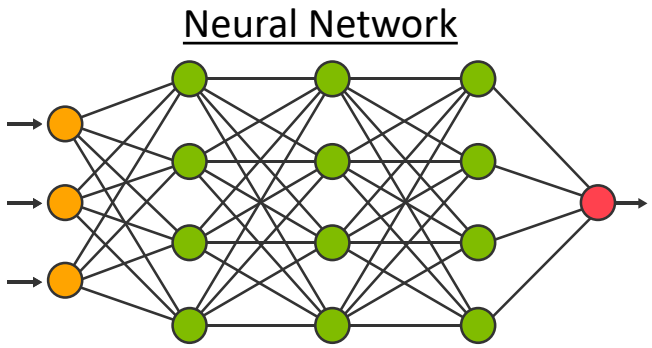
STPA APPLIED TO A NEURAL NETWORK-CONTROLLED AIRCRAFT

RYAN BOWERS – 40TH FLIGHT TEST SQUADRON

DR. JOHN THOMAS – MASSACHUSETTS INSTITUTE OF TECHNOLOGY

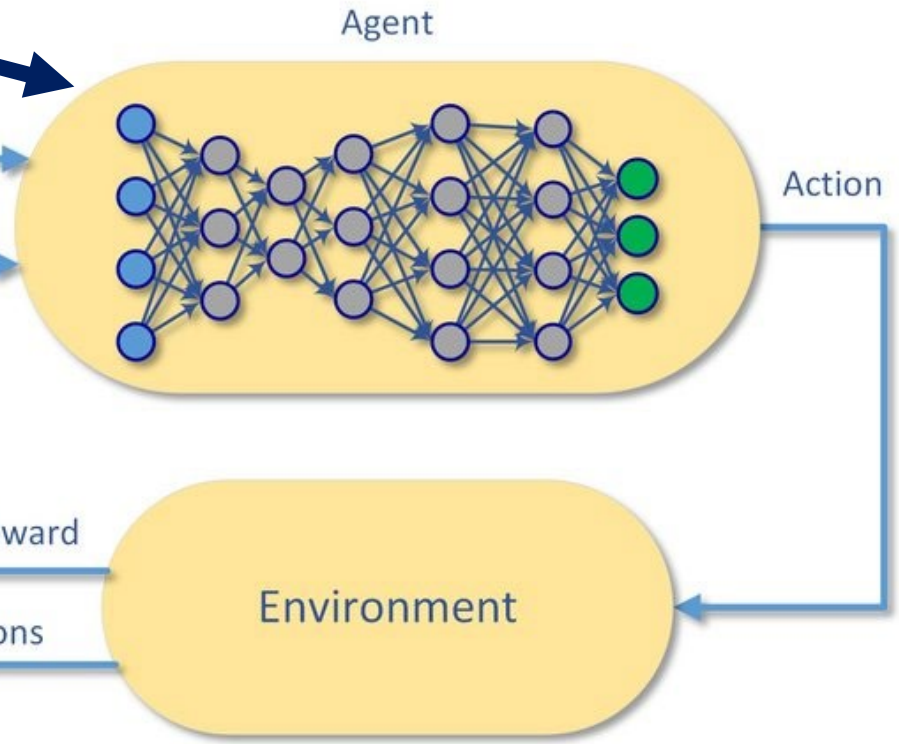
Overview

- 40th FLTS: flight test for AI-enabled autonomous aircraft
- July 2023: First flight test of a group-5 UAV flown by machine learning agents
- Agents trained using deep reinforcement learning
- Applied STPA before flight test



Deep Reinforcement Learning Agents

Computes actions based on environment observations



Flight controls
n_z, roll rate, throttle

Environment state
- UAV state
- Safety limits
- External entities

System Considerations for Safety

- ML agents can be difficult to explain
- Agents trained in simulation, then transitioned to real life
- UAV and agents developed under completely separate programs before integrating

Three-Pronged Flight Test Safety Approach

(1) UAV Mechanisms

Envelope trips: Disables agent if speed/altitude limits exceeded

Command Limiters: Agent control inputs are clipped to stay within min and max bounds.

(2) Autonomy Mechanisms

Simulation Training: Agents were trained to stay within limits.

Redundant envelope trips: Agent disables itself if limits exceeded.

(3) Test Procedures

Manual Disable: Remote pilot can disable agent at anytime.

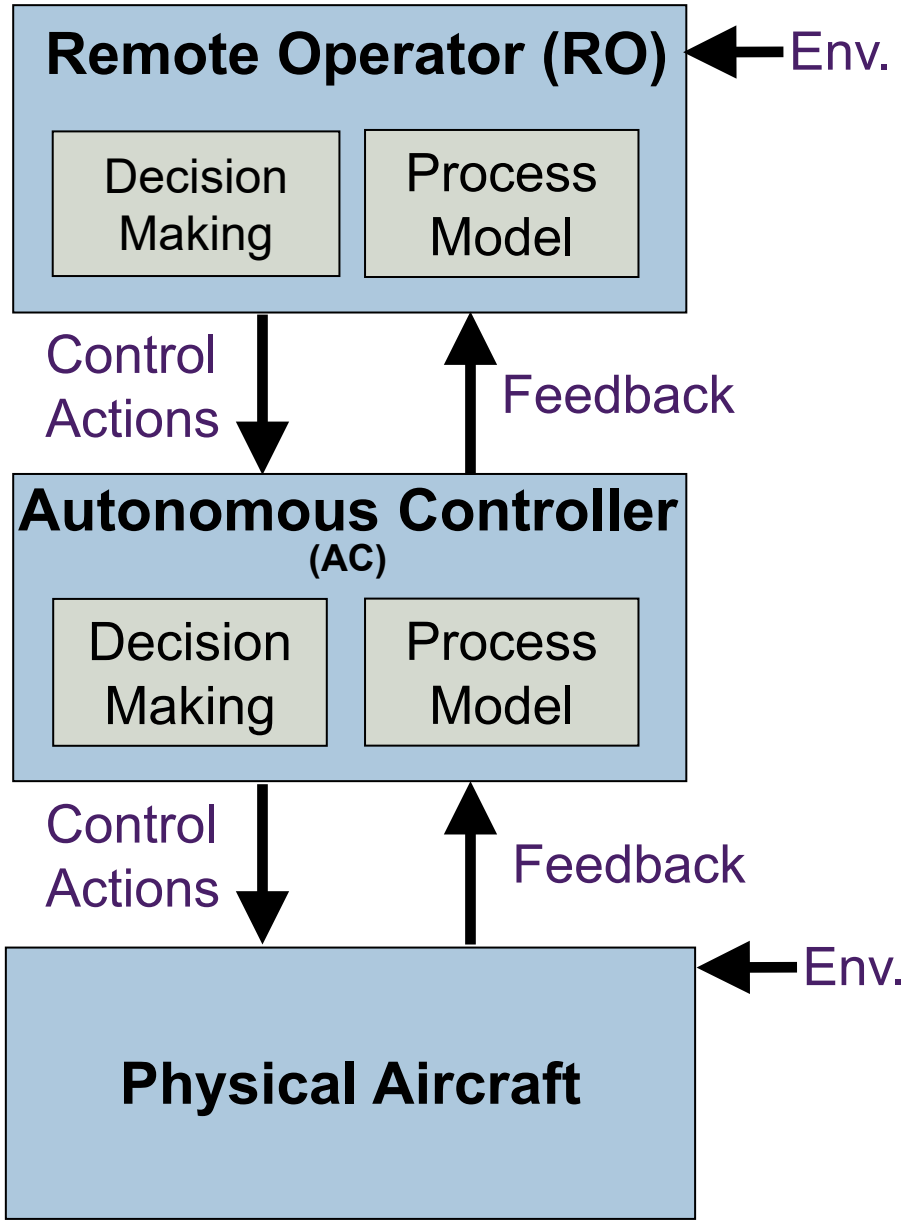
Abort Limits: Manually disable if any limits exceeded.

