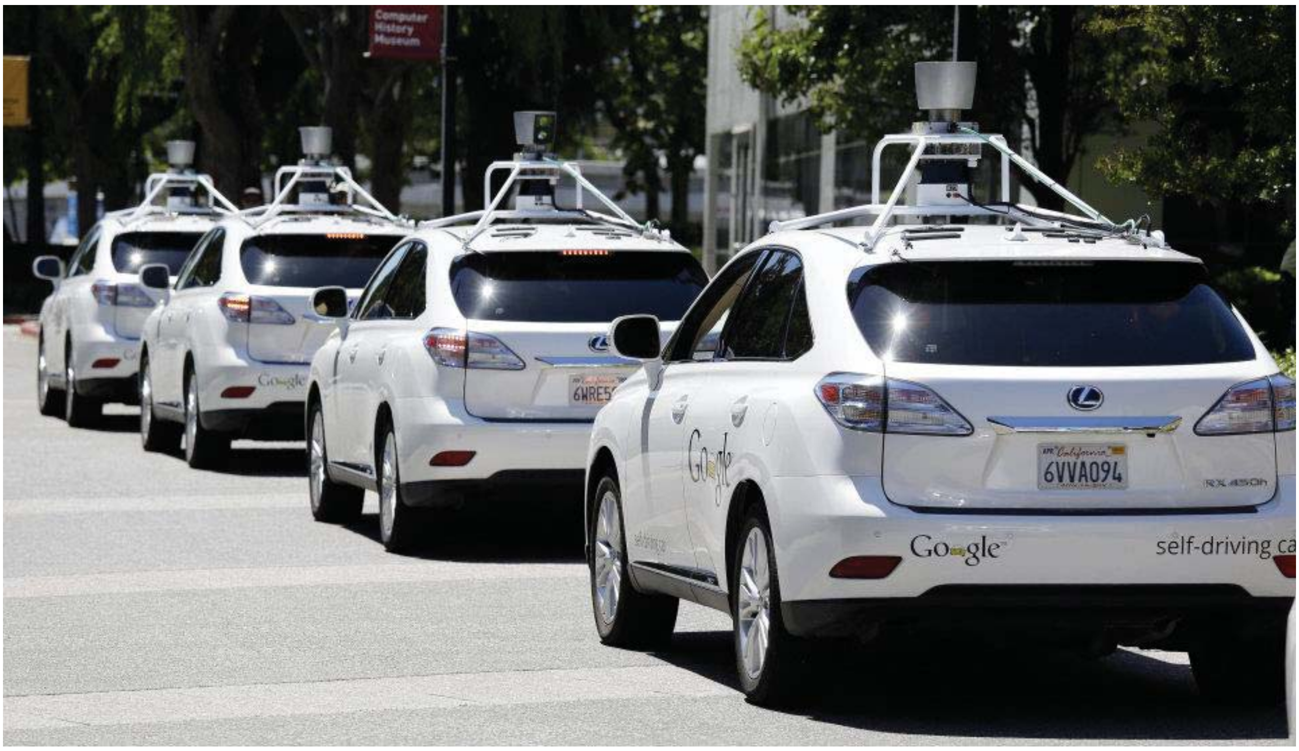
- Complex automation
- No training



Everything in this presentation also being used in aviation, oil & gas, nuclear, chemical, etc.

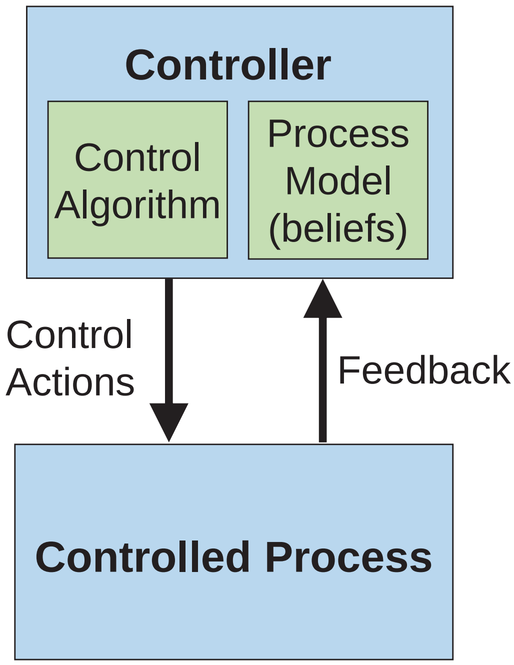
Chart: https://hbr.org/2010/06/why-dinosaurs-will-keep-ruling-the-auto-industry/ar/1

Google Self-Driving Car



Google Self-Driving Car

A different view



Discuss application to AI

Unintended Acceleration

- **2004-2009:** 102 incidents

Operated exactly as designed!
No component failure, no reverse flow, etc.
System behavior unexpected, unsafe

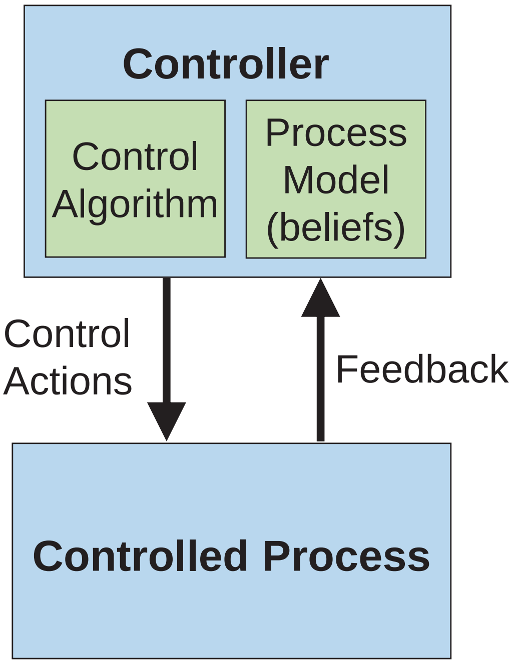
Unintended Acceleration

- **2004-2009:** 102 incidents

Human and technical
considerations cannot be isolated!

© Copyright John Thomas 2018

Another view



Applicable to Computers

Applicable to Humans

Monostable shifter design

NHTSA: "operation of the Monostable shifter is not intuitive and provides poor tactile and visual feedback to the driver, increasing the potential for unintended gear selection."

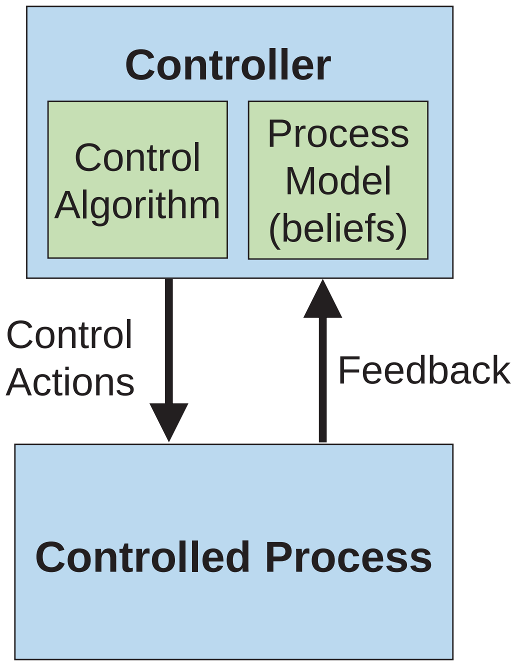
Monostable shifter design

Designed by German supplier
OEM still responsible for integration

Monostable shifter design

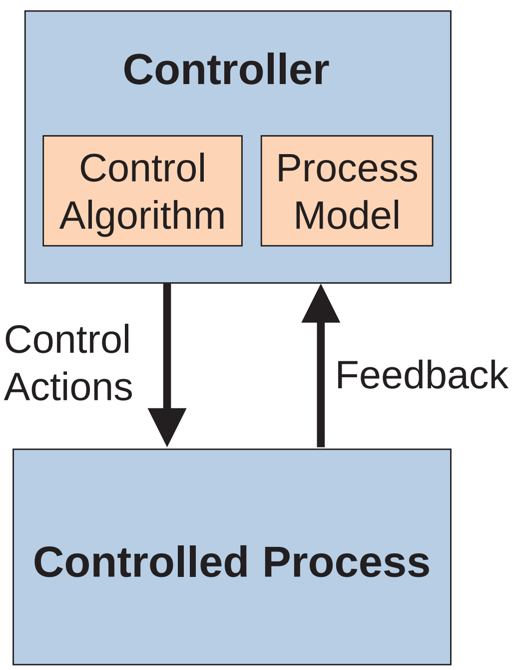
Audi A8: Similar design, but SW will automatically activate electronic park brake if driver exits

Another view

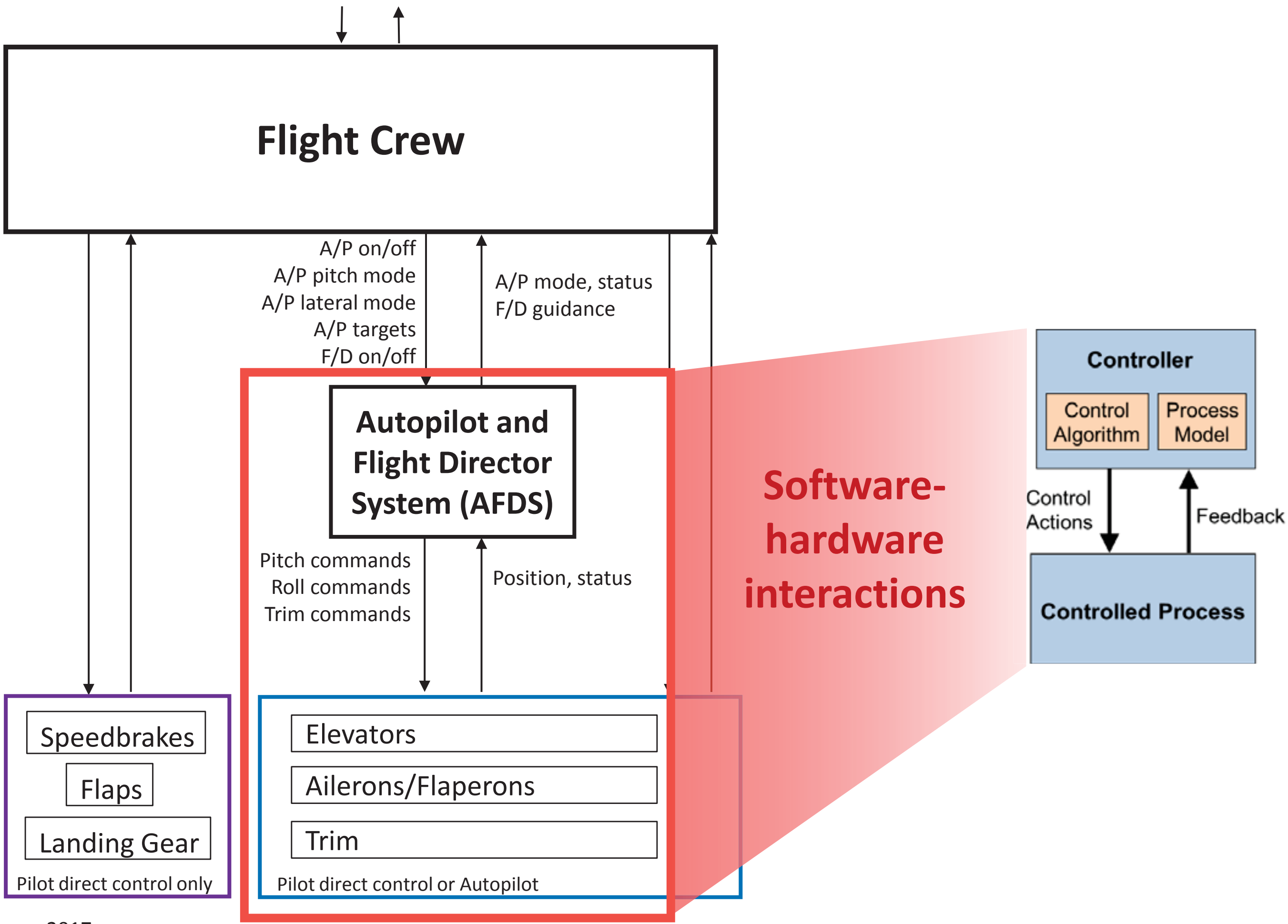


- Can be used in engineering to anticipate and prevent these problems earlier, before simulators or detailed models are available

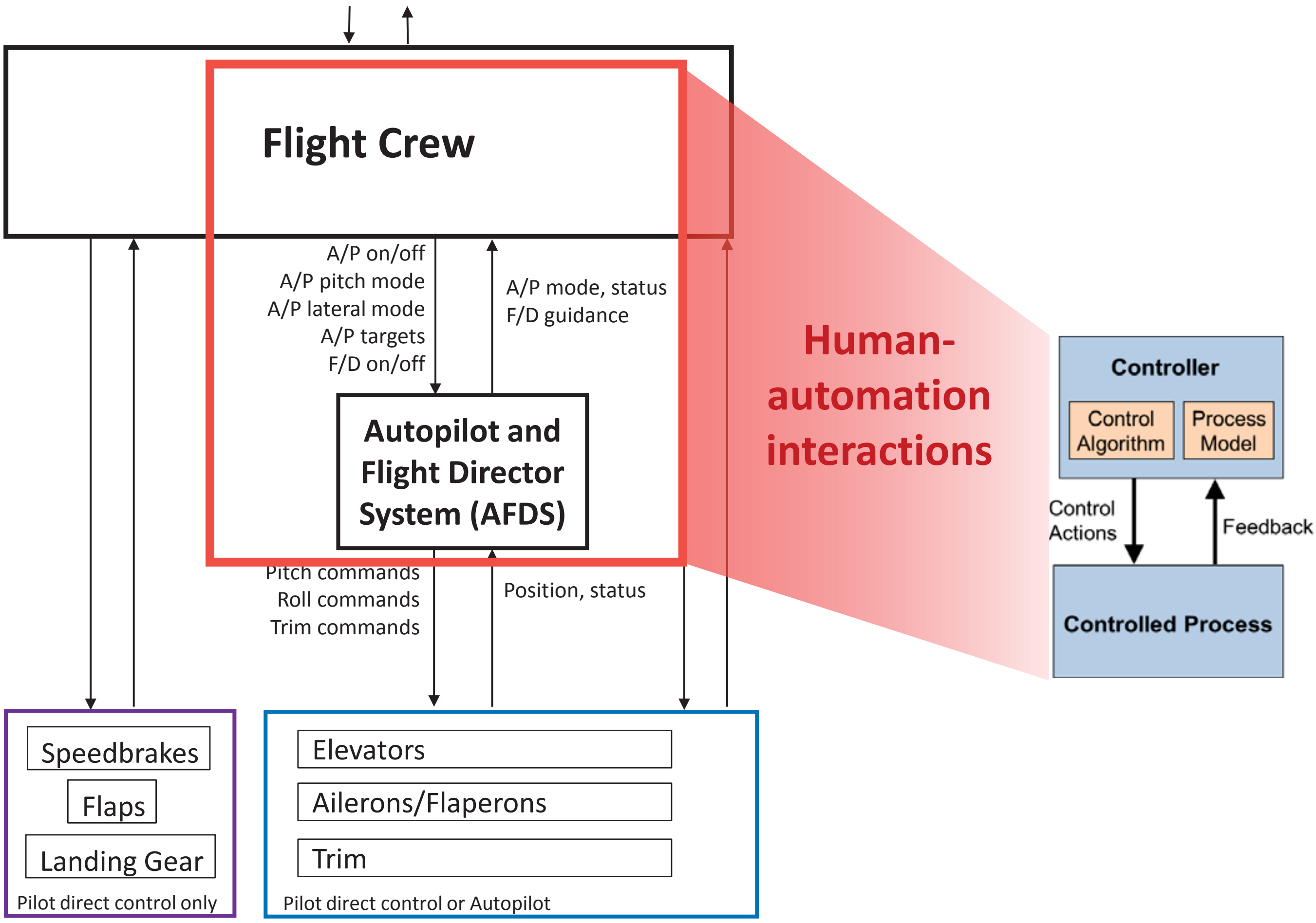
Another view



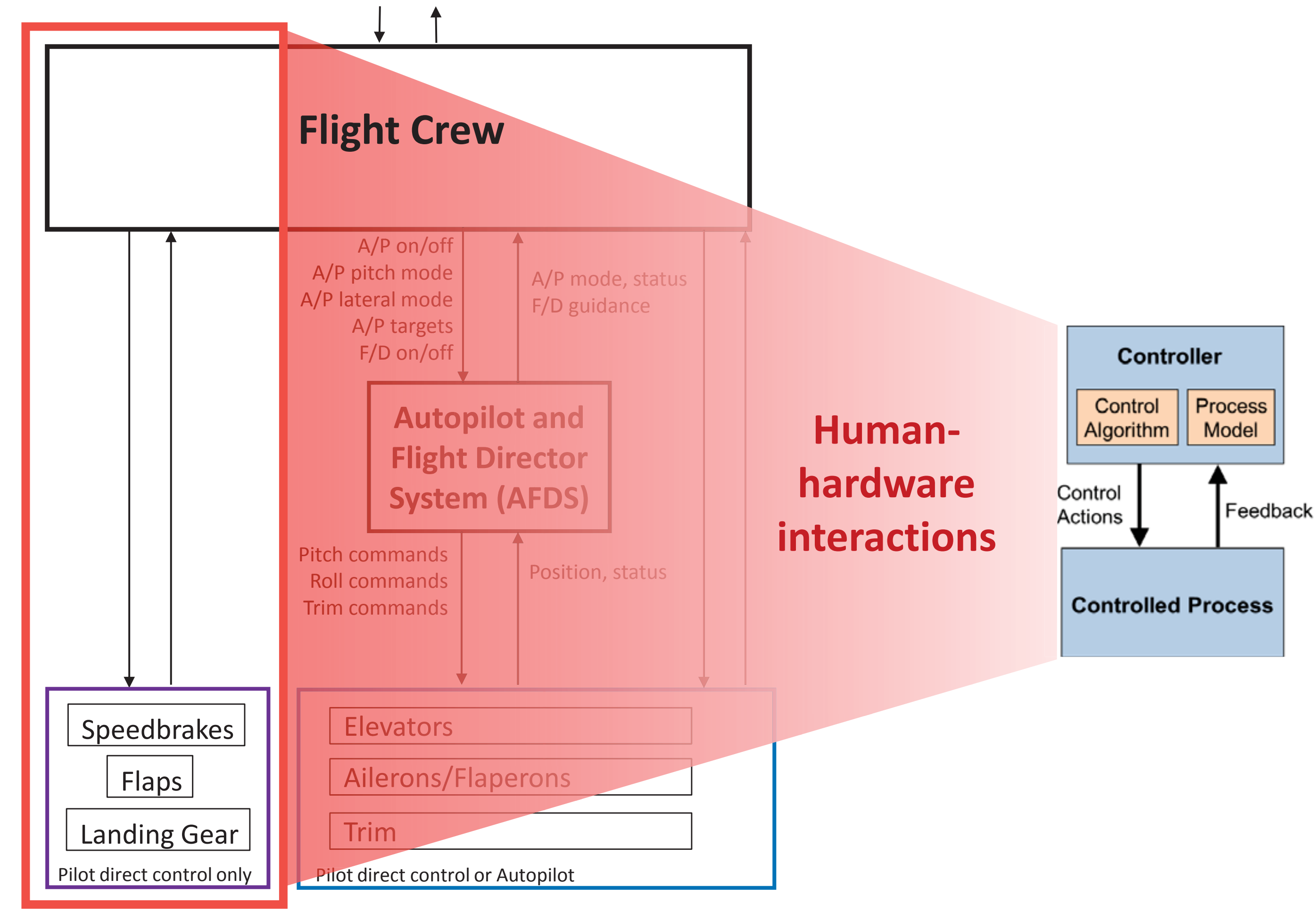
- **Control actions** are provided to affect a controlled process
- **Feedback** may be used to monitor the process
- **Process model** (beliefs) formed based on feedback and other information
- **Control algorithm** determines appropriate control actions given current beliefs