Briefing Items: Team briefed on possible unsafe agent behavior.

STPA: System Theoretic Process Analysis

E.g., X-47B UAV integrated into carrier operations alongside manned aircraft. Provides autonomous launch, flight, follow manned A/C, carrier landing, etc.

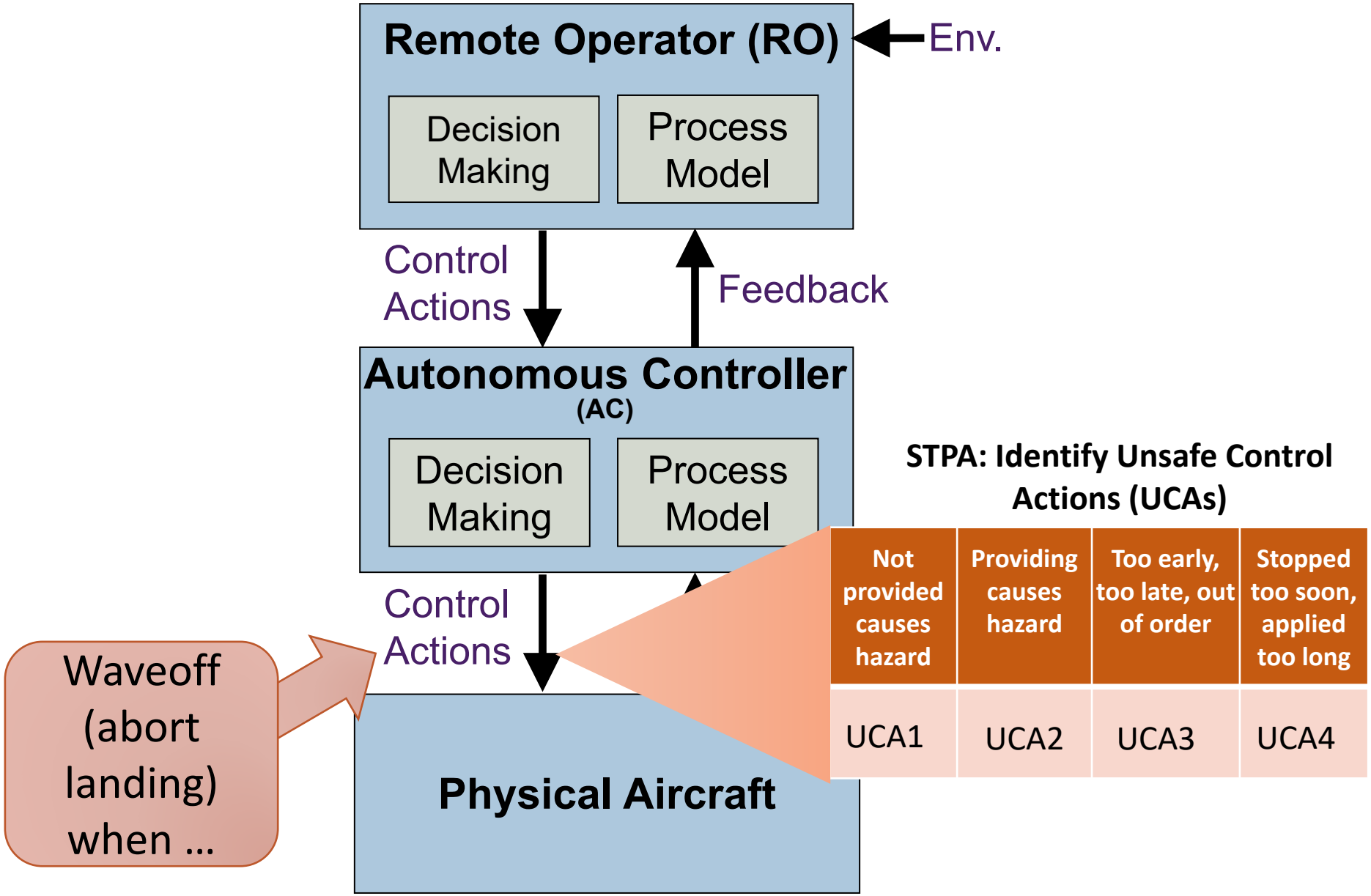
What can go wrong in flight test?



(John Thomas, 2021)

STPA Step 3: Identify UCAs

E.g., X-47B UAV integrated into carrier operations alongside manned aircraft. Provides autonomous launch, flight, follow manned A/C, carrier landing, etc.



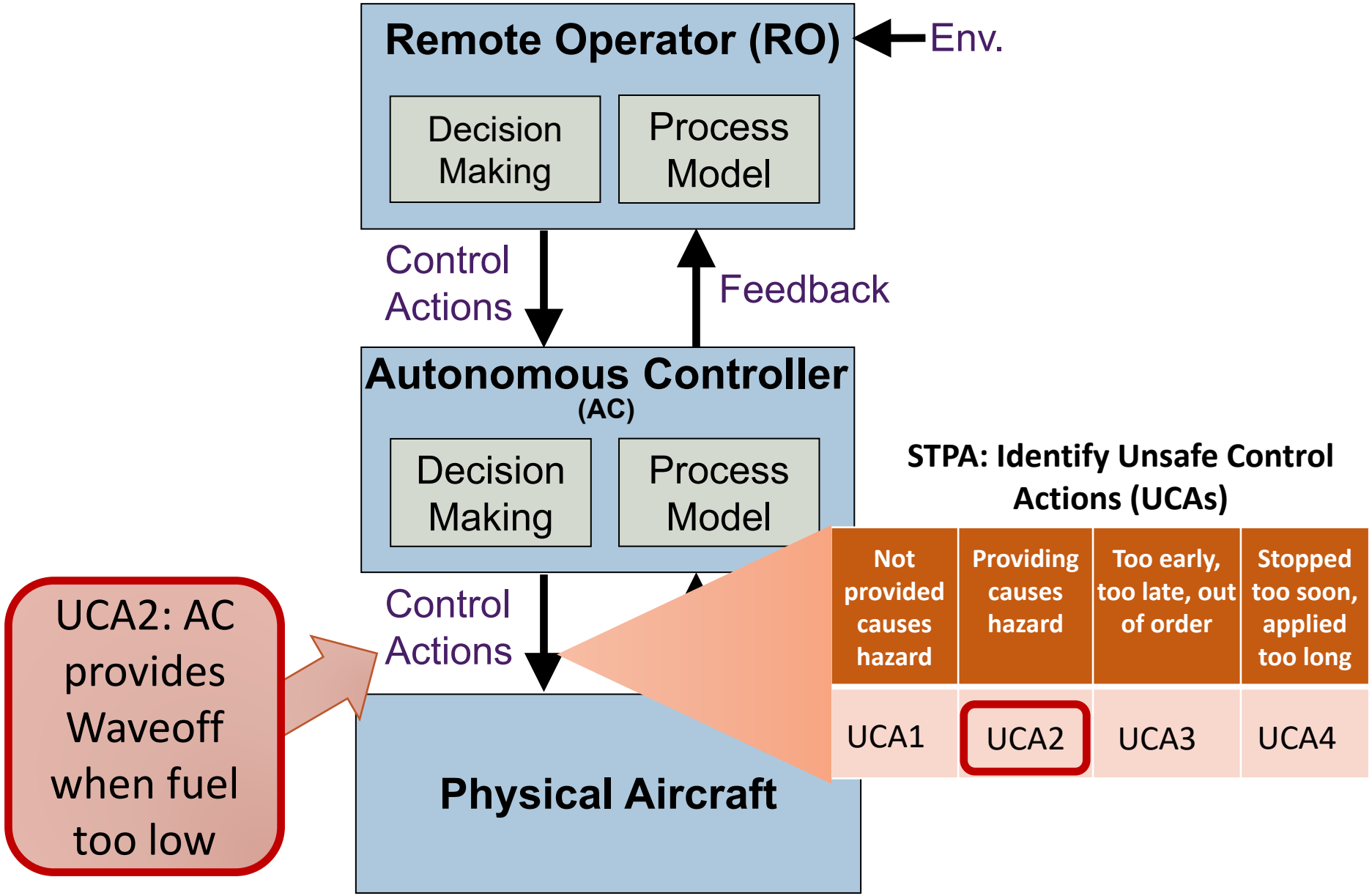
STPA: Identify Unsafe Control Actions (UCAs)

Not provided causes hazard	Providing causes hazard	Too early, too late, out of order	Stopped too soon, applied too long
UCA1	UCA2	UCA3	UCA4

(John Thomas, 2021)

STPA Step 3: Identify UCAs

E.g., X-47B UAV integrated into carrier operations alongside manned aircraft. Provides autonomous launch, flight, follow manned A/C, carrier landing, etc.



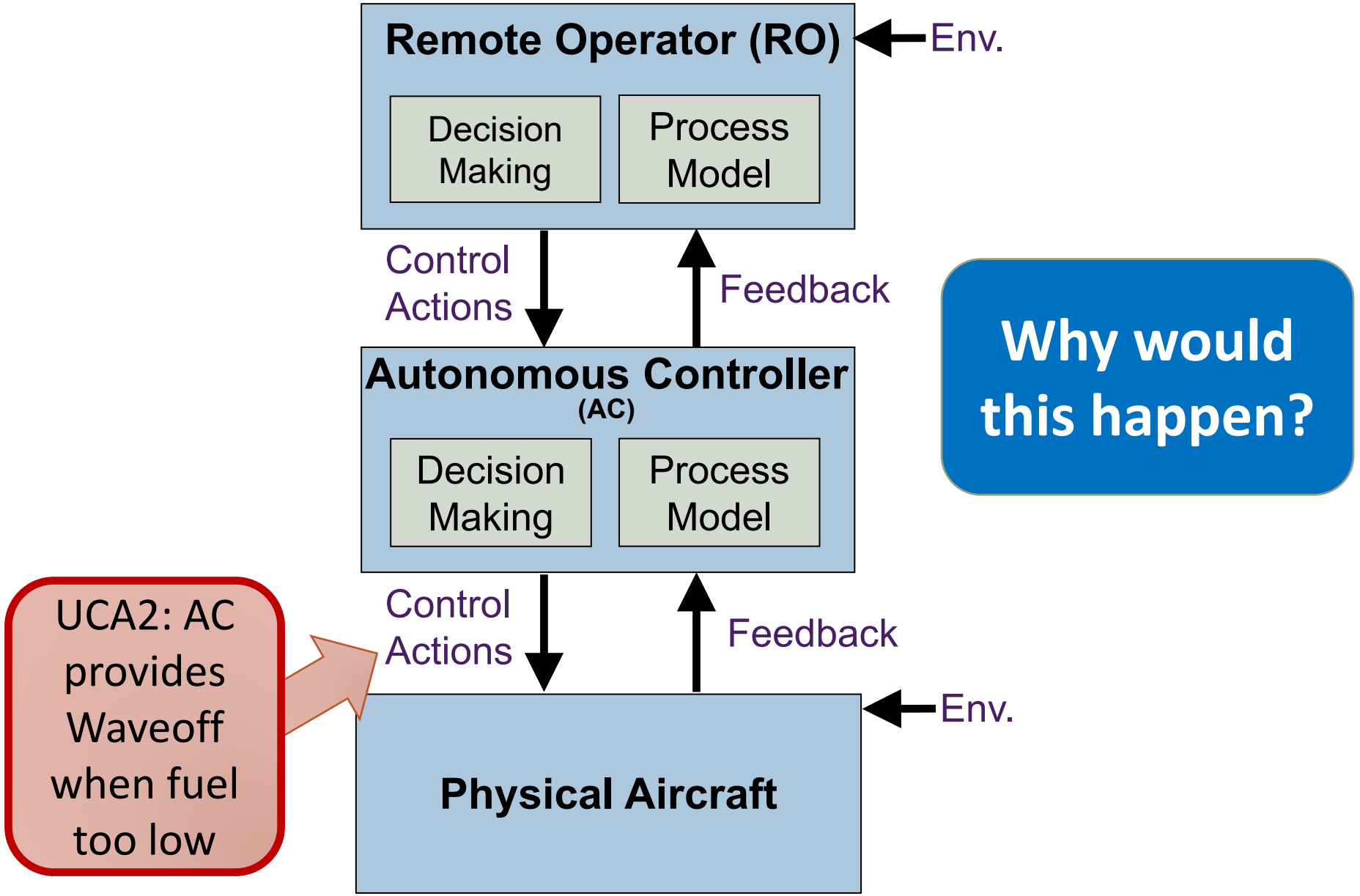
STPA: Identify Unsafe Control Actions (UCAs)

Not provided causes hazard	Providing causes hazard	Too early, too late, out of order	Stopped too soon, applied too long
UCA1	UCA2	UCA3	UCA4

(John Thomas, 2021)