Thomas, 2017

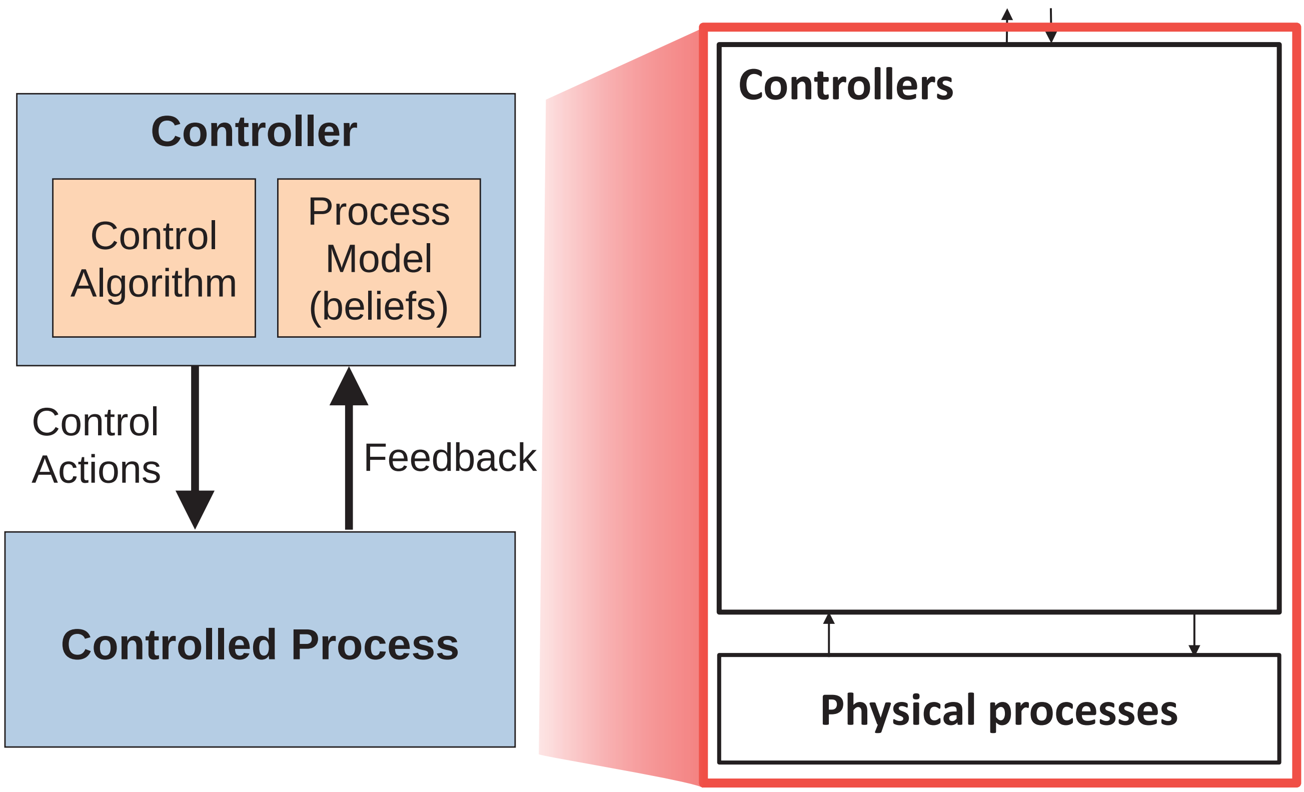


Thomas, 2017

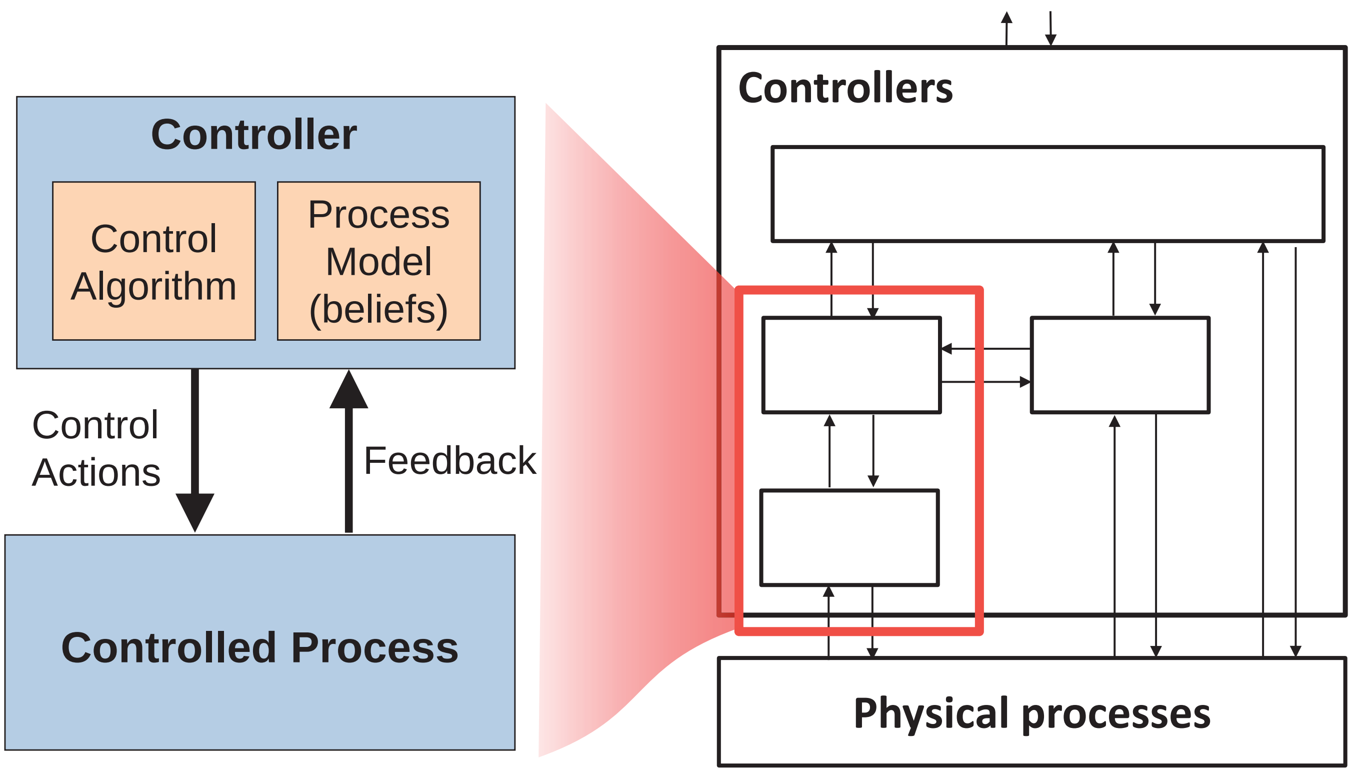


Thomas, 2017

Abstraction

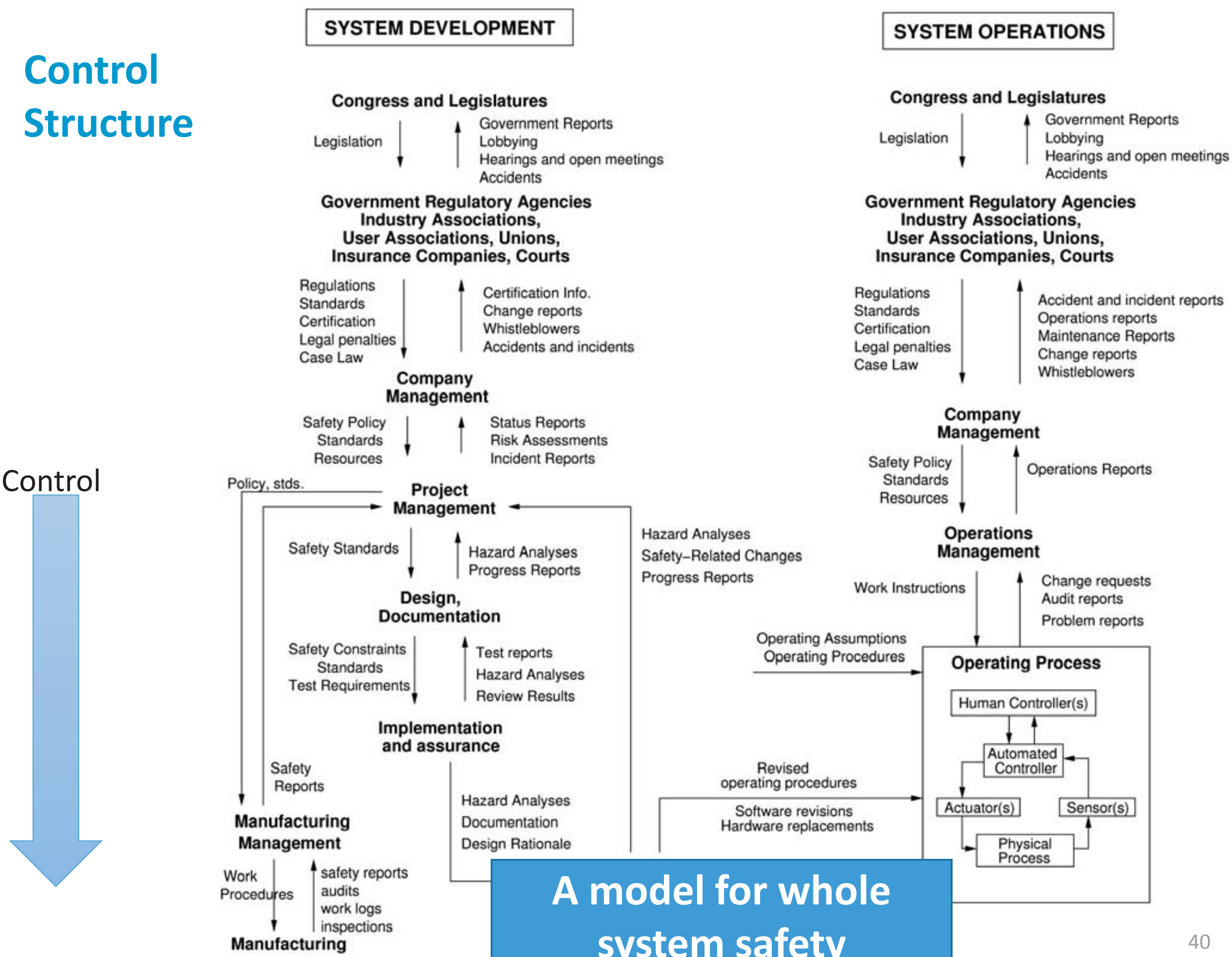


Refinement



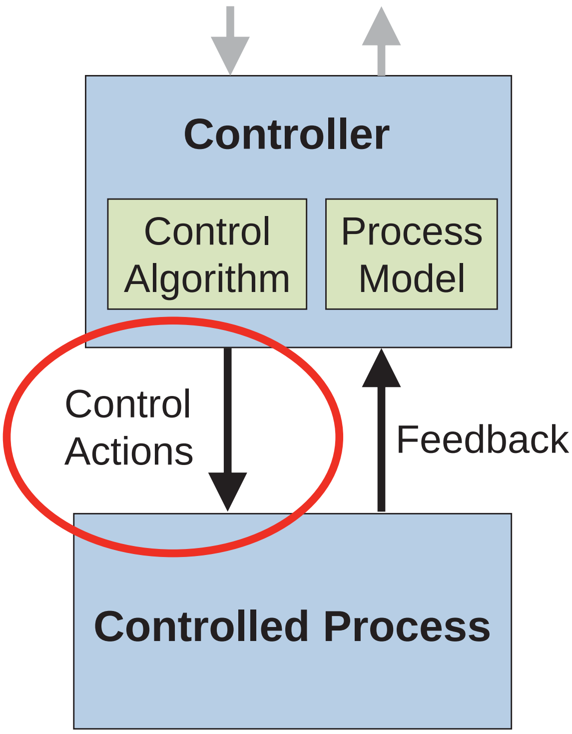
Control Structure

Control



A model for whole system safety

(Leveson, 2012)



Four types of **unsafe control actions**:

1) Control actions required for safety are not given
2) Unsafe ones are given
3) Potentially safe control actions but given too early, too late
4) Control action stops too soon or applied too long

(Leveson, 2012)

Application to Engineering

STPA

Systems Theoretic Process Analysis

Basic STPA

1. Identify accidents, hazards — Losses to prevent
2. Draw control structure — Model
3. Identify unsafe control actions — Behavior to prevent
4. Identify accident scenarios — How could behavior occur

(Leveson, 2012)

System-Theoretic Process Analysis (STPA)

- Identify Accidents, hazards
- Draw functional control structure
- Identify unsafe control actions
- Identify accident scenarios

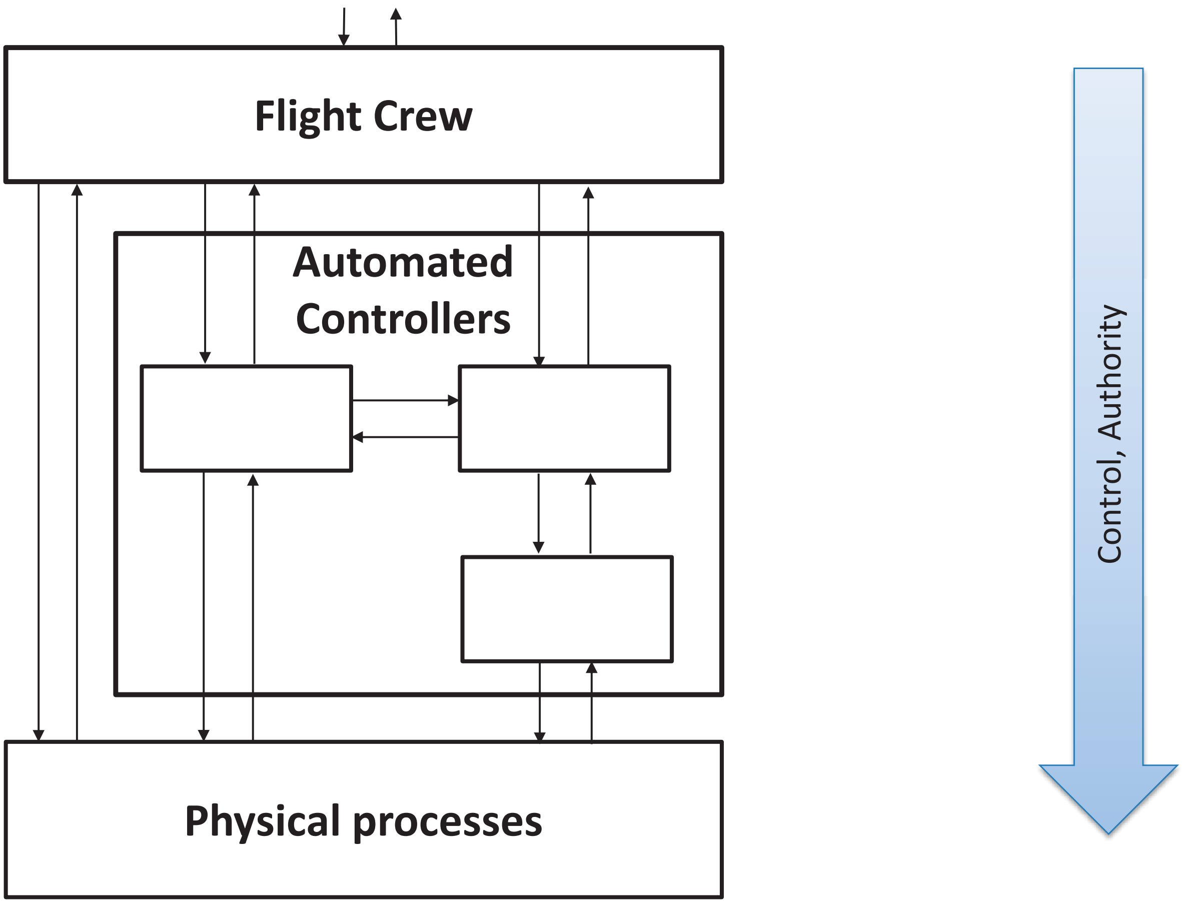
(Leveson, 2012)

System-Theoretic Process Analysis (STPA)

- Identify Accidents, hazards
- Draw functional control structure
- Identify unsafe control actions
- Identify accident scenarios

(Leveson, 2012)