STPA Step 3: Identify UCAs

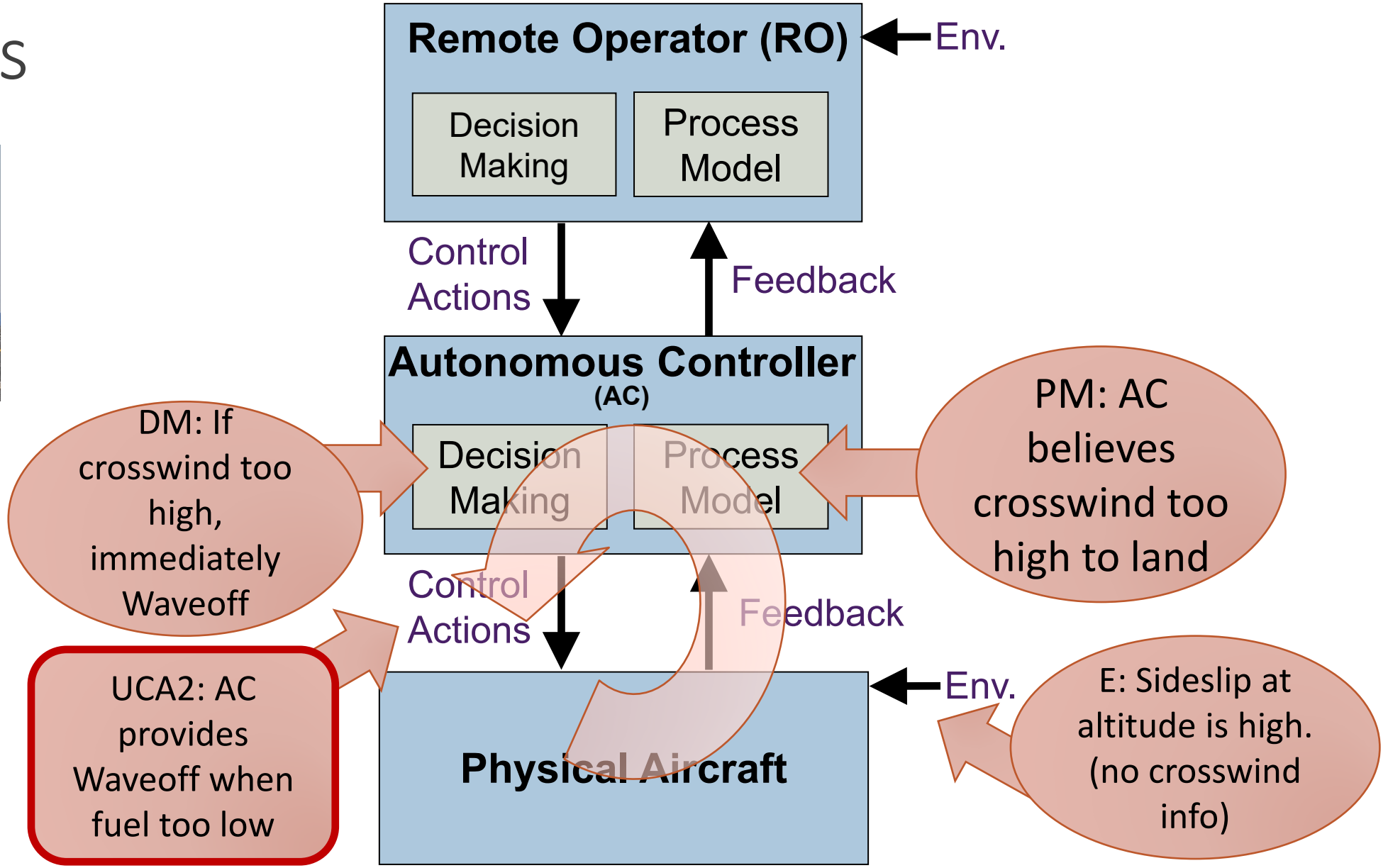
E.g., X-47B UAV integrated into carrier operations alongside manned aircraft. Provides autonomous launch, flight, follow manned A/C, carrier landing, etc.



(John Thomas, 2021)

STPA Step 3: Identify UCAs

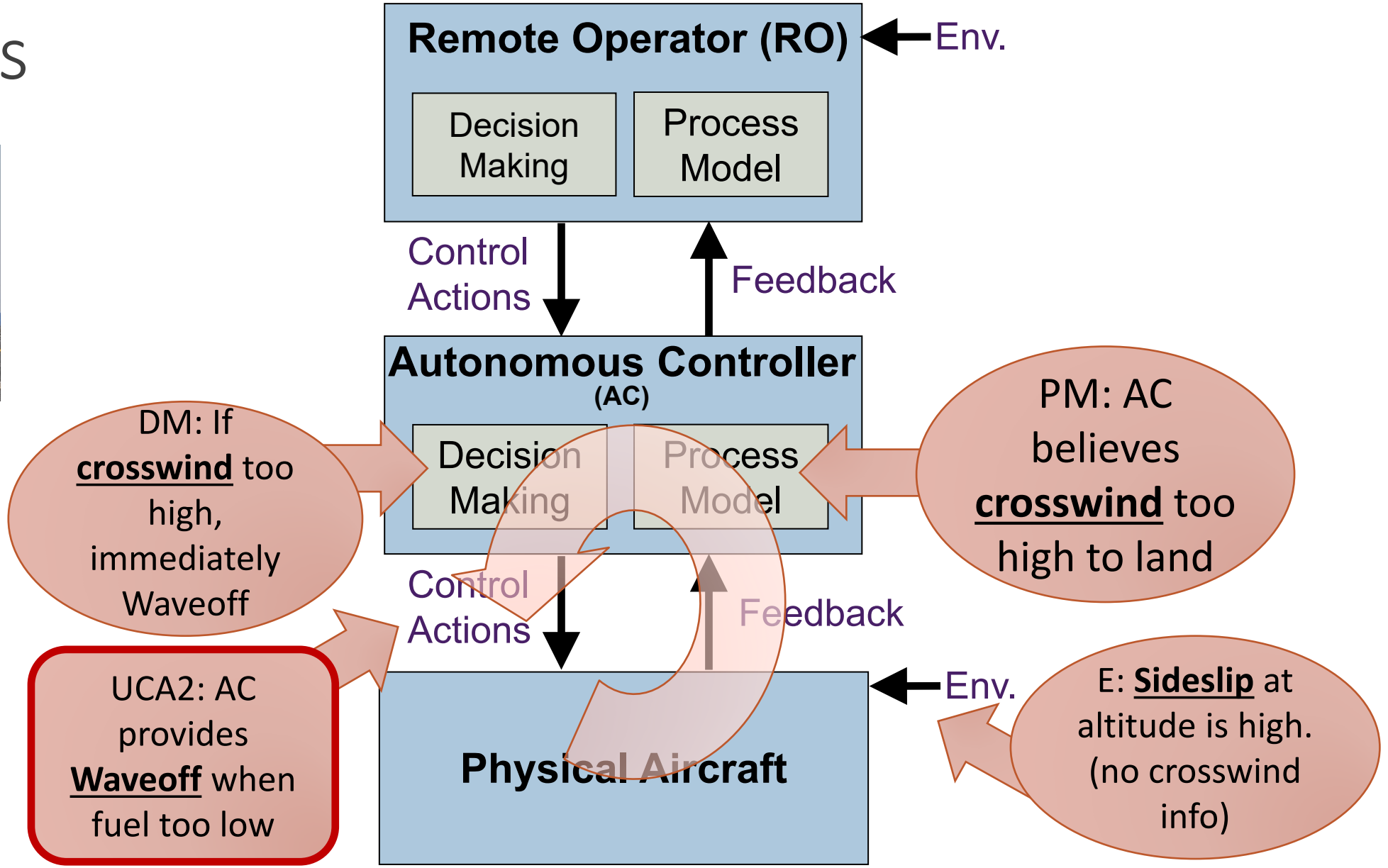
E.g., X-47B UAV integrated into carrier operations alongside manned aircraft. Provides autonomous launch, flight, follow manned A/C, carrier landing, etc.



(John Thomas, 2021)

STPA Step 3: Identify UCAs

E.g., X-47B UAV integrated into carrier operations alongside manned aircraft. Provides autonomous launch, flight, follow manned A/C, carrier landing, etc.

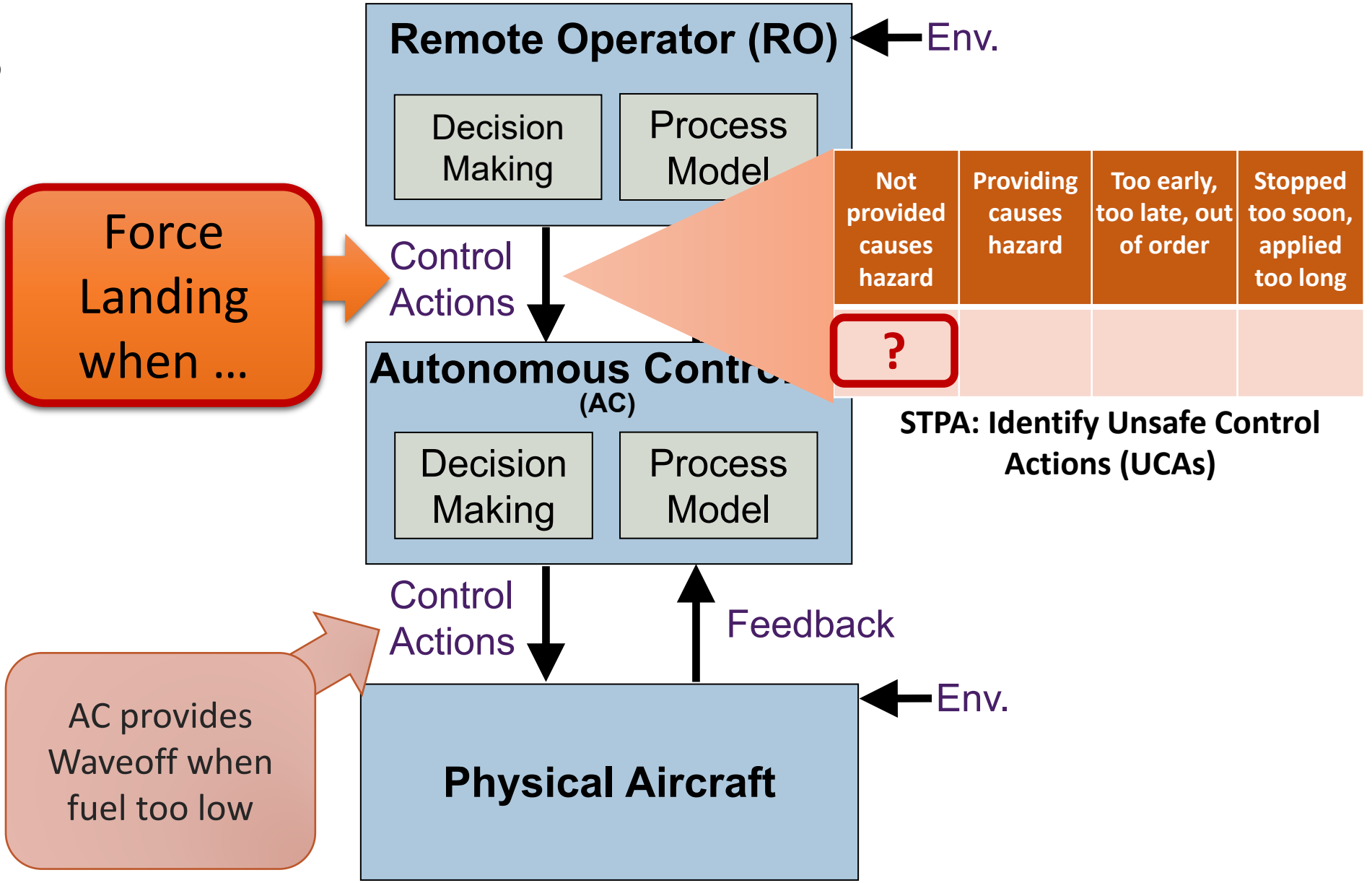


(John Thomas, 2021)

STPA applied to humans (FTEs, Pilots, etc.)

STPA Step 3: Identify UCAs

E.g., X-47B UAV integrated into carrier operations alongside manned aircraft. Provides autonomous launch, flight, follow manned A/C, carrier landing, etc.