Basic STPA: (2) Control Structure



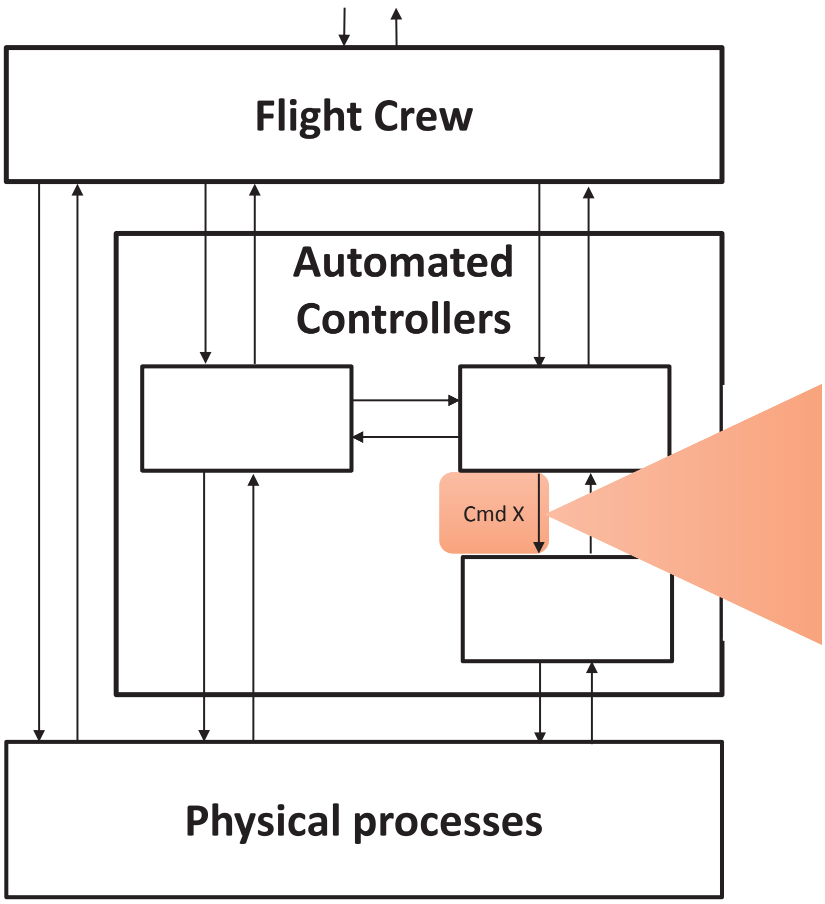
Thomas, 2017

System-Theoretic Process Analysis (STPA)

- Identify Accidents, hazards
- Draw functional control structure
- Identify unsafe control actions
- Identify accident scenarios

(Leveson, 2012)

Basic STPA: (3) Unsafe Control Actions (UCA)



Not provided causes hazard	Providing causes hazard	Too early, too late, out of order	Stopped too soon, applied too long

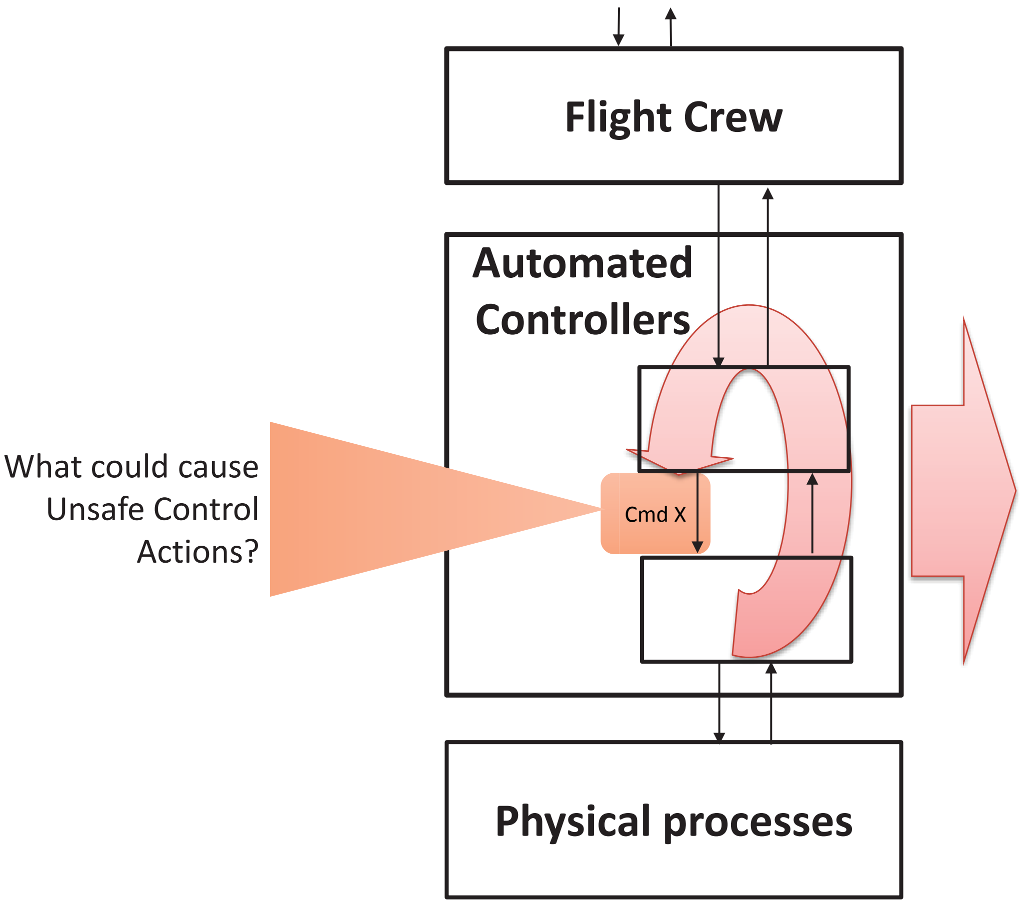
Thomas, 2017

System-Theoretic Process Analysis (STPA)

- Identify accidents, hazards
- Draw functional control structure
- Identify unsafe control actions
- **Identify accident scenarios**

(Leveson, 2012)

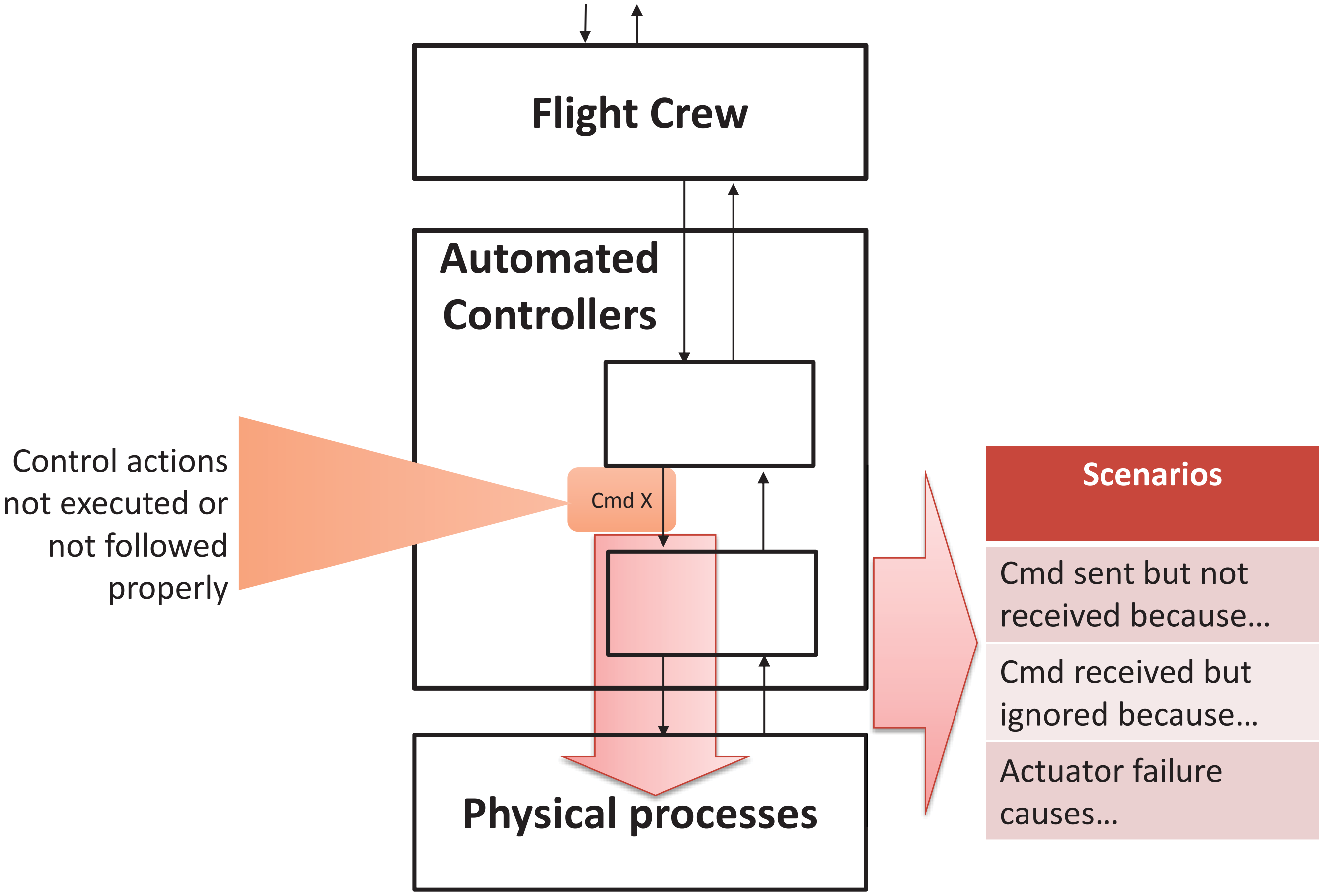
Basic STPA: (4) Identify Accident Scenarios



Scenarios
Controller incorrectly believes X because …
Controller control algorithm does not enforce Y because …
Incorrect feedback Z received because …
Sensor failure causes…
Etc.

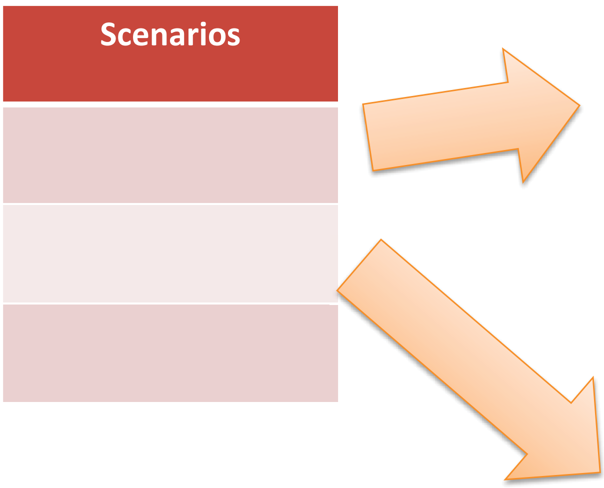
Thomas, 2017

Identify Accident Scenarios



(Thomas, 2017)

Design recommendations and component requirements



Design recommendations
Component A must be able to respond within B seconds <u>to avoid C</u>
Controller X must take into consideration D <u>to prevent E</u>
Etc.

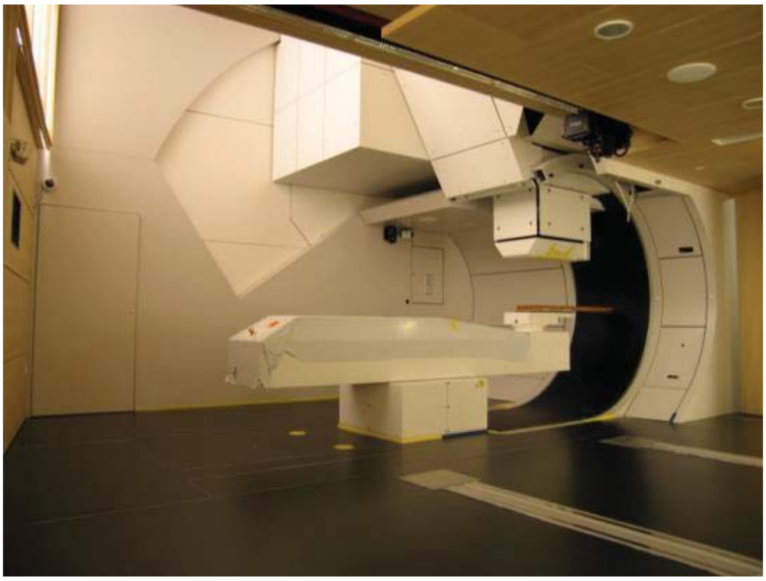
Rationale and assumptions identified

Component requirements
Component F shall automatically operate within G seconds <u>when H</u>
Component I and J shall be operated at the same time <u>to prevent K</u>
Etc.

Every recommendation and requirement is traceable

(Thomas, 2017)

PSI Proton Therapy Machine High-level Control Structure



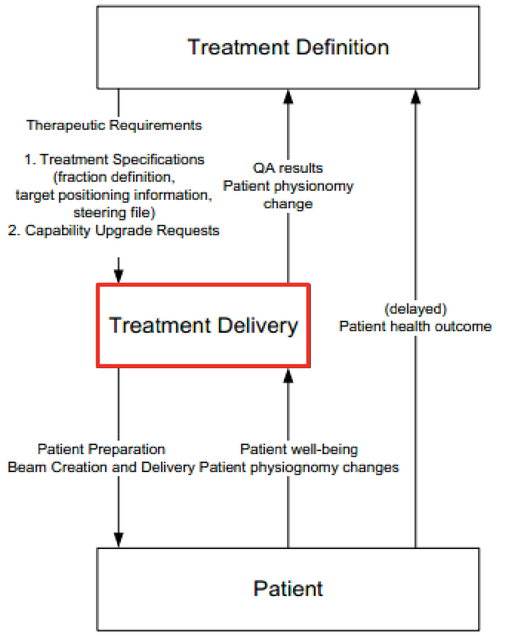
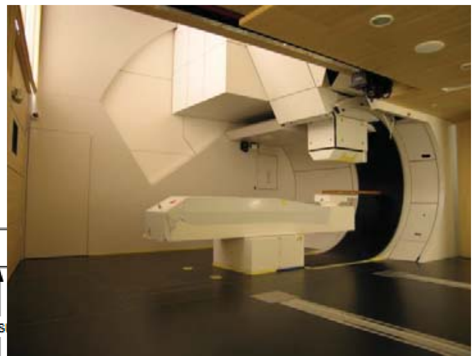
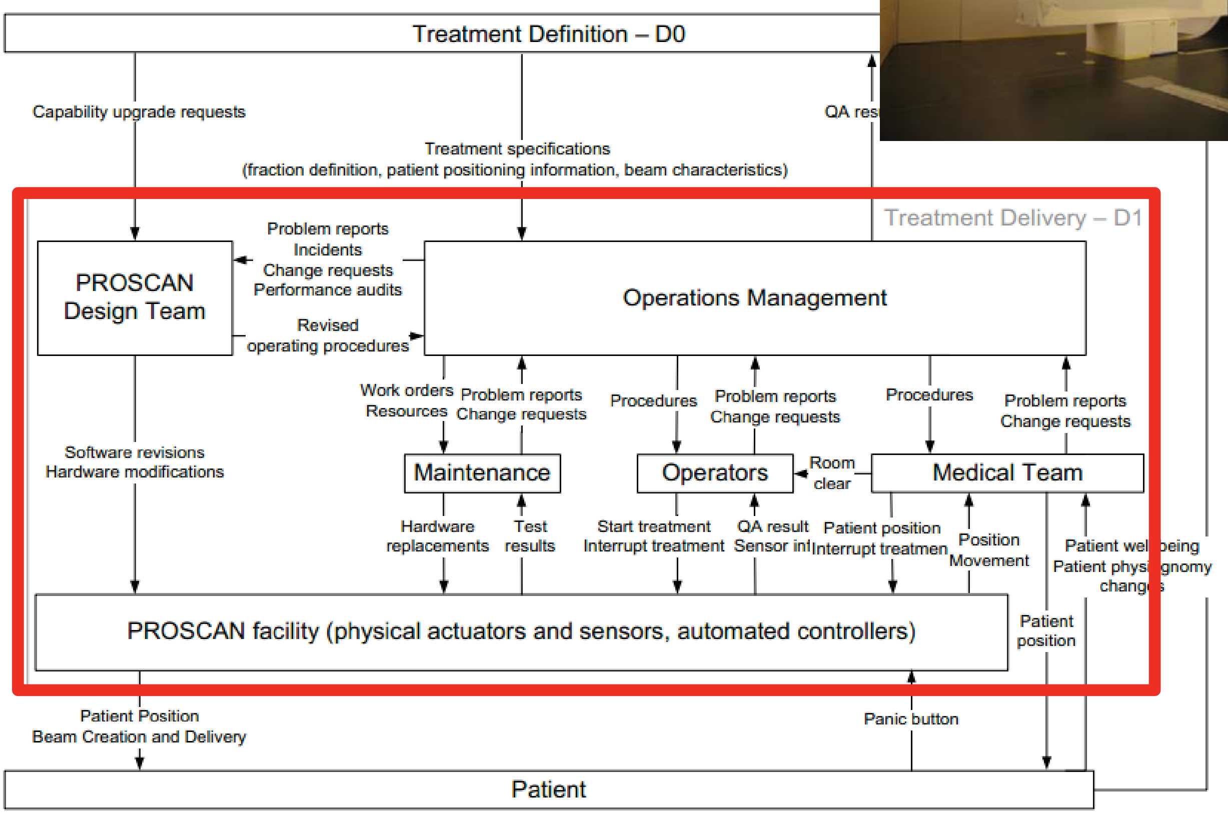


Figure 11 - High-level functional description of the PROSCAN facility (D0)

Antoine PhD Thesis, 2012

PSI Proton Therapy Machine Control Structure

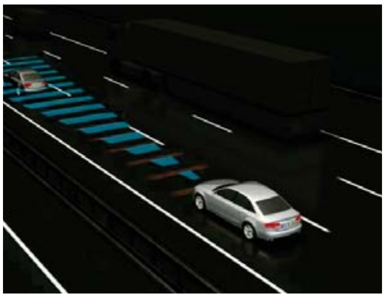


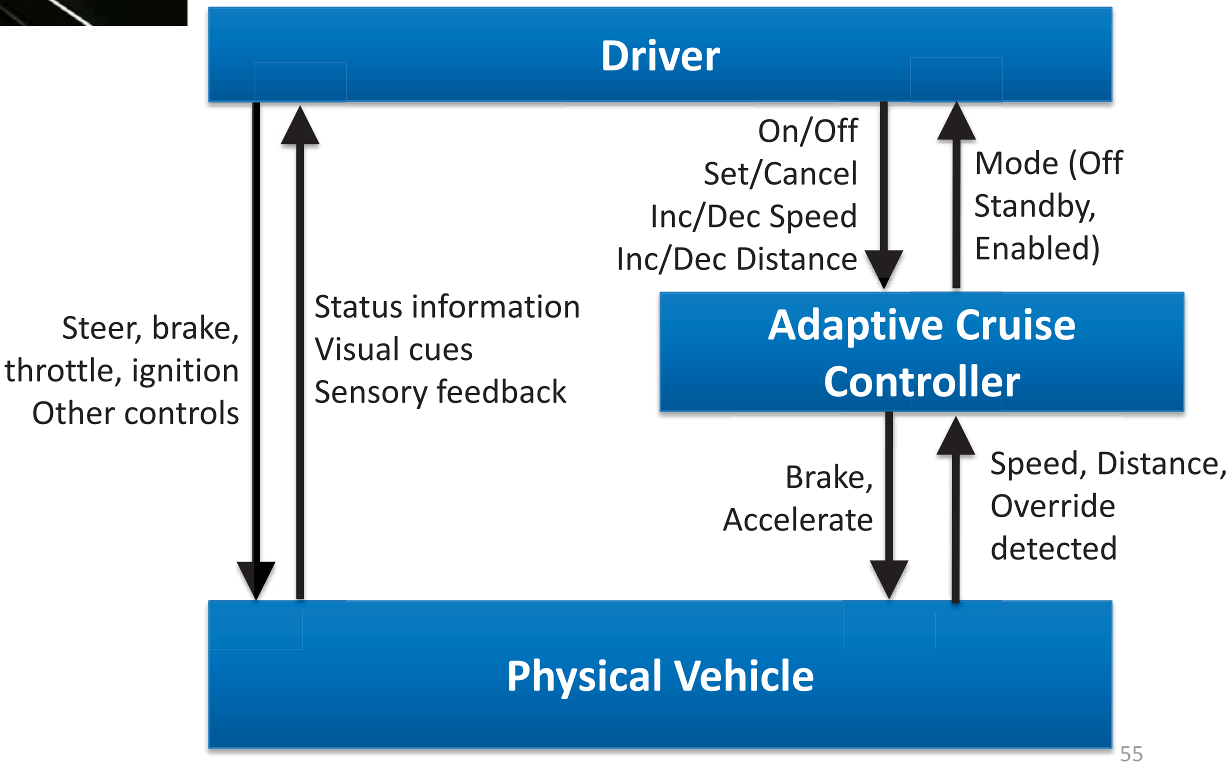


Antoine PhD Thesis, 2012

Figure 13 - Zooming into the Treatment Delivery group (D1)