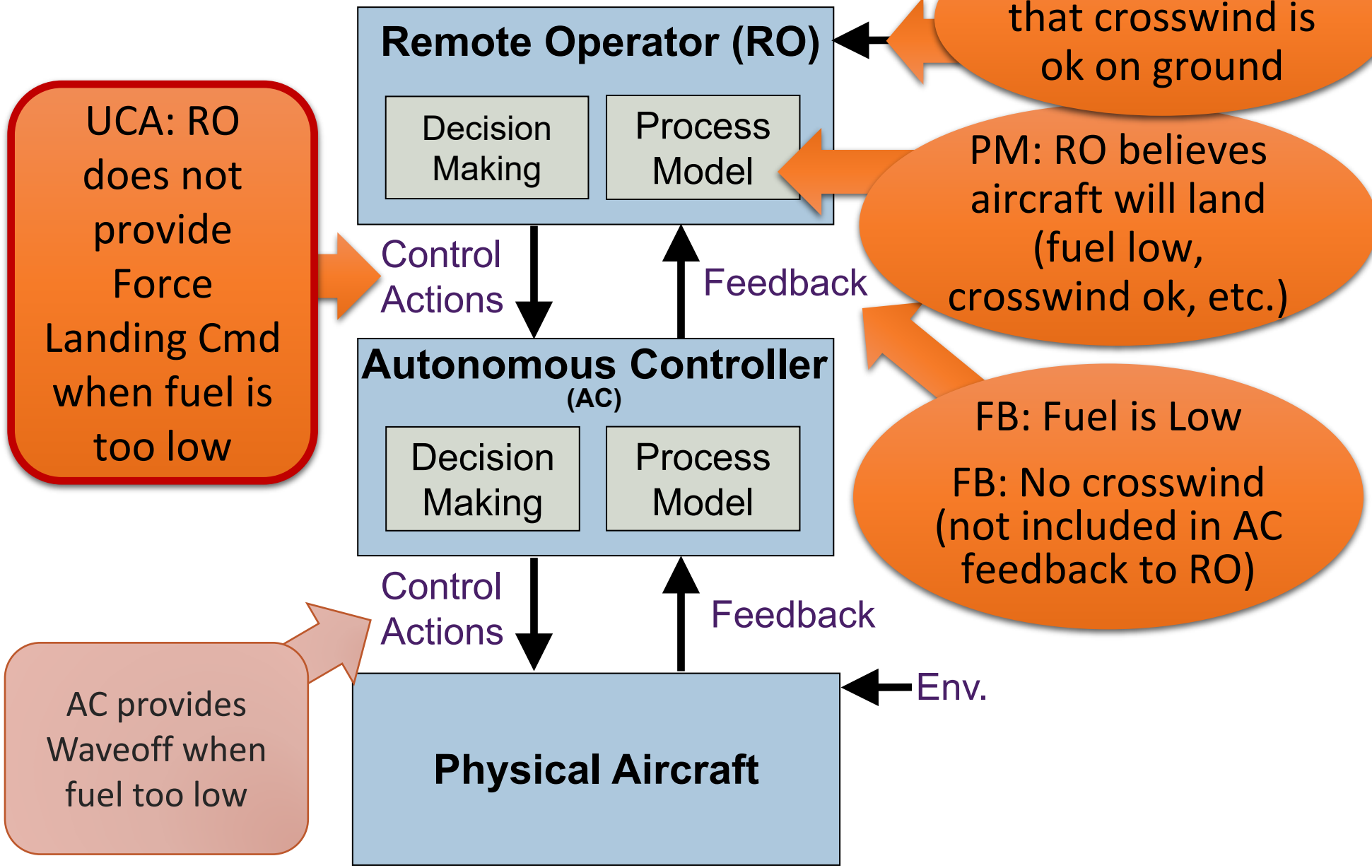


STPA: Identify Unsafe Control Actions (UCAs)

(John Thomas, 2021)

STPA Step 3: Identify UCAs

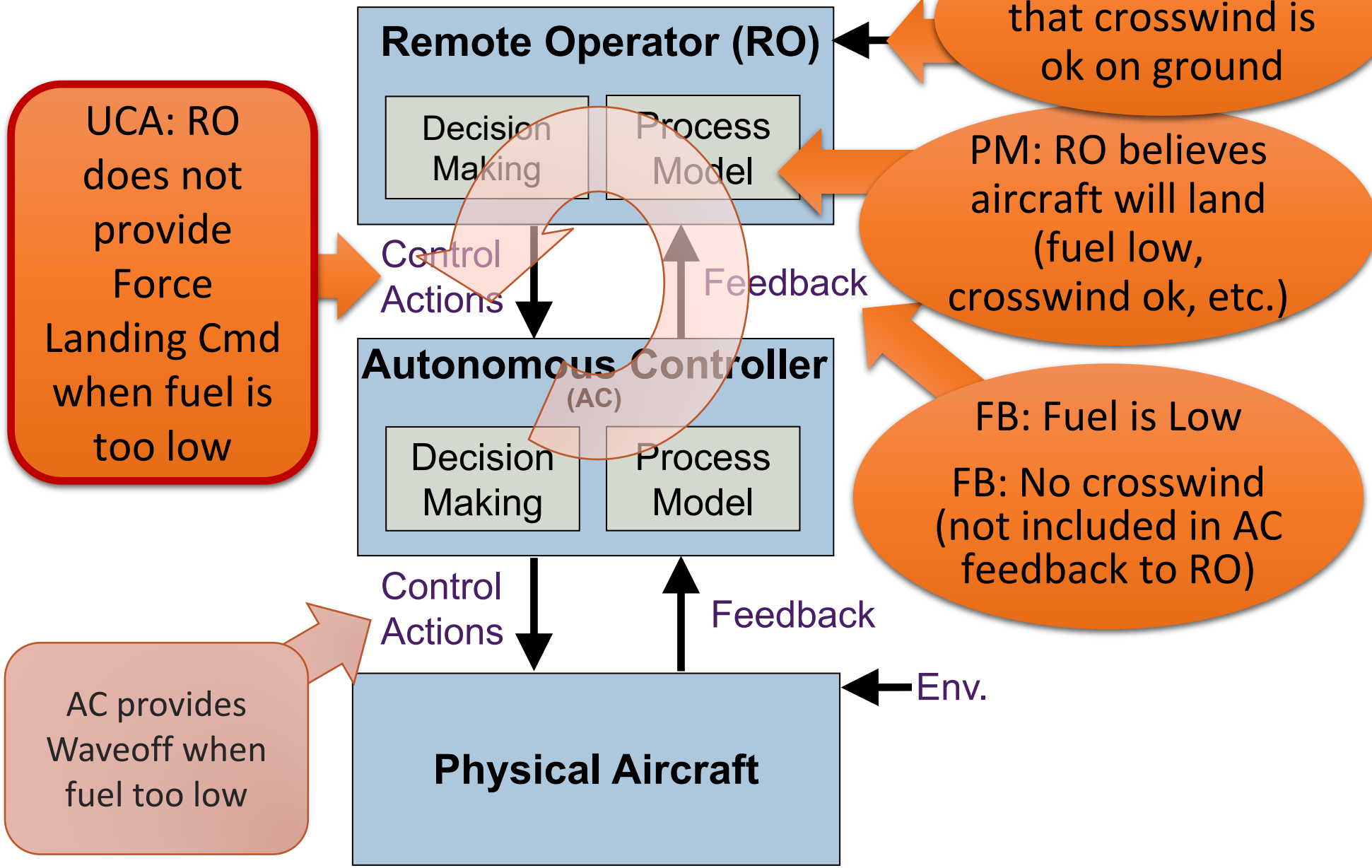
E.g., X-47B UAV integrated into carrier operations alongside manned aircraft. Provides autonomous launch, flight, follow manned A/C, carrier landing, etc.