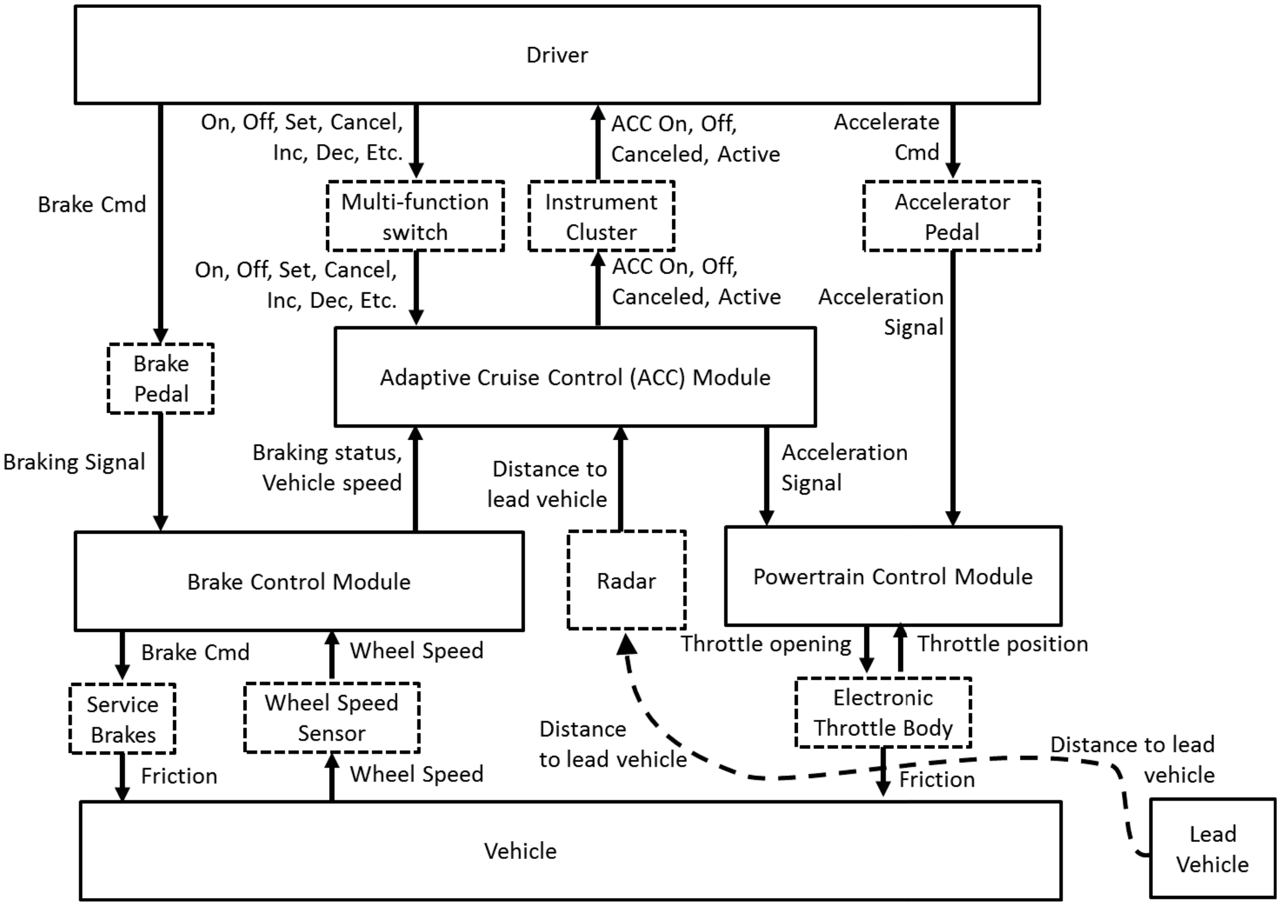
Adaptive Cruise Control





Thomas, 2012

Refined Control Structure

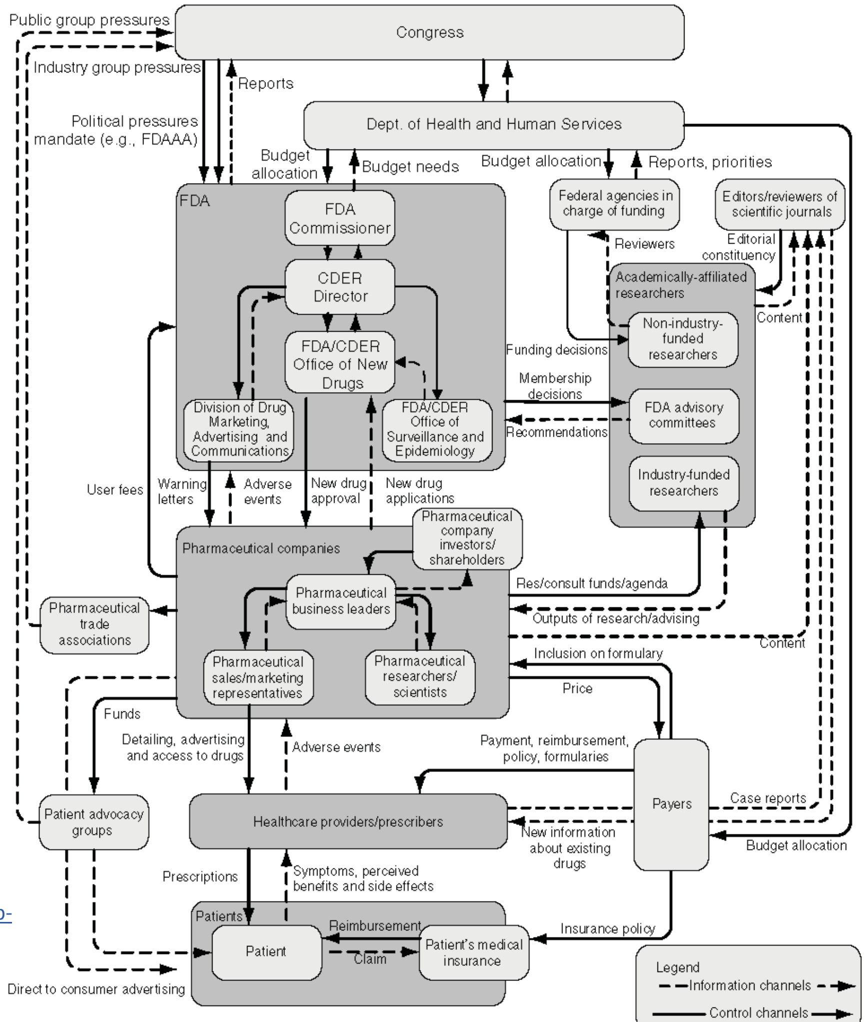


U.S. pharmaceutical safety control structure

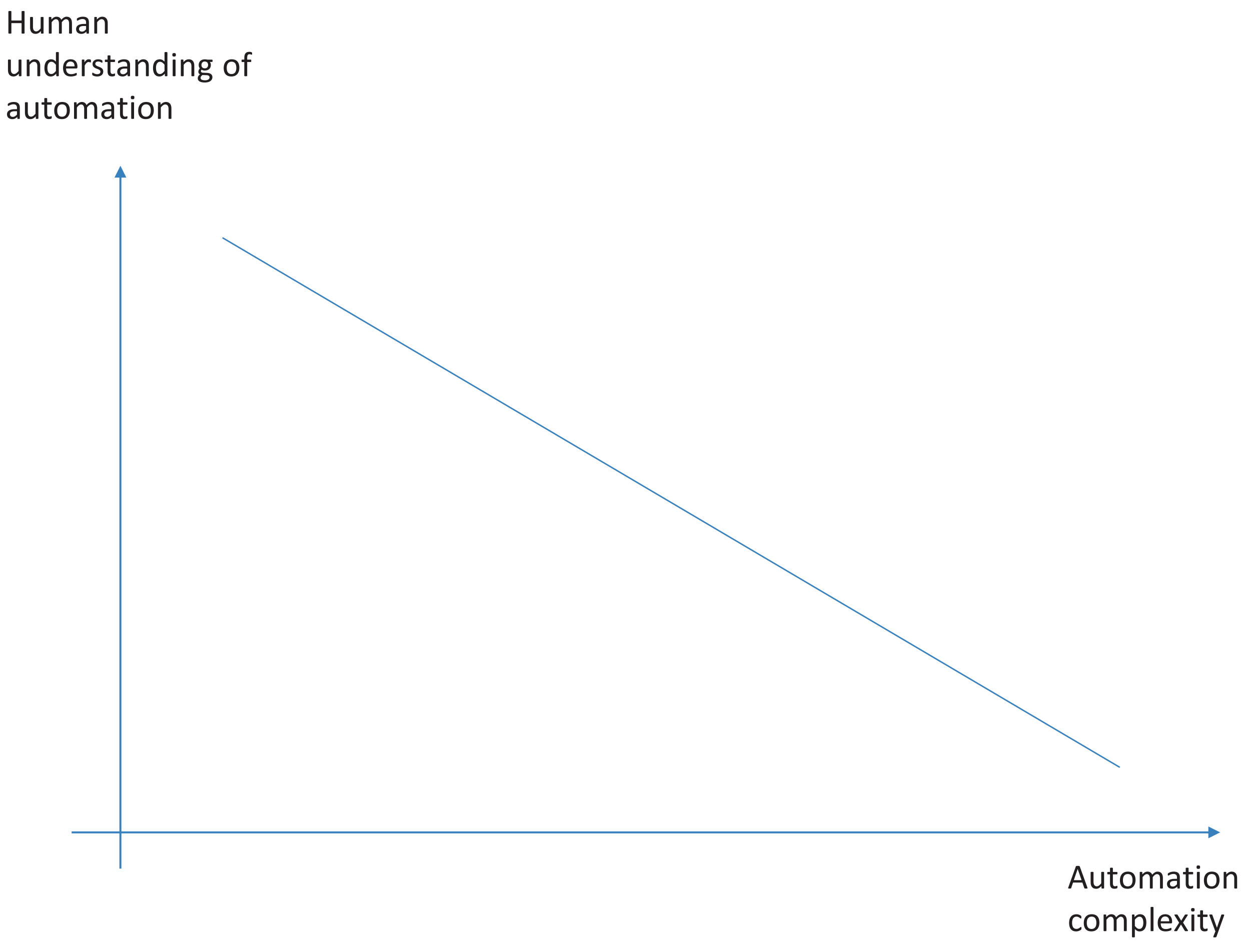
(a purely human/organizational system)

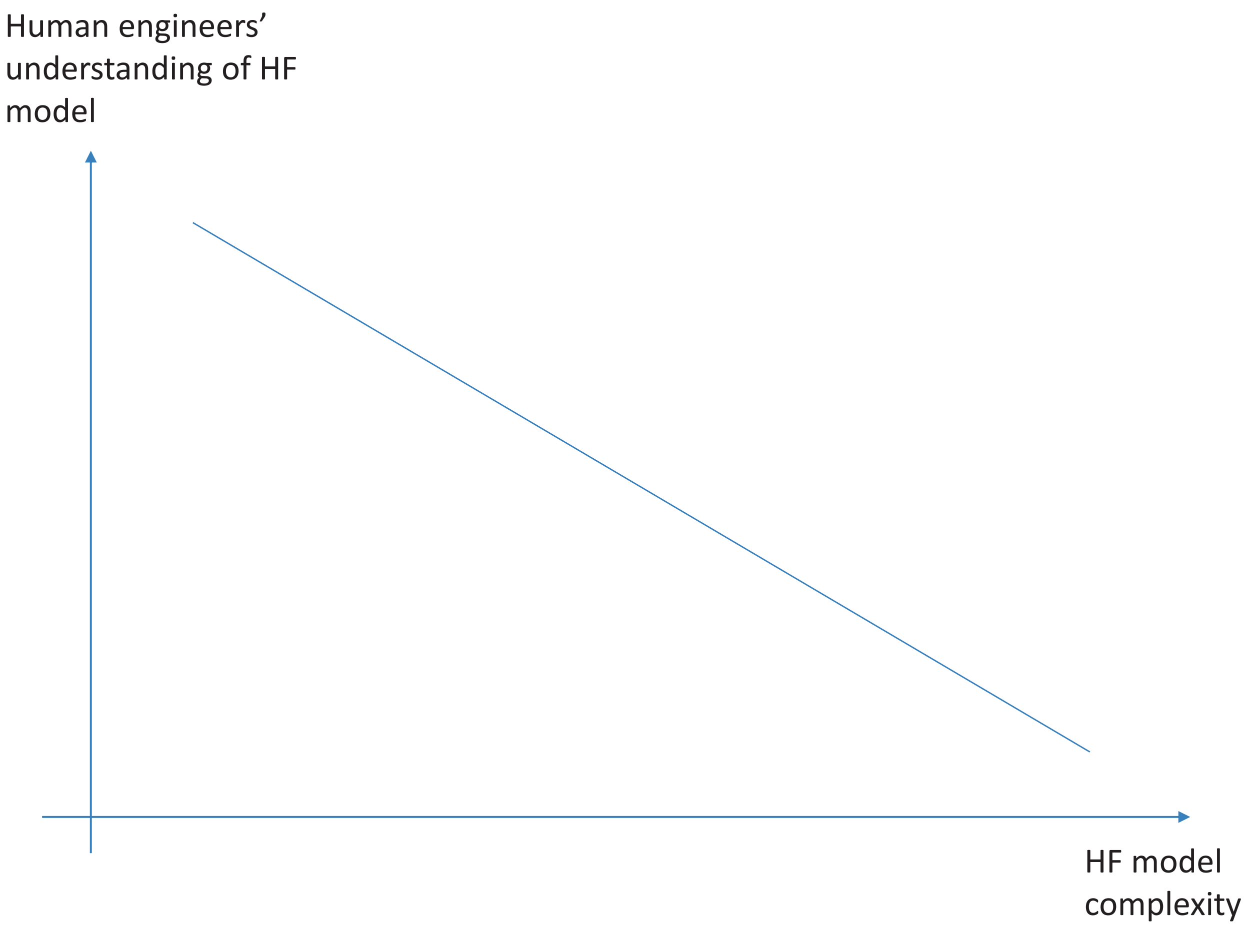
Image from: http://www.kleantreatmentcenter.com/wp-content/uploads/2012/07/vioxx.jpeg

Leveson, Couturier, Thomas, Dierks, Wierz, Psaty, Finkelstein, Applying System Engineering to Pharmaceutical Safety

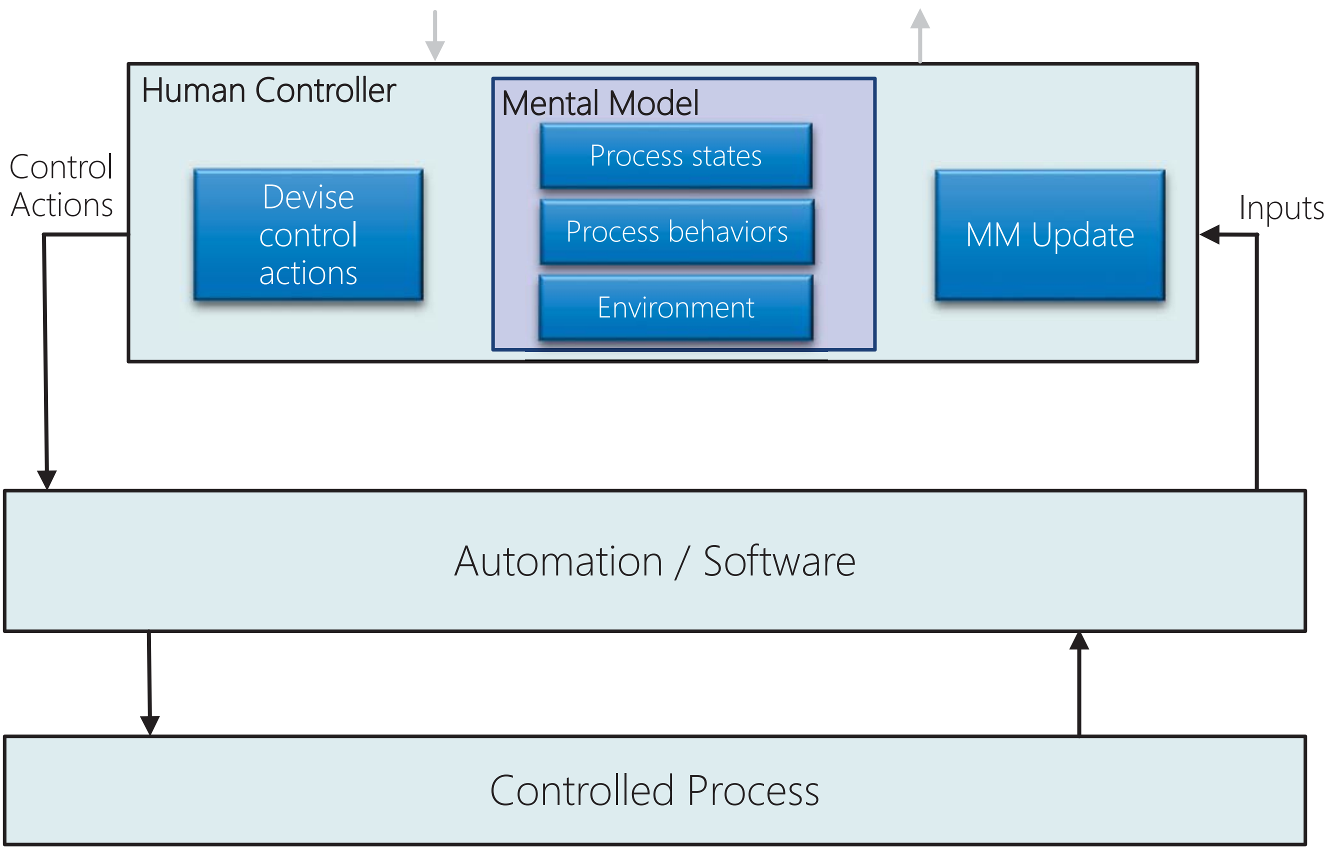


Application to human factors


Human understanding of automation
Automation complexity


Human engineers' understanding of HF model
HF model complexity

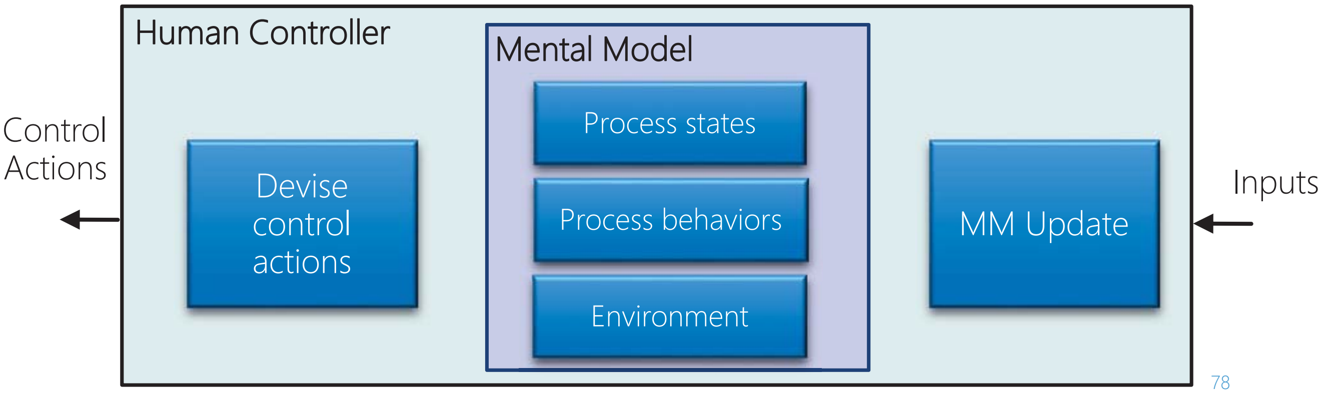
HUMAN CONTROL MODEL



ENGINEERING/ANALYSIS METHOD

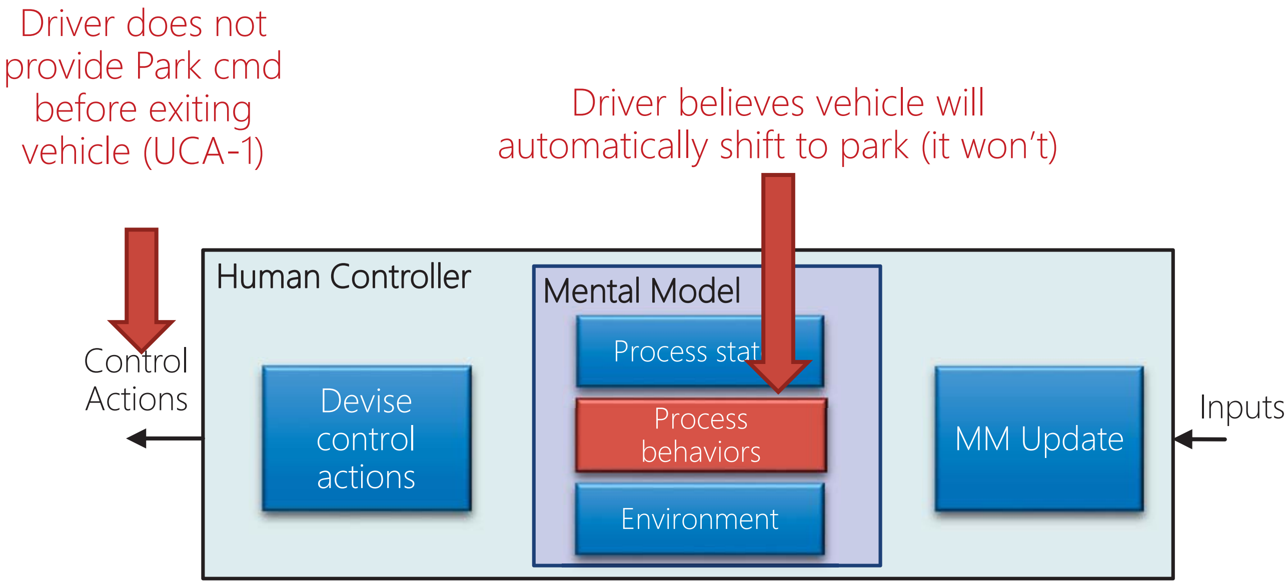
- Accidents (Losses), Hazards
- Control structure
- UCAs
- Build scenarios
  - Identify Mental Model variables
  - Identify Mental Model Flaws
  - Identify flaws in Mental Model Updates
  - Identify unsafe decisions (Control Action Selections)

Model is based on accidents



ACCIDENTS/INCIDENTS

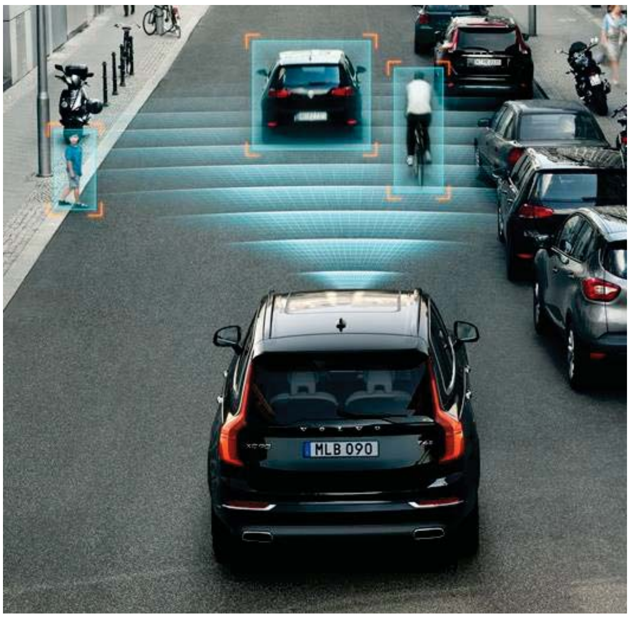
MENTAL MODEL OF BEHAVIOR, CAPABILITY



VOLVO CITY SAFETY SYSTEM

From Volvo website:

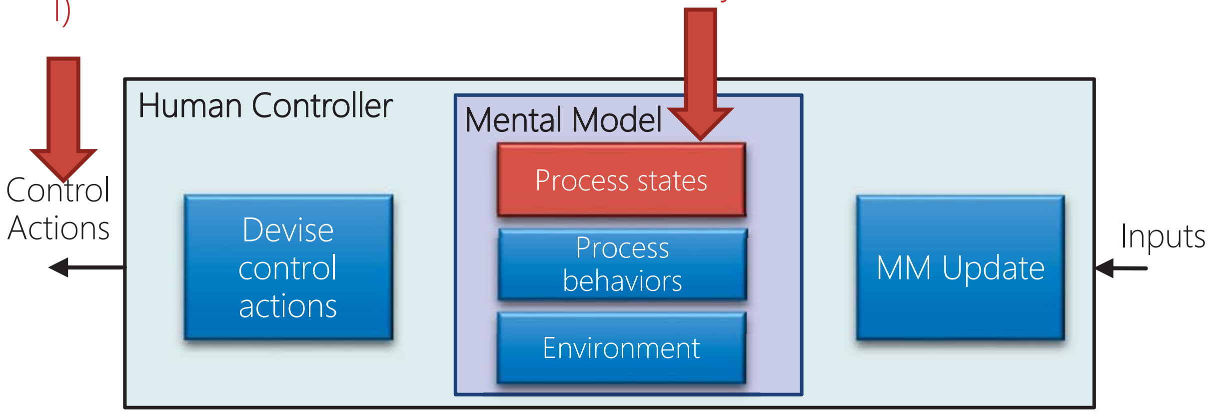
- City Safety is a support system designed to help the driver avoid low speed collisions when driving in slow-moving, stop-and-go traffic.
- City Safety triggers brief, forceful braking if a low-speed collision is imminent.



MENTAL MODEL OF STATE

Driver does not brake for pedestrian (UCA-1)

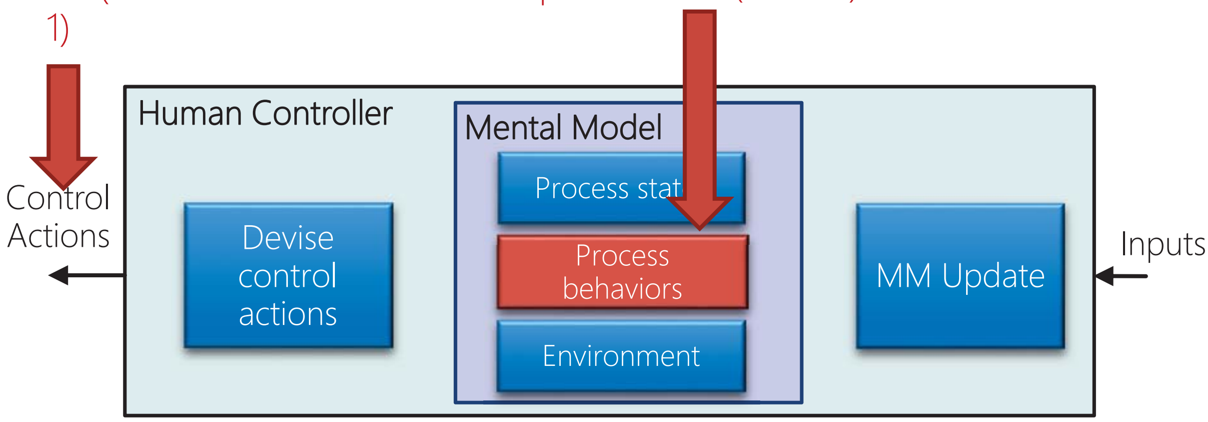
Driver thinks City Safety System is on (it is really off)



MENTAL MODEL OF BEHAVIOR, CAPABILITY

Driver does not brake for pedestrian (UCA-1)

Driver thinks City Safety System can automatically brake for pedestrians (it can't)



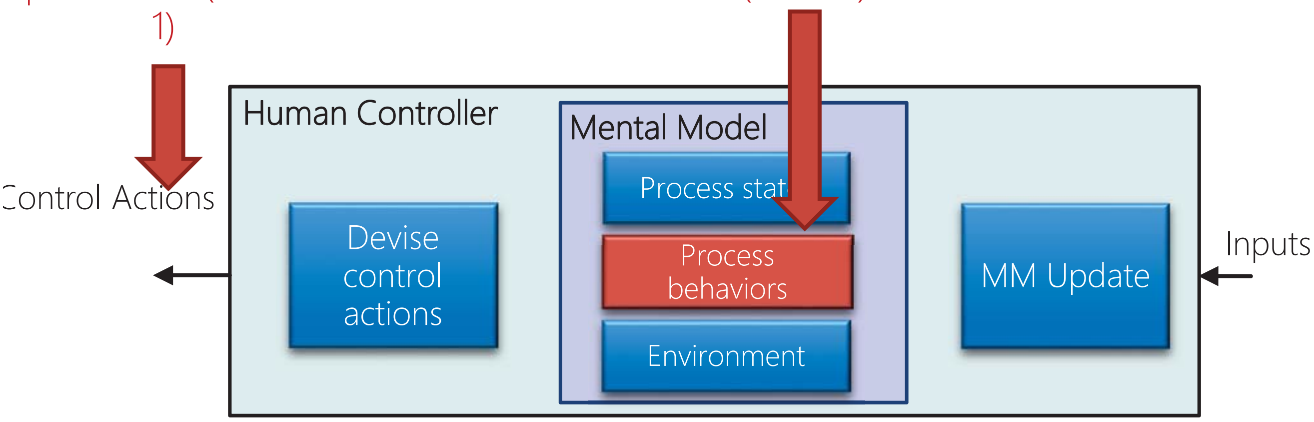
VOLVO RESPONSE

- "The Volvo XC60 comes with City Safety as a standard feature ...
- "however this does not include the Pedestrian detection functionality ... this is sold as a separate package."
- Optional pedestrian detection functionality costs $3,000
- Even with pedestrian detection, it mostly likely would not have worked because the driver accelerated

MENTAL MODEL OF BEHAVIOR, CAPABILITY

Driver does not brake for pedestrian (UCA-1)

Driver thinks City Safety System can intervene during acceleration (it can't)



Application to Engineering

Automated Parking Assist

Massachusetts Institute of Technology

John Thomas

Megan France

Collaboration with General Motors

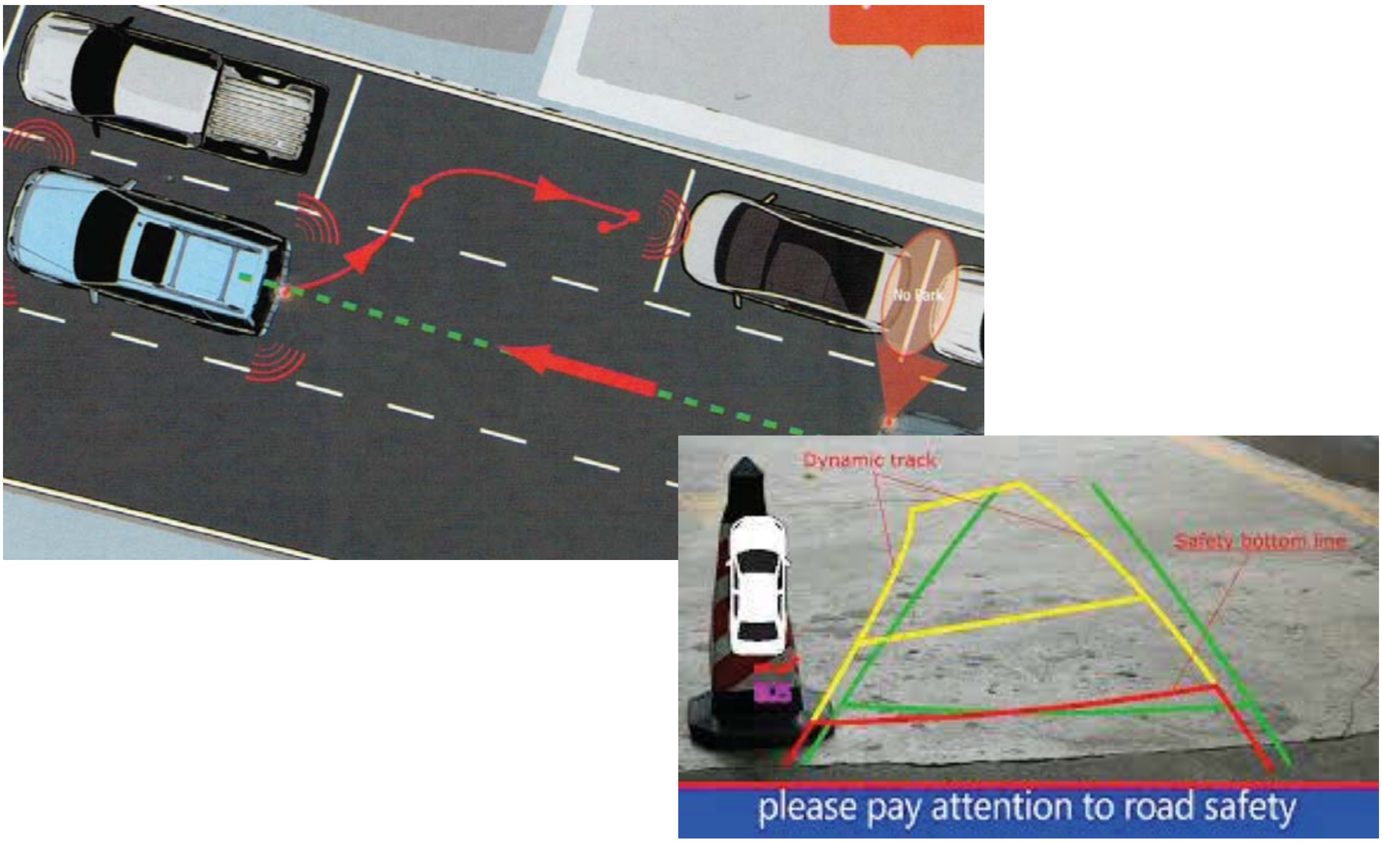
Charles A. Green

Mark A. Vernacchia

Padma Sundaram

Joseph D'Ambrosio

AUTOMATED PARKING ASSIST

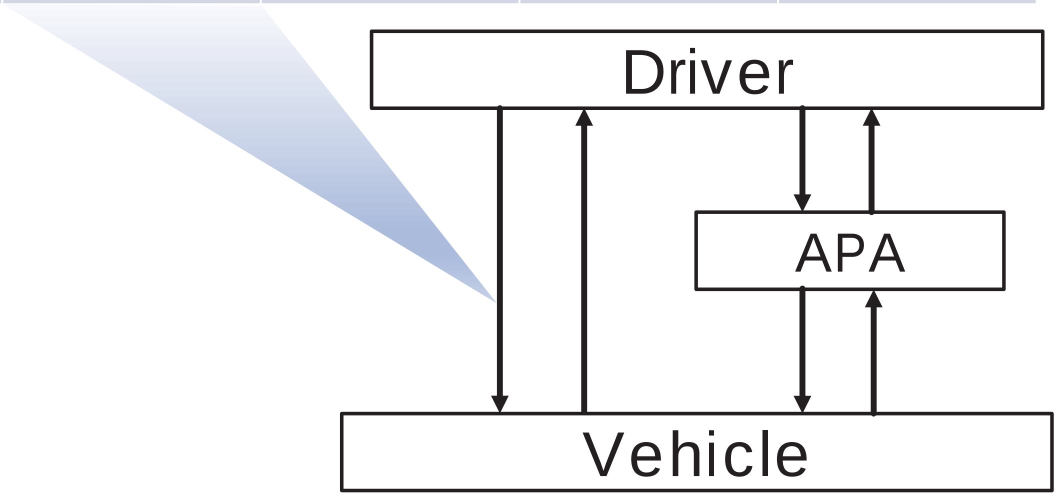


NEW PROCESS

- Identify UCAs
- Identify Mental Model variables
- Identify Mental Model Flaws
- Identify flaws in Mental Model Updates
- Identify unsafe decisions (Control Action Selections)