(John Thomas, 2021)

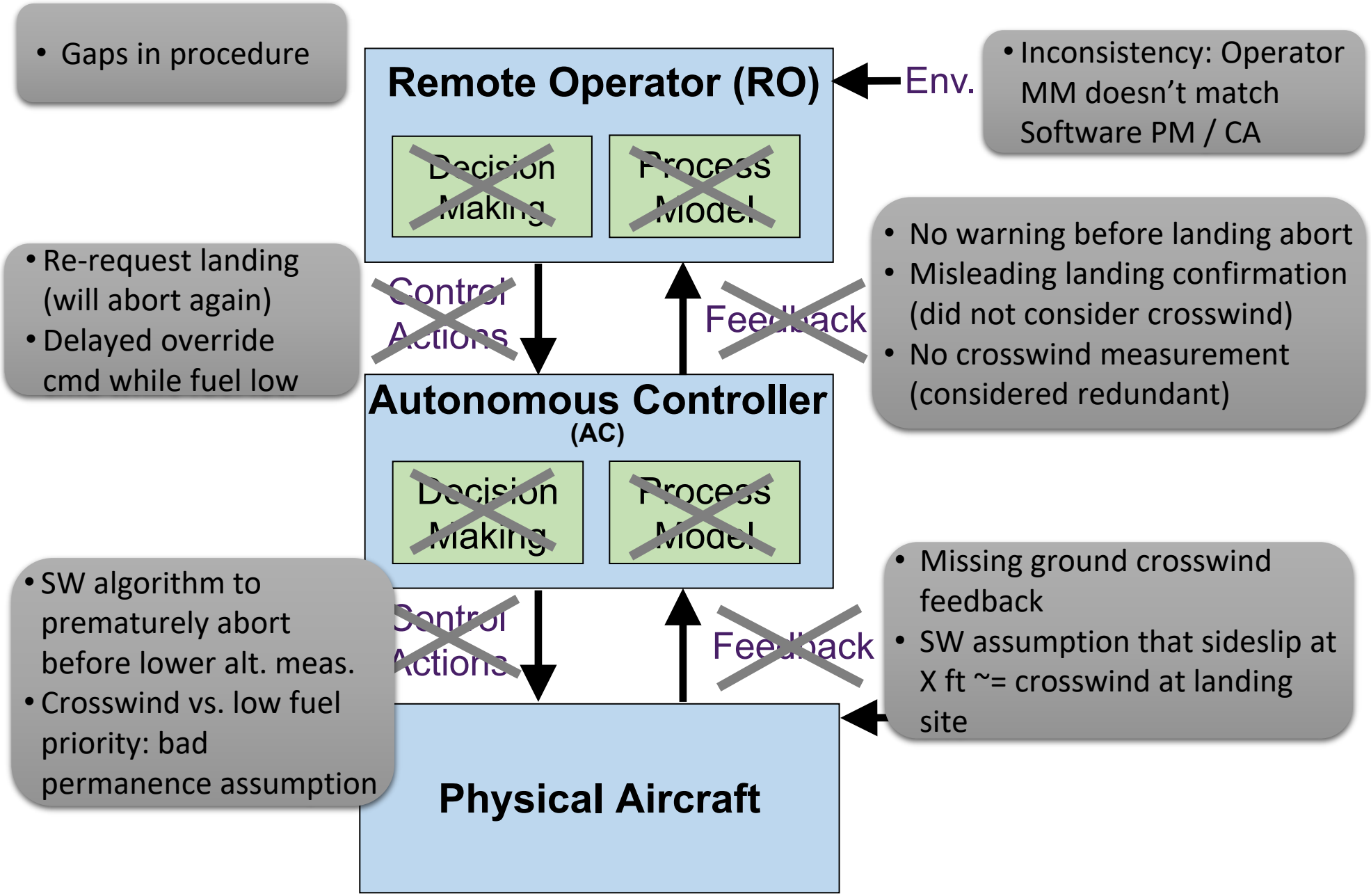
STPA Step 3: Identify UCAs

E.g., X-47B UAV integrated into carrier operations alongside manned aircraft. Provides autonomous launch, flight, follow manned A/C, carrier landing, etc.



(John Thomas, 2021)

Results



(John Thomas, 2021)

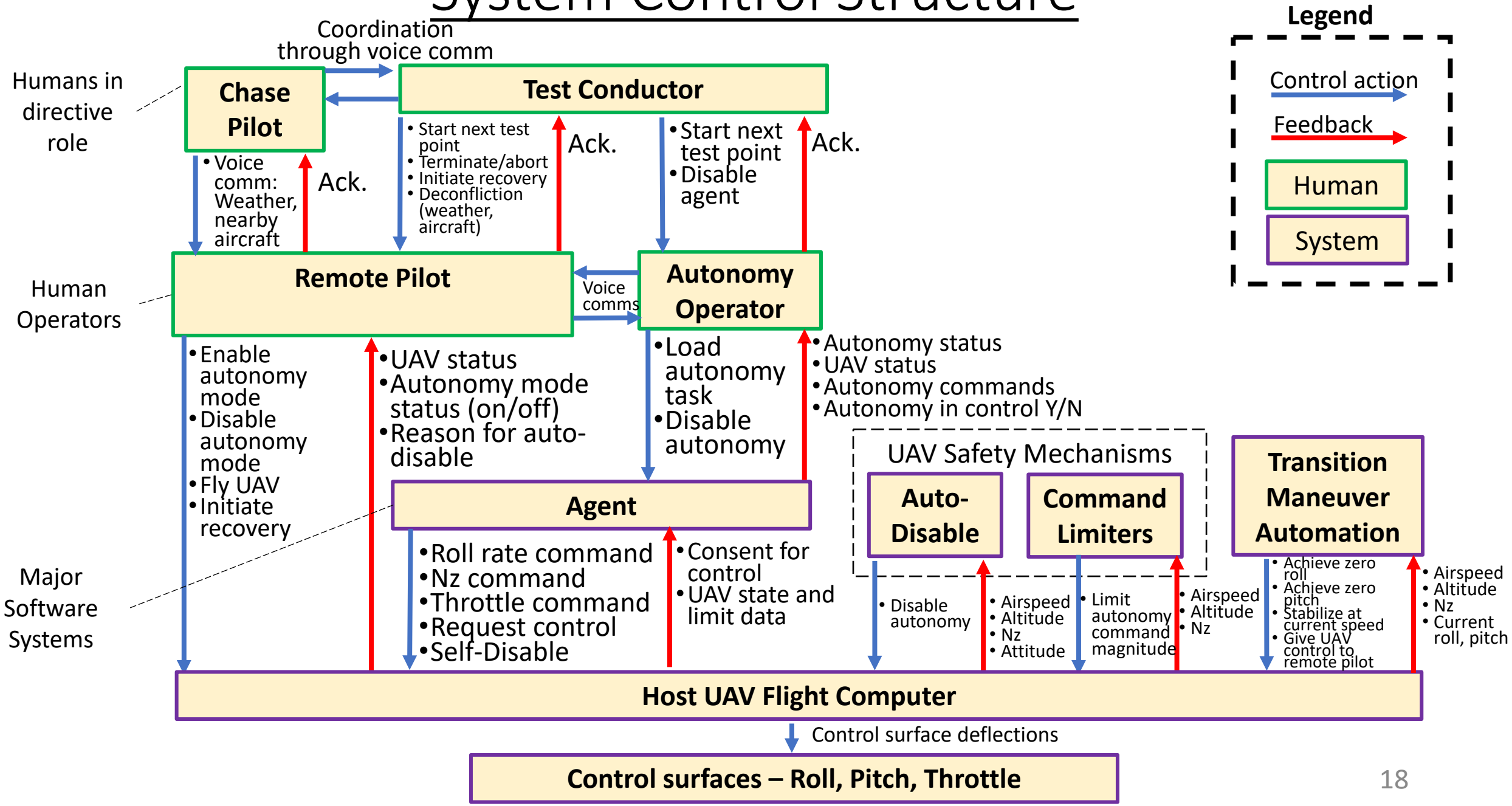
STPA Application

Session 1: 4-day training + application (UAV only)
Session 2: 5-day training + application (UAV + AI)
Session 3: 2-day application (UAV + more AI detail)