UNSAFE CONTROL ACTIONS

	Not Provided	Provided	Too early, too late, out of order	Stopped too soon, applied too long
Brake	UCA-1: Driver does not when auto-parking and computer doesn't react an obstacle			

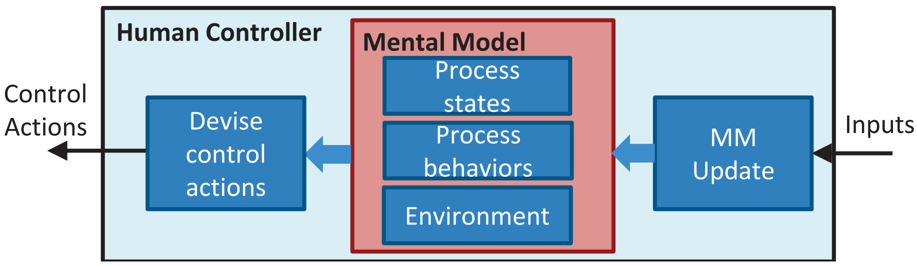


NEW PROCESS

- Identify UCAs
  - UCA-1: Driver does not brake when auto-parking and computer doesn't react to an obstacle
- Identify Mental Model variables
  - PM-1: APA is enabled/disabled
  - PM-2: APA computer reacting appropriately/inappropriately
  - PM-3: Obstacle on collision path
- Identify Mental Model Flaws
- Identify flaws in Mental Model Updates
- Identify unsafe Control Action Selections

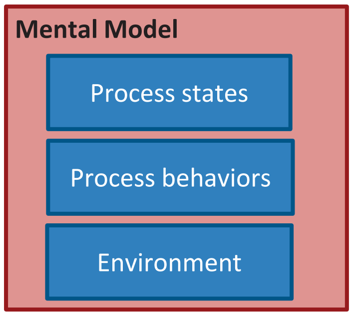
NEW PROCESS

- Identify UCAs
  - UCA-1: Driver does not brake when auto-parking and computer doesn't react to an obstacle
- Identify Mental Model variables
  - PM-1: APA is enabled/disabled
  - PM-2: APA computer reacting appropriately/inappropriately
  - PM-3: Obstacle on collision path
- Identify Mental Model Flaws
- Identify flaws in Mental Model Updates
- Identify unsafe Control Action Selections



NEW PROCESS

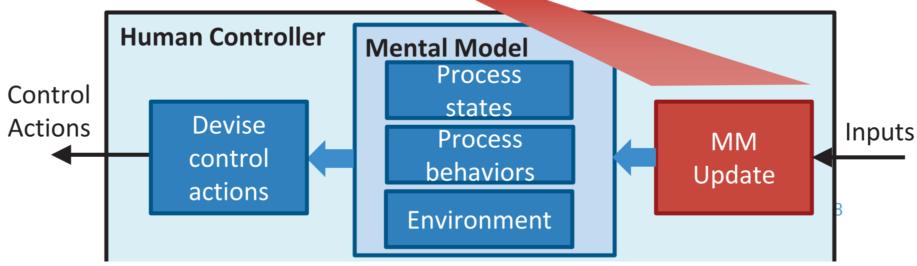
- Identify UCAs
- Identify Mental Model variables
  - PM-1: APA is enabled/disabled
  - PM-2: APA computer reacting appropriately/inappropriately
  - PM-3: Obstacle on collision path
- Identify Mental Model Flaws
- Identify unsafe decisions (Control Action Selections)
- Identify inadequate Mental Model Updates



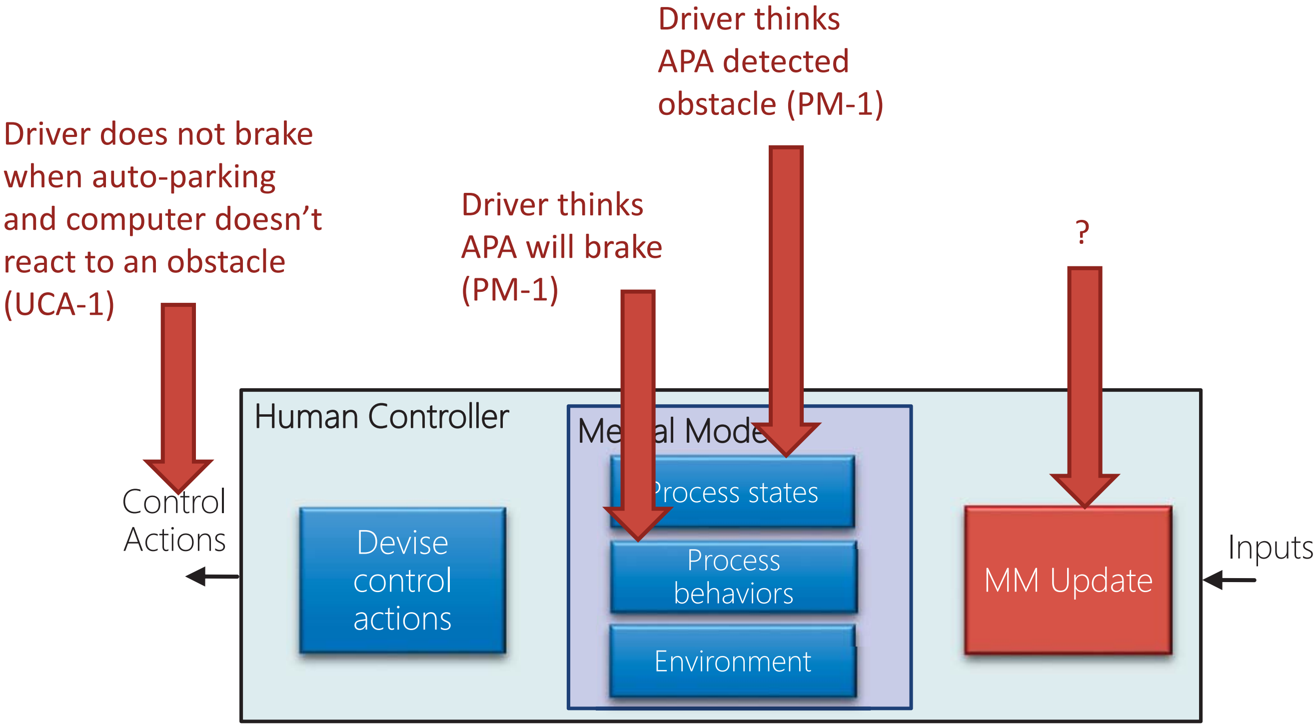
Type of MM flaw	Examples
Incorrect beliefs about process state (including modes)	Driver thinks APA is enabled when APA is really disabled
Incorrect beliefs about process behaviors	Driver thinks APA is reacting properly and will brake automatically
Incorrect beliefs about environment	Driver thinks there is no obstacle when there is one Driver knows there is an obstacle but doesn't know it's on a collision path

NEW PROCESS

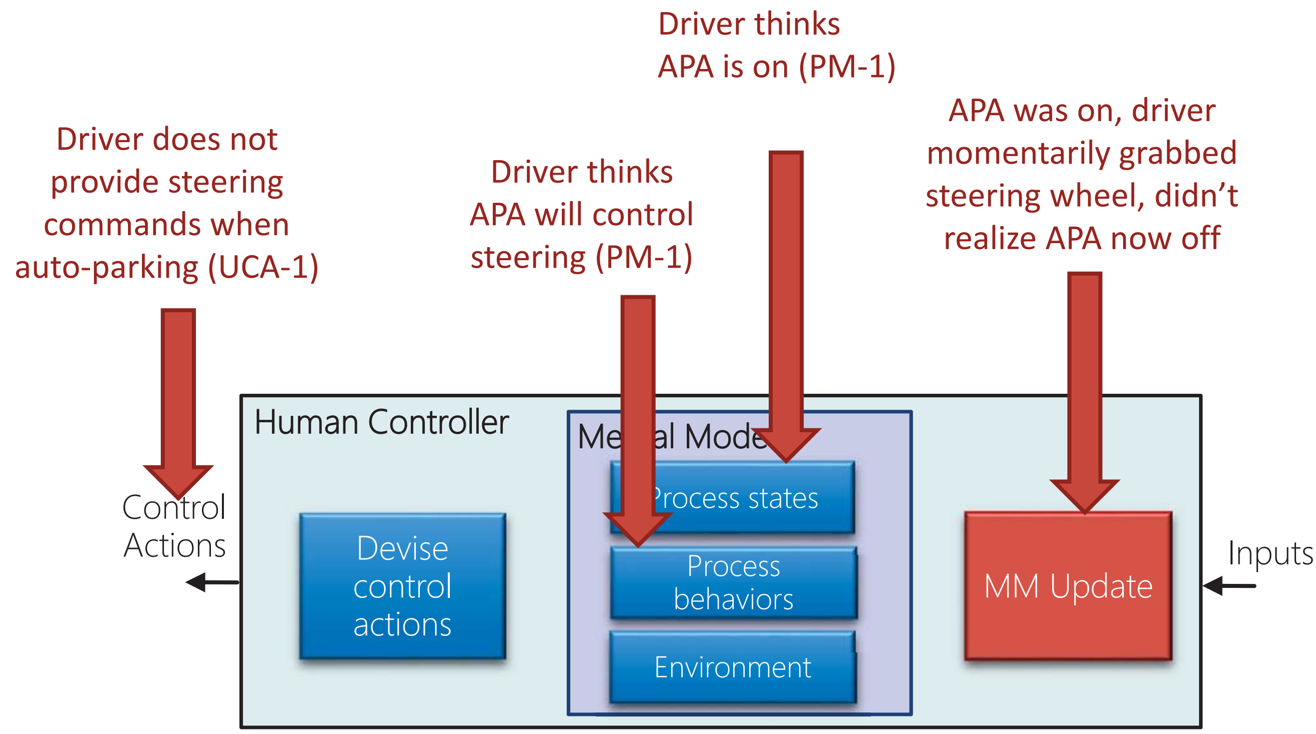
- Identify UCAs
  - UCA-1: Driver does not brake when auto-parking and computer doesn't react to an obstacle
- Identify Mental Model variables
  - PM-1: APA is enabled/disabled
  - PM-2: APA computer reacting appropriately/inappropriately
  - PM-3: Obstacle on collision path
- Identify Mental Model Flaws
- Identify flaws in Mental Model Updates
- Identify unsafe Control Action Selections



NEW PROCESS

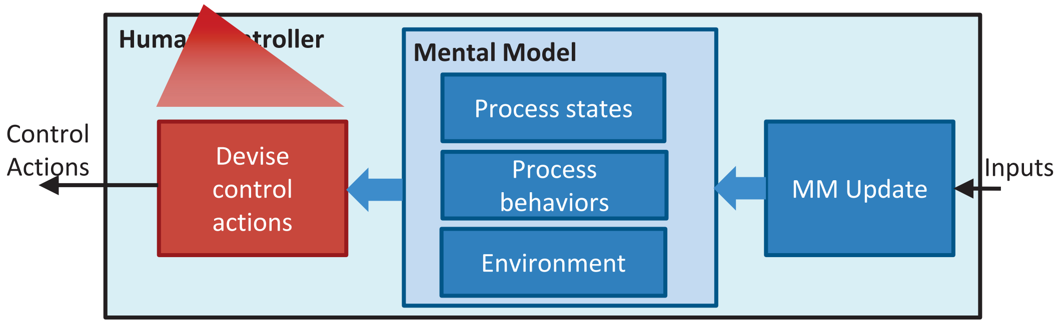


NEW PROCESS



NEW PROCESS

- Identify UCAs
  - UCA-1: Driver does not brake for an obstacle when computer does not react appropriately to the obstacle
- Identify Mental Model variables
  - PM-1: APA reacting appropriately/inappropriately
  - PM-2: Obstacle on collision path
- Identify Mental Model Flaws
- Identify flaws in Mental Model Updates
- Identify unsafe Control Action Selections



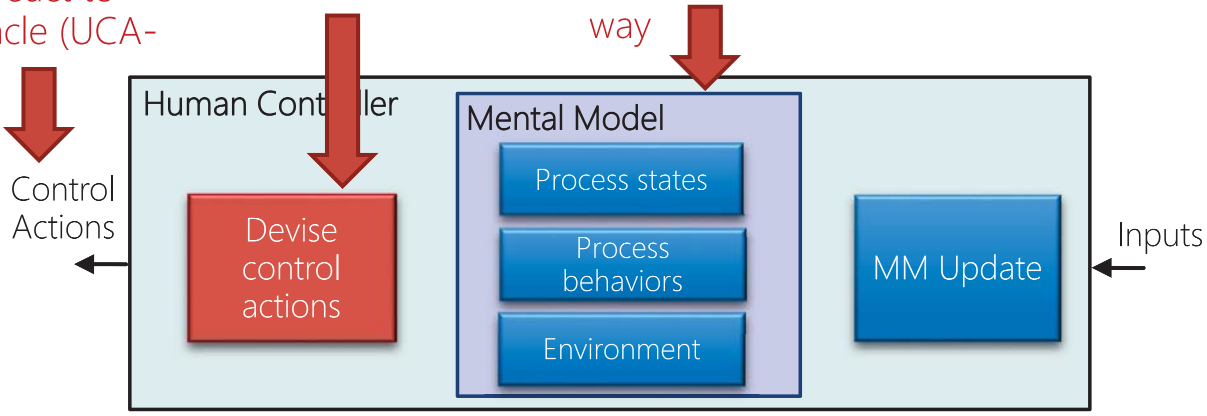
NEW PROCESS

- Identify unsafe Control Action Selections

Driver does not brake when auto-parking and computer doesn't react to an obstacle (UCA-1)

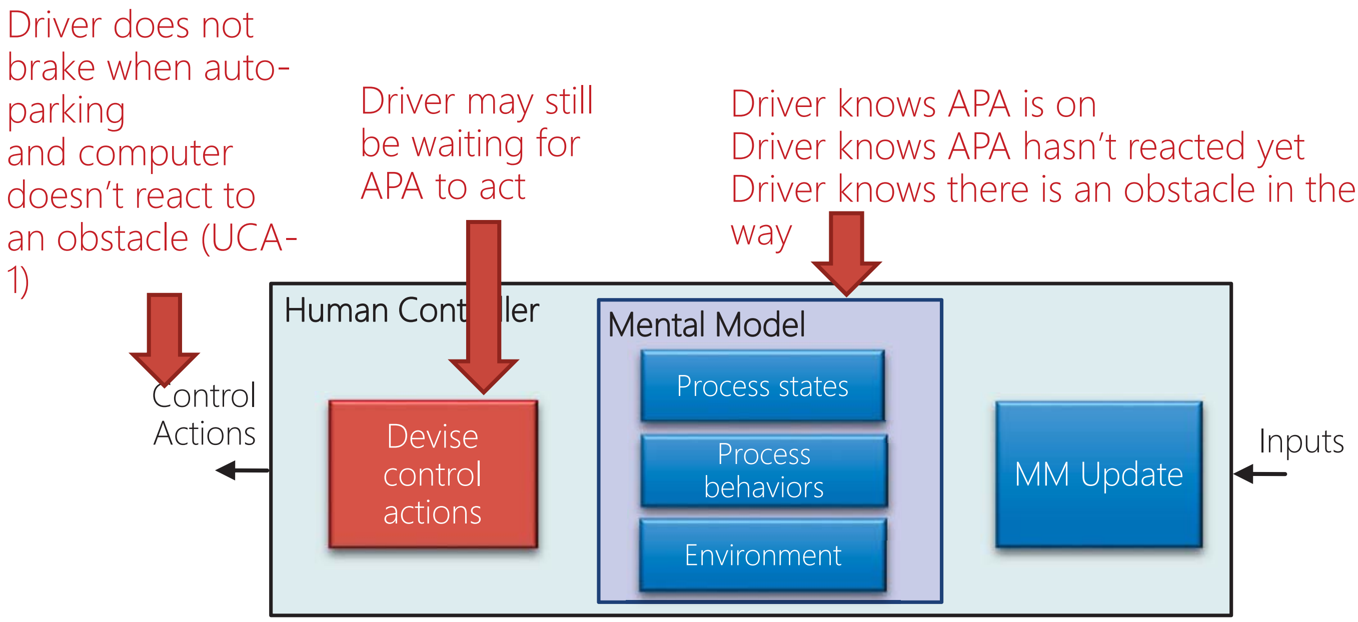
Maybe driver decides to disable APA instead

Driver knows APA is on
Driver knows APA hasn't reacted yet
Driver knows there is an obstacle in the way



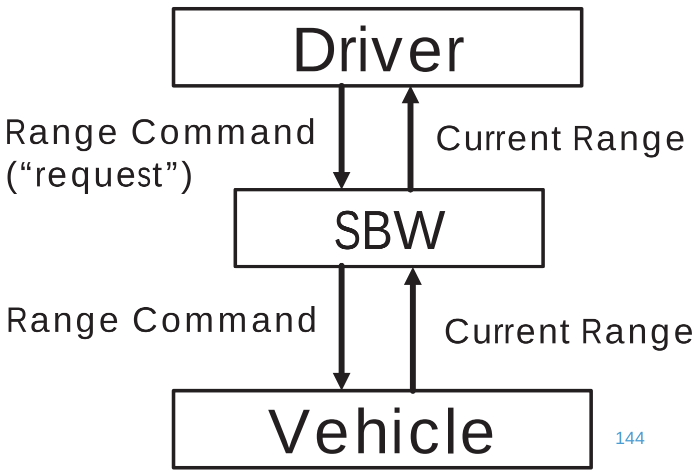
NEW PROCESS

- Identify unsafe Control Action Selections

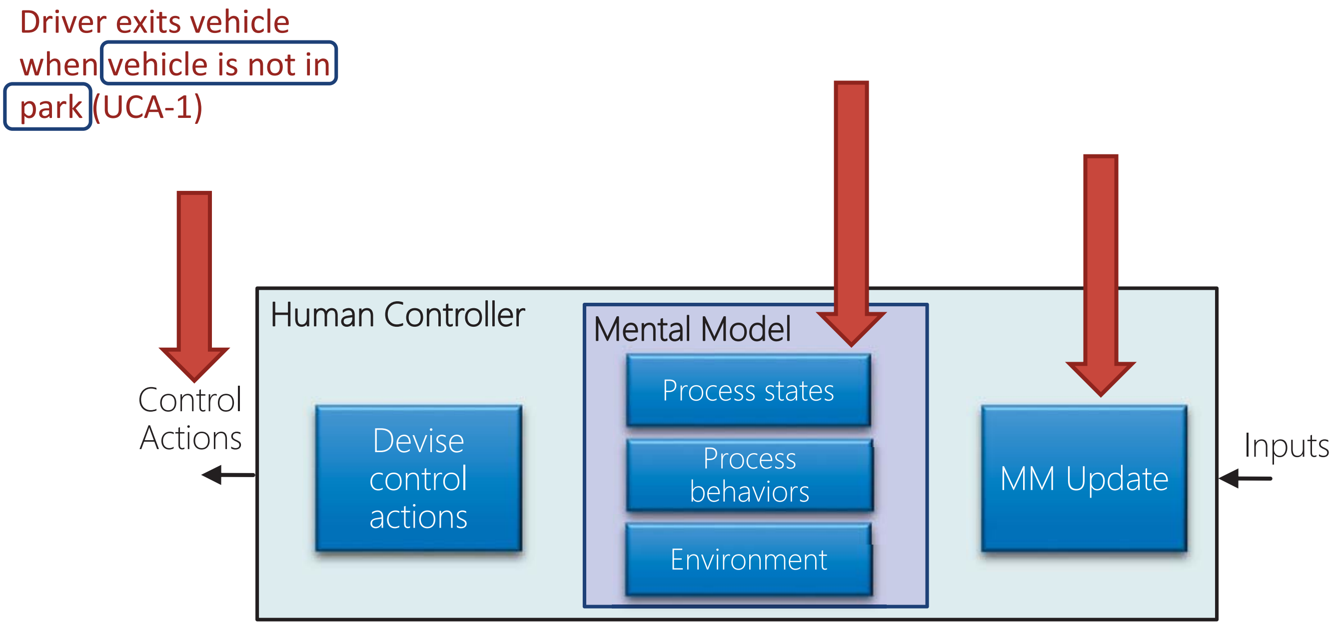


Range =

- Park
- Reverse
- Neutral
- Drive
- Etc.

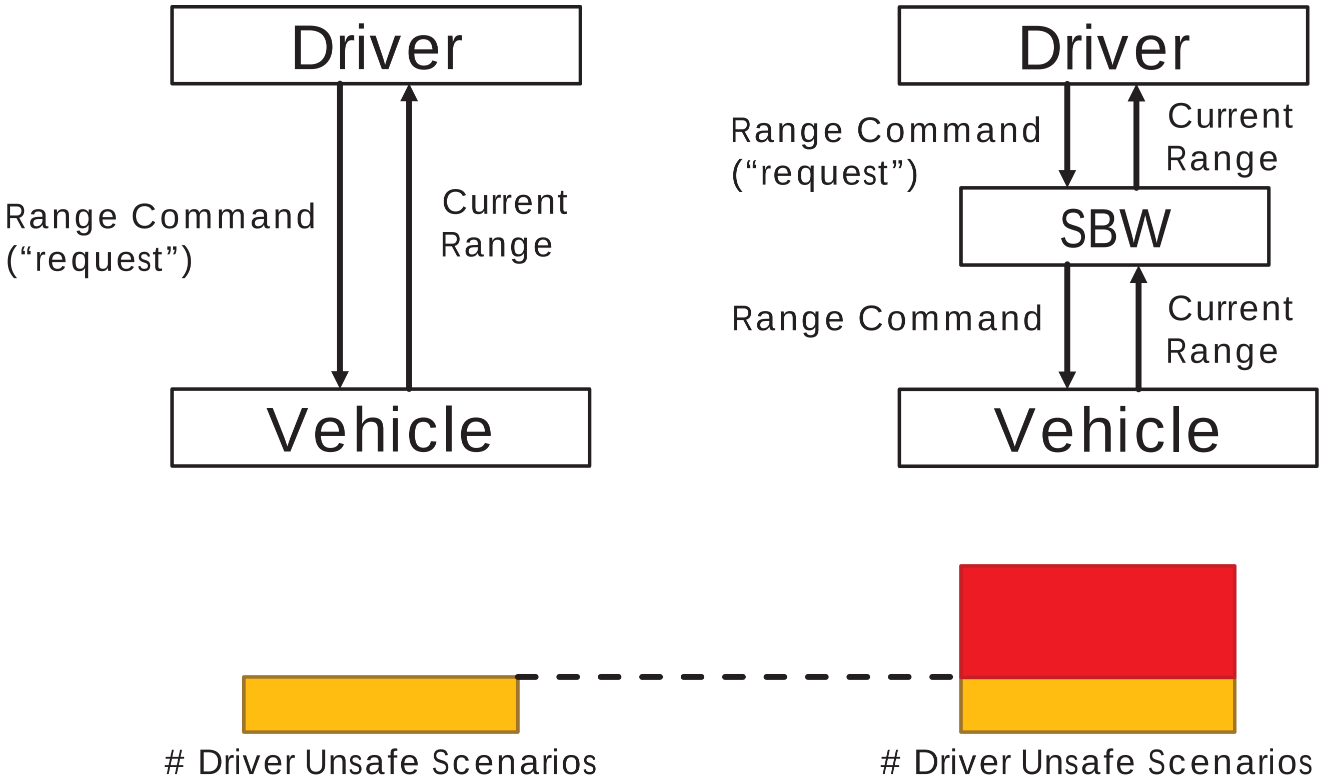


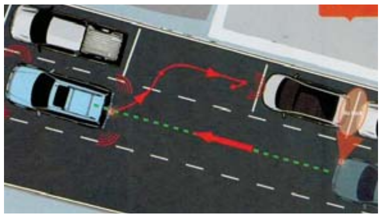
NEW PROCESS



Old System

New System





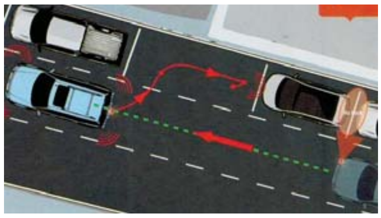
AUTOMATED PARKING

Features of each system considered for this analysis:

	Level 0* No Driving Automation	Level 1 "Driver Assistance"	Level 2a "Partial Automation"	Level 2b "Partial Automation"	Level 3 "Conditional Automation"
Steering	-	✓	✓	✓	✓
Braking	-	-	✓	✓	✓
Shifting and Acceleration	-	-	-	✓	✓
Object and Event Detection and Response	-	-	-	-	✓

*System numbering is consistent with SAE definitions for levels of automation, while "a" and "b" indicate different implementations which are classified within the same SAE level.

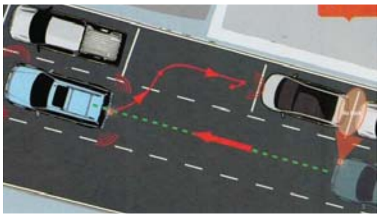
Analysis reuse



AUTOMATED PARKING

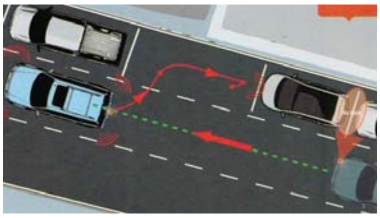
	Level 1 "Driver Assistance"	Level 2a "Partial Automation"	Level 2b "Partial Automation"	Level 3 "Conditional Automation"
Driver UCAs				
APA Computer UCAs				
Total				

AUTOMATED PARKING



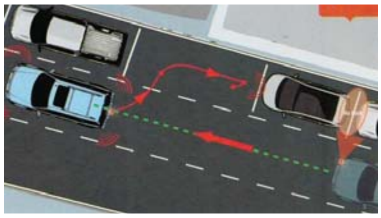
	Level 1 "Driver Assistance"	Level 2a "Partial Automation"	Level 2b "Partial Automation"	Level 3 "Conditional Automation"
Driver UCAs				
APA Computer UCAs	5	13	28	28
Total				

AUTOMATED PARKING



	Level 1 "Driver Assistance"	Level 2a "Partial Automation"	Level 2b "Partial Automation"	Level 3 "Conditional Automation"
Driver UCAs	42	41	38	44
APA Computer UCAs	5	13	28	28
Total				

AUTOMATED PARKING



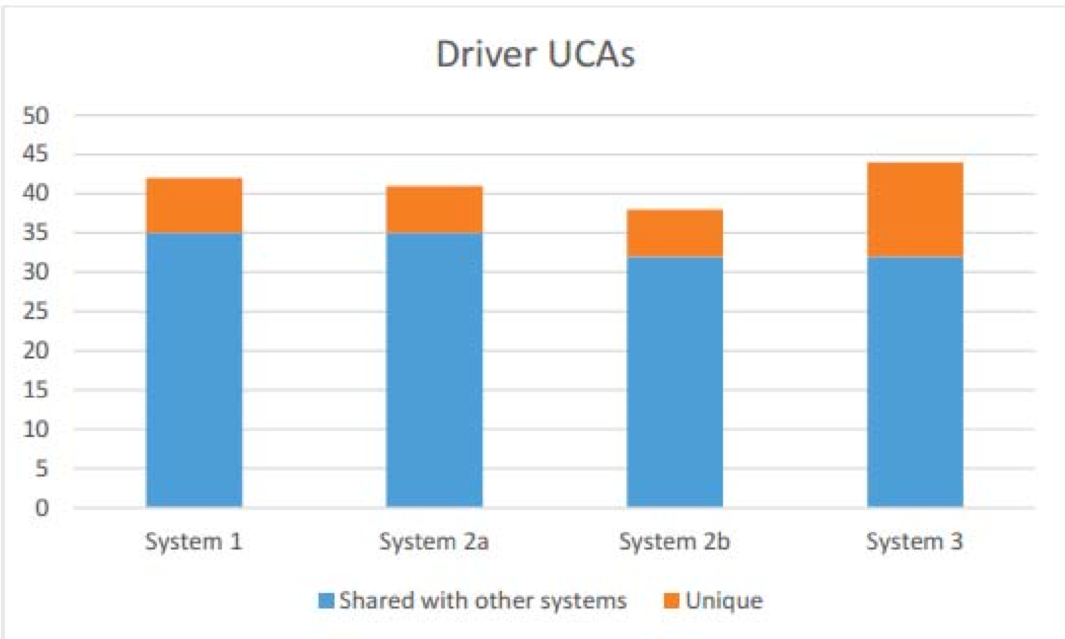
	Level 1 "Driver Assistance"	Level 2a "Partial Automation"	Level 2b "Partial Automation"	Level 3 "Conditional Automation"
Driver UCAs	42	41	38	44
APA Computer UCAs	5	13	28	28
Total	47	54	66	72

Driver UCAs: 35 in common (Level 1 / Level 2a); 30 in common (Level 2a / Level 2b); 32 in common (Level 2b / Level 3)

APA Computer UCAs: 5 in common (Level 1 / Level 2a); 13 in common (Level 2a / Level 2b); 28 in common (Level 2b / Level 3)

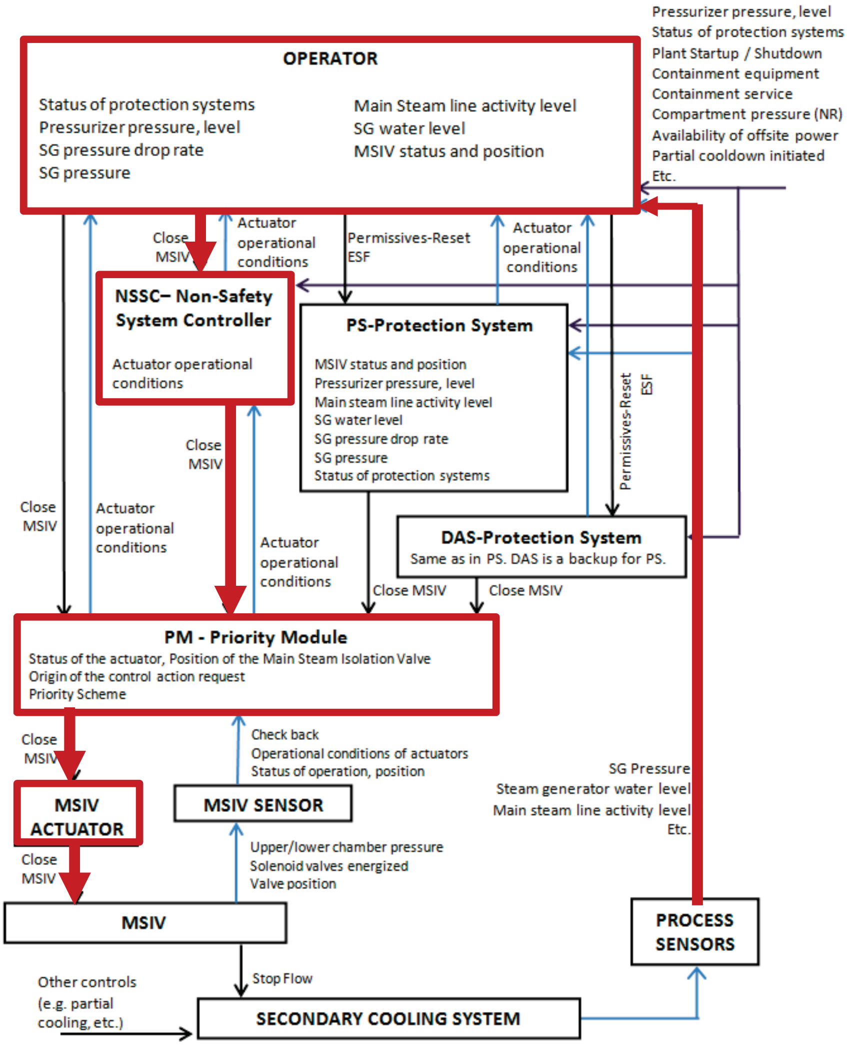
Total: 40 in common (Level 1 / Level 2a); 43 in common (Level 2a / Level 2b); 60 in common (Level 2b / Level 3)

	Level 1	Level 2a	Level 2b	Level 3
Driver UCAs	42	41	38	44
APA Computer UCAs	5	13	28	28
Total	47	54	66	72



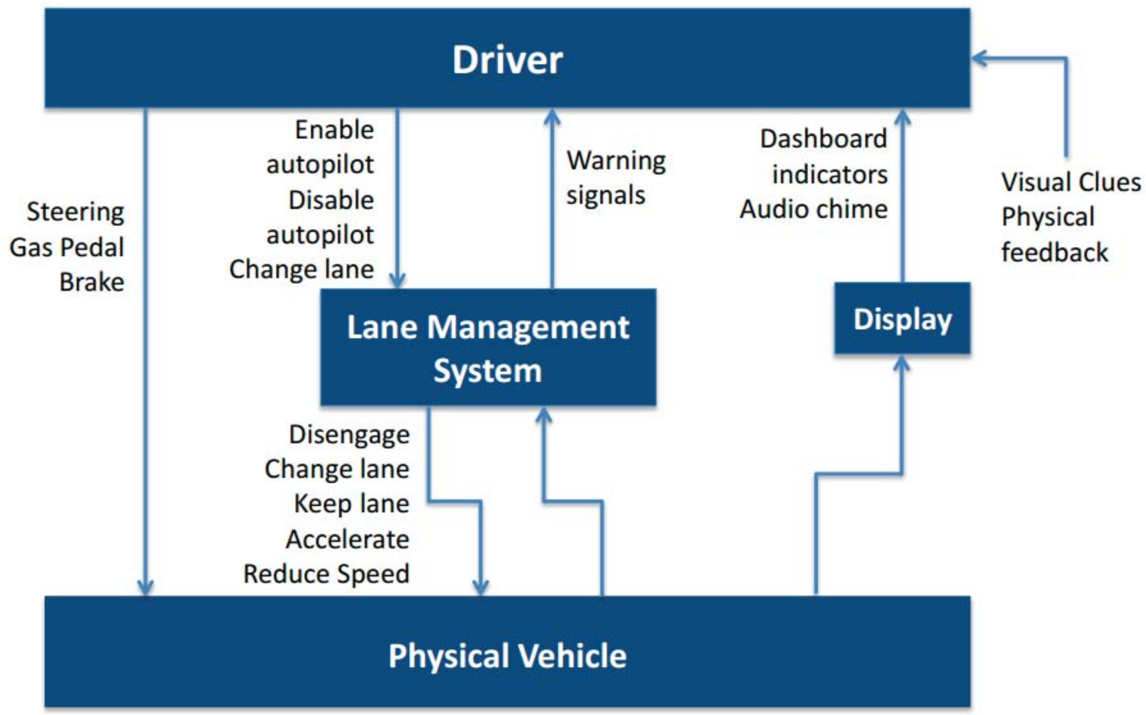
Nuclear power example

Real
safety & security
issues identified



Tesla Autopilot example

Tesla Autopilot



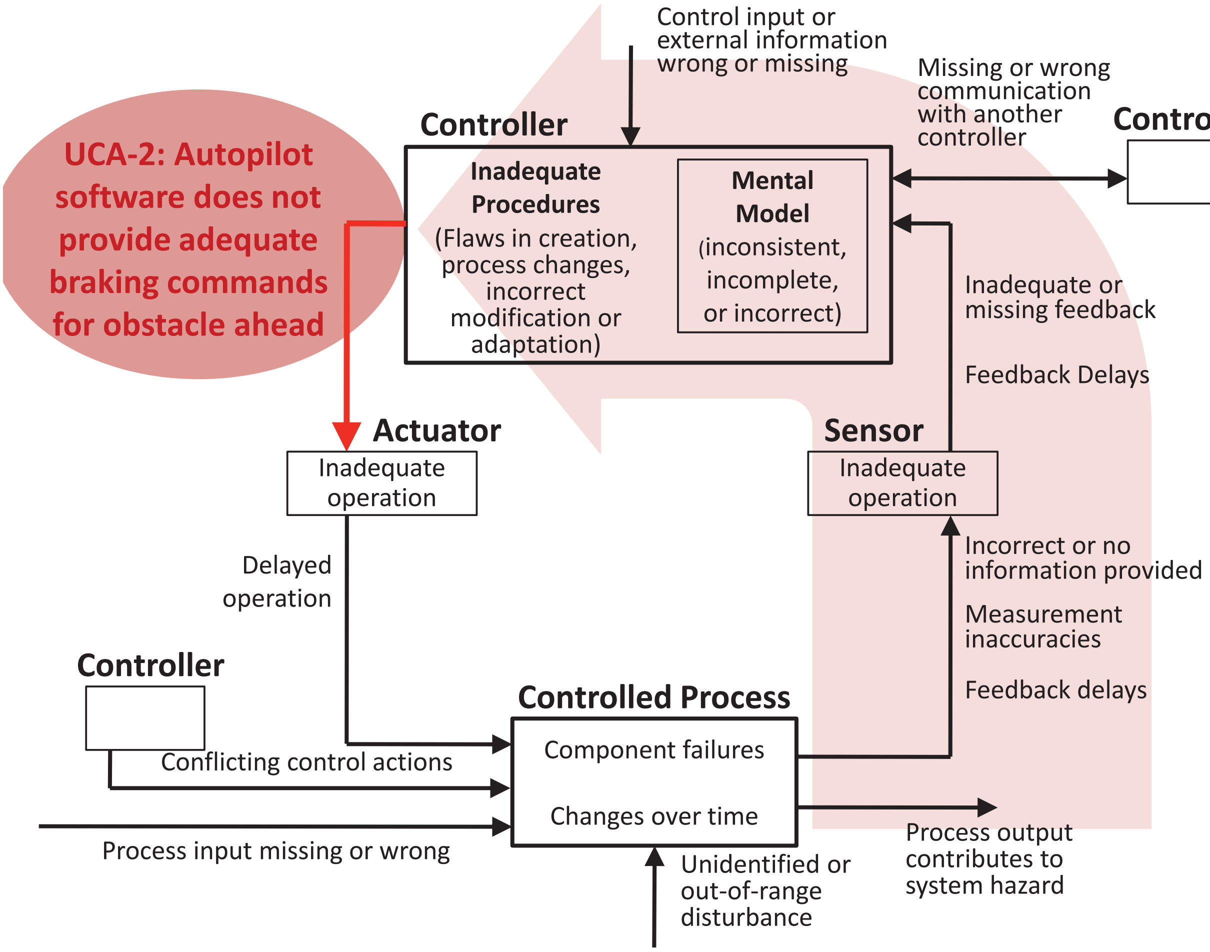
Spring 2016 Student project: Diogo Castilho, Megan France

Tesla Autopilot

Controller	Control Action	Not providing causes hazards	Providing causes hazards	Incorrect Timing / Order	Stopped too soon / Applied too long
Driver	**Steering**	-	UCA-7: Driver provides steering can cause hazards if autopilot is changing the lane to the opposite direction	-	-
Driver	**Steering**	UCA-8: Driver does not provide steering to avoid obstacles when autopilot does not react	-	-	-
Auto-Pilot	**Lane changing**	UCA-13: Auto-pilot Not providing lane changing automatically causes hazards	-	-	-
Auto-Pilot	**Reduce Speed**	UCA-17: Auto-pilot does not provide reducing speed can cause hazards if range and range rate of current vehicle is above the limit	-	-	-

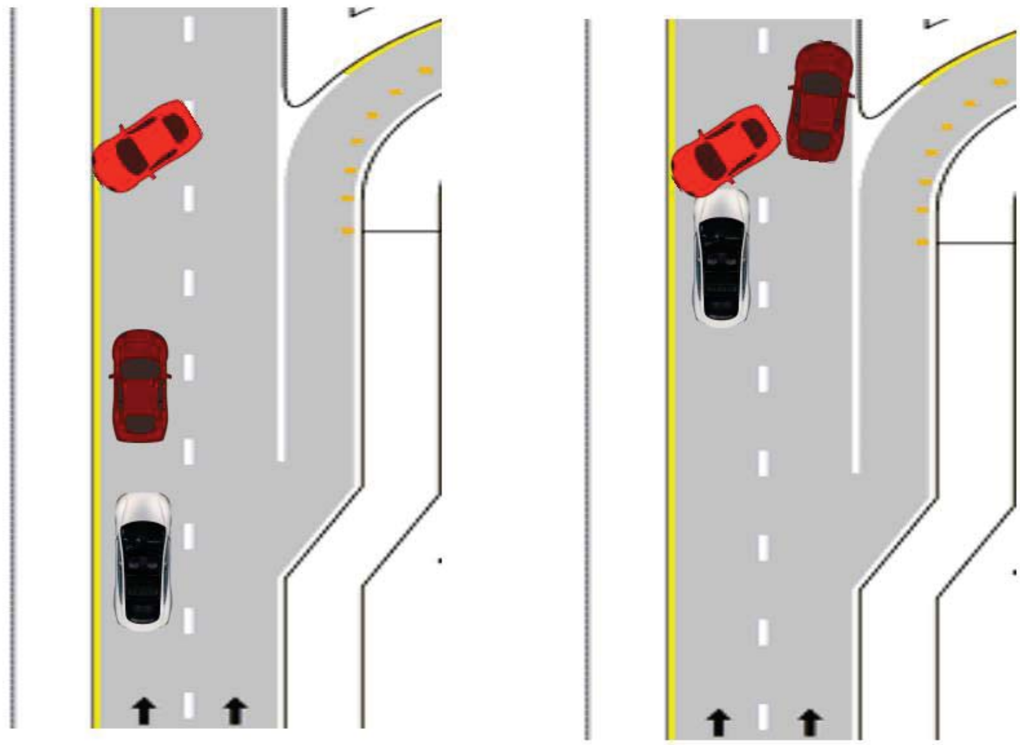
Spring 2016 Student project: Diogo Castilho, Megan France

Step 2A: Potential causes of UCAs



Tesla Autopilot

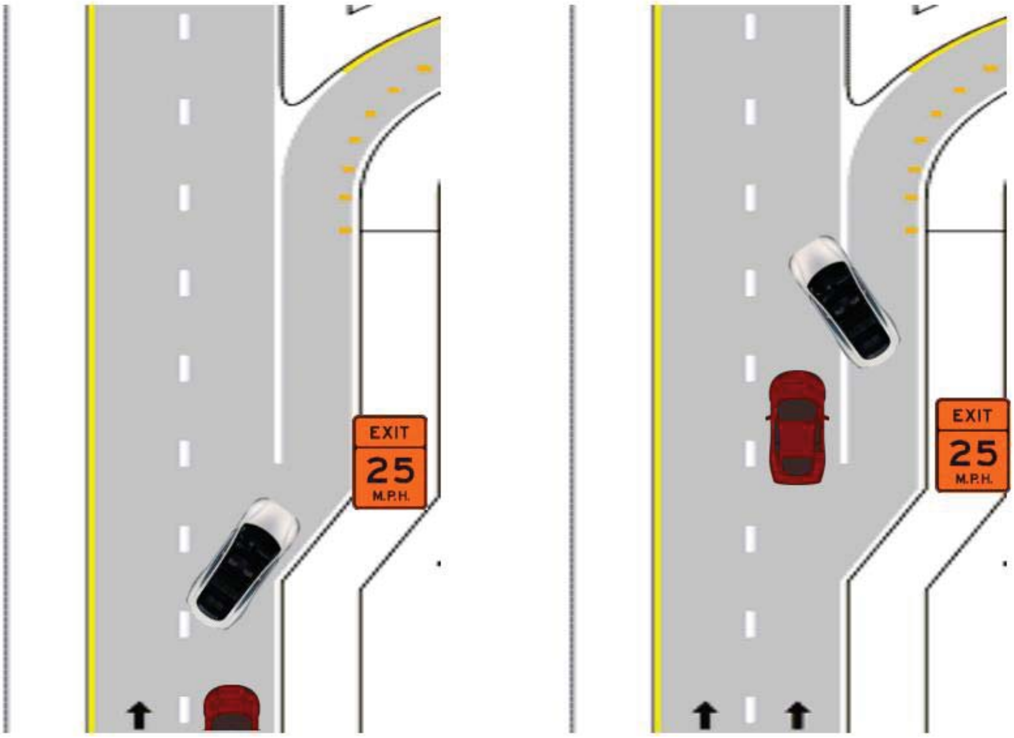
UCA-2: Autopilot does not provide adequate braking commands for obstacle ahead



Spring 2016 Student project: Diogo Castilho, Megan France

Tesla Autopilot

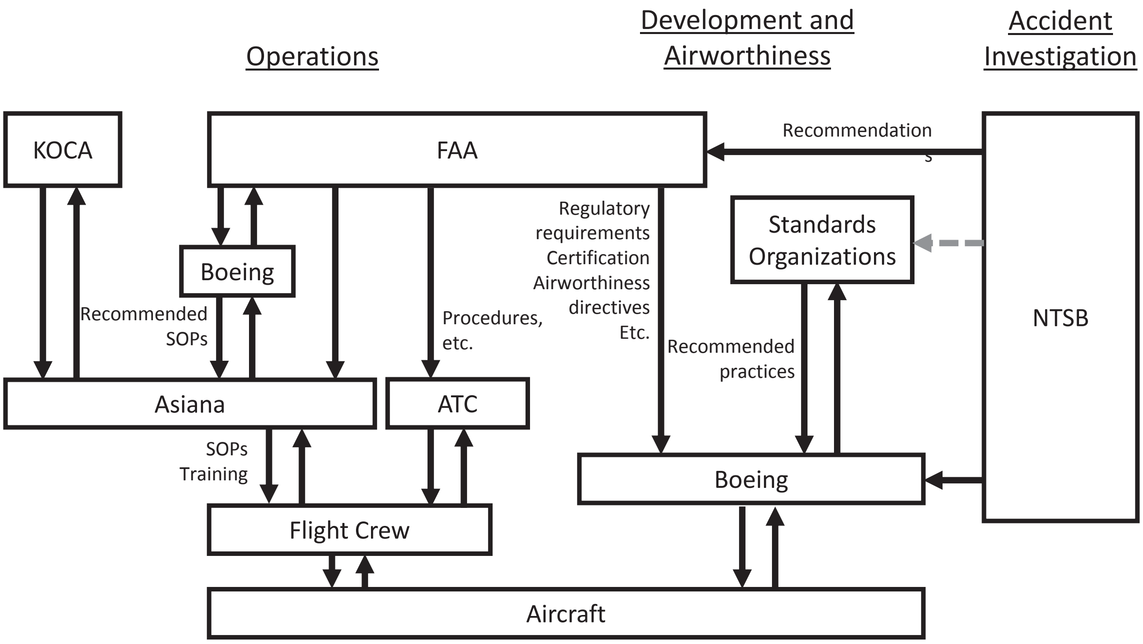
UCA-1: Driver provides unsafe steering override commands when autopilot is engaged



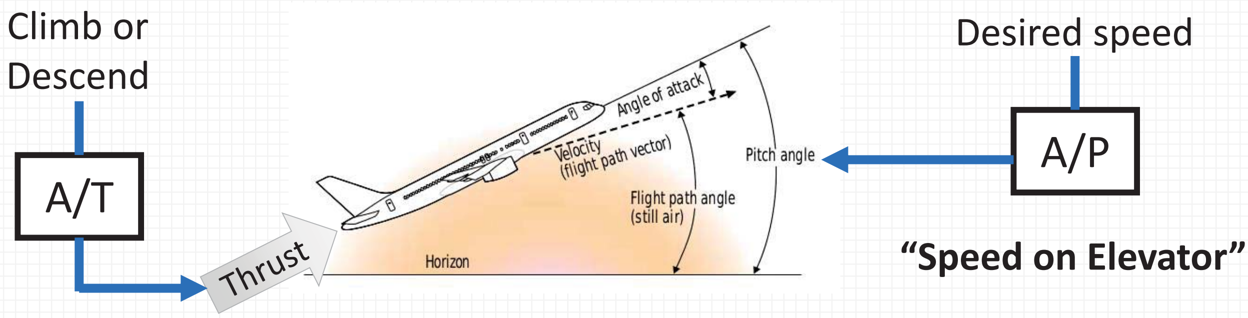
Spring 2016 Student project: Diogo Castilho, Megan France

Accident/Incident Analysis

Accident Analysis: Asiana 214



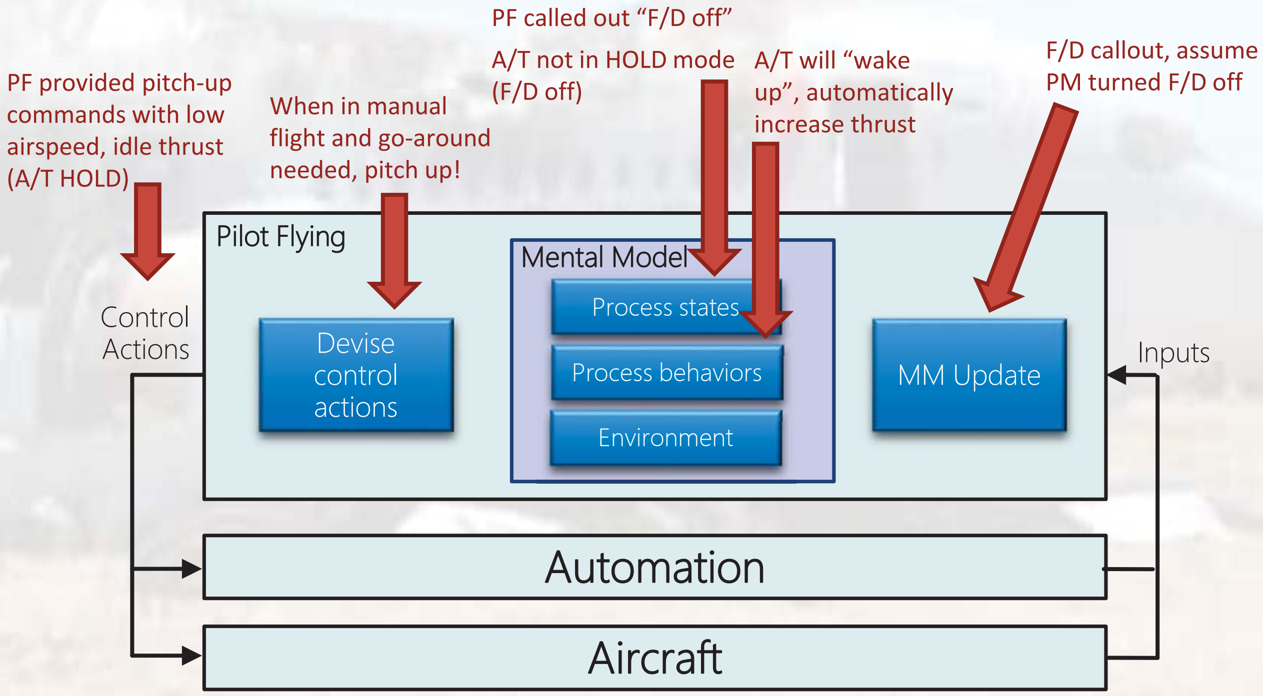
Autopilot (A/P) and Autothrottle (A/T) Pairing



A/T will remain in HOLD mode until one of the following conditions is met:

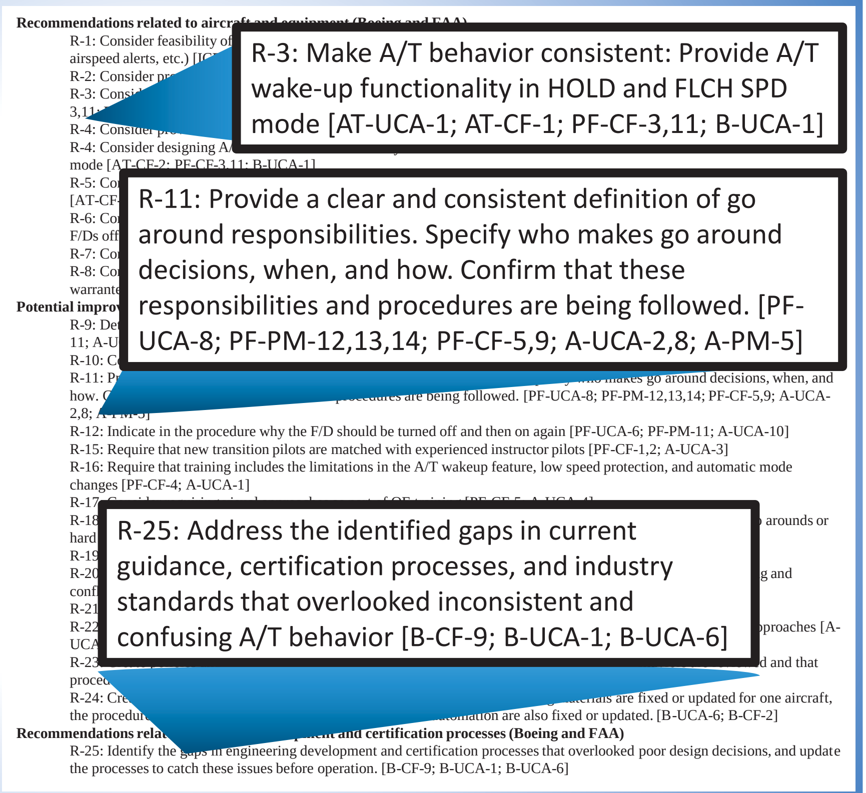
- The airplane reaches the MCP target altitude
- The pilot engages a new AFDS pitch mode or new A/T mode
- The A/T arm switches are turned off
- The thrust is manually commanded to increase past the thrust limit
- The A/P is disconnected, and both F/D switches are turned off

Analyzing controllers: Pilot Flying



CAST Recommendations

- 25 recommendations
  - Technical design
  - Procedural
  - Regulatory

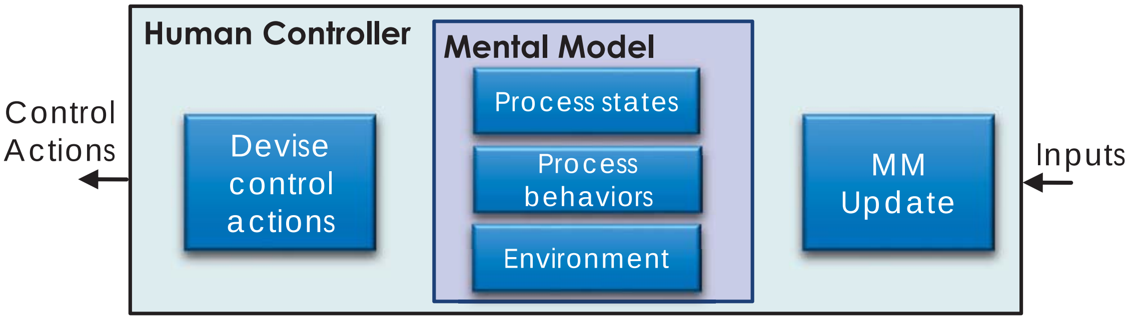


Leading indicators

Findings

- Most CAST rec's not included in NTSB rec's
  - Exception: low energy alerting system recommended by both
- Systematic methodology to:
  - Organize, make sense of complex accidents
  - Ensure deeper systemic factors are examined
  - Help guide less experienced teams
  - Help overcome human biases
  - Ensure causal factors and recommendations aren't overlooke

CONCLUSIONS



New human engineering extension strengths:

- Easy to learn, use
- Applicable to accident analysis and engineering
- Use early to drive requirements and concepts from the start
- Applicable earlier than detailed simulations or prototypes
- Successful in industry, adoption