Scope of analysis:

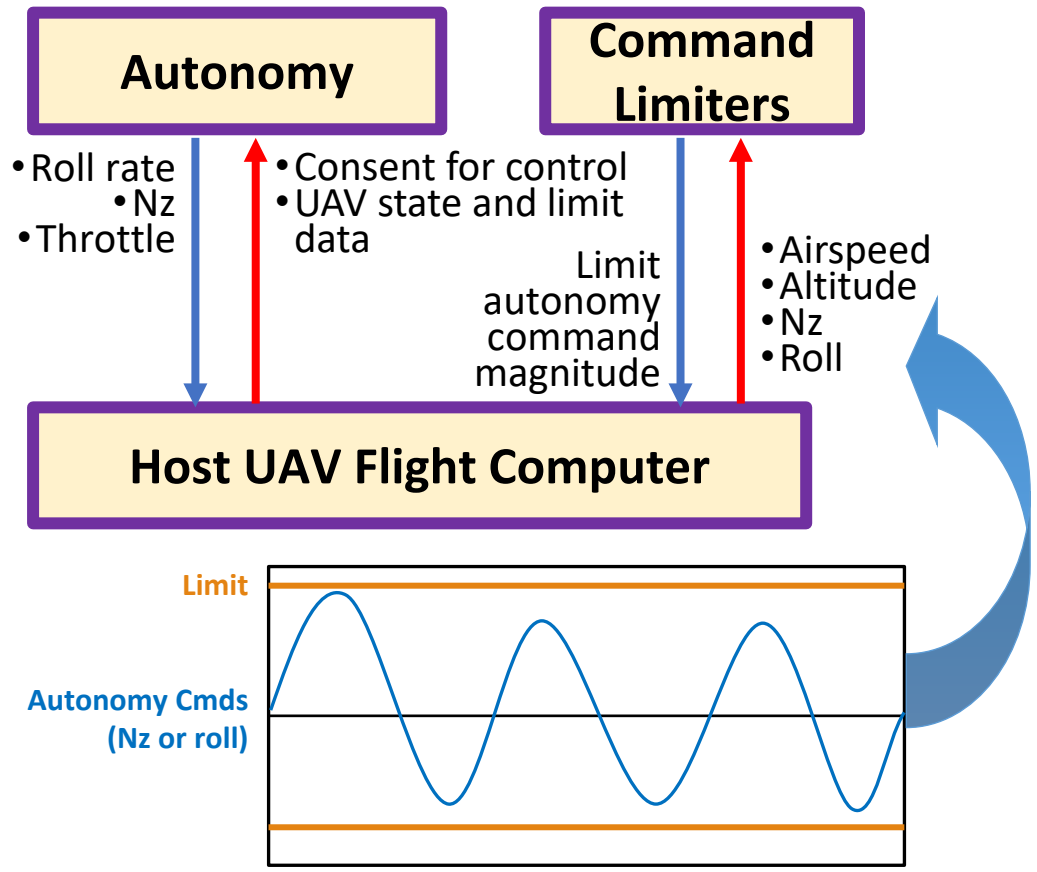
- Focus on flight test ops rather than internal system design
- Black-box AI – could do anything at any time
- Can the Autonomy Safety Sandbox handle all situations?

System Control Structure



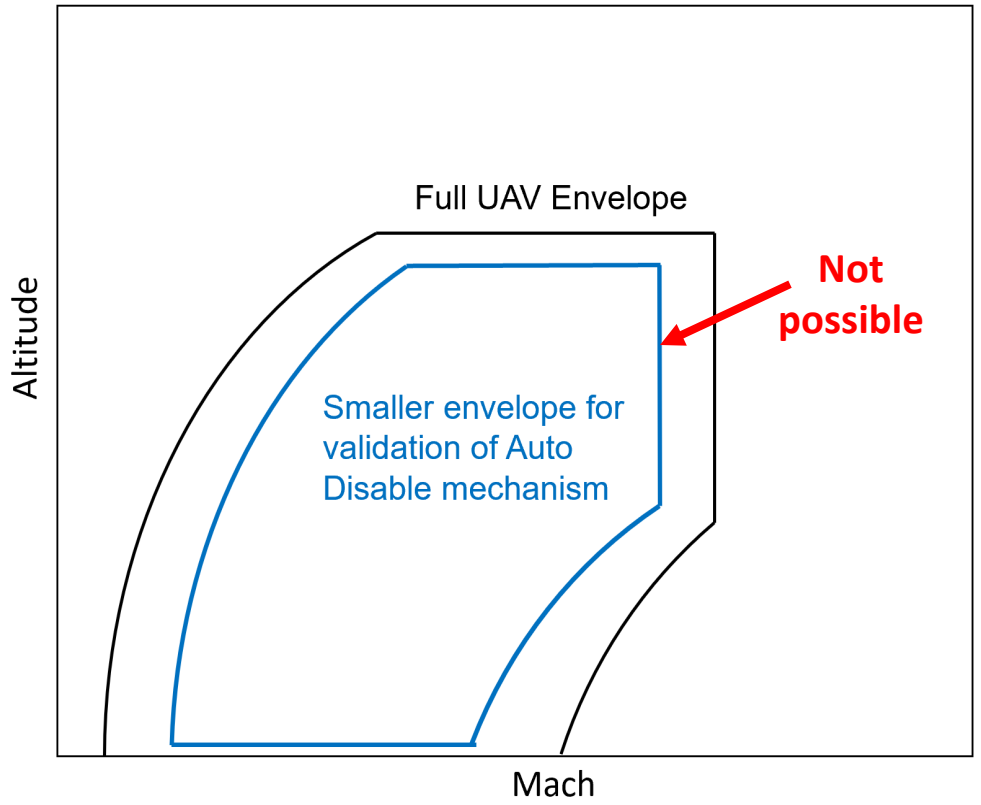
Finding 1: Limitations of Command Limiters

- Command limiters not complex enough to prevent some unsafe/inefficient commands
- No prevention of unsafe input **combinations**
- No awareness of time history – divergent **oscillatory** control inputs possible
- Recommendation: implement mechanism to prevent unsafe **maneuvers**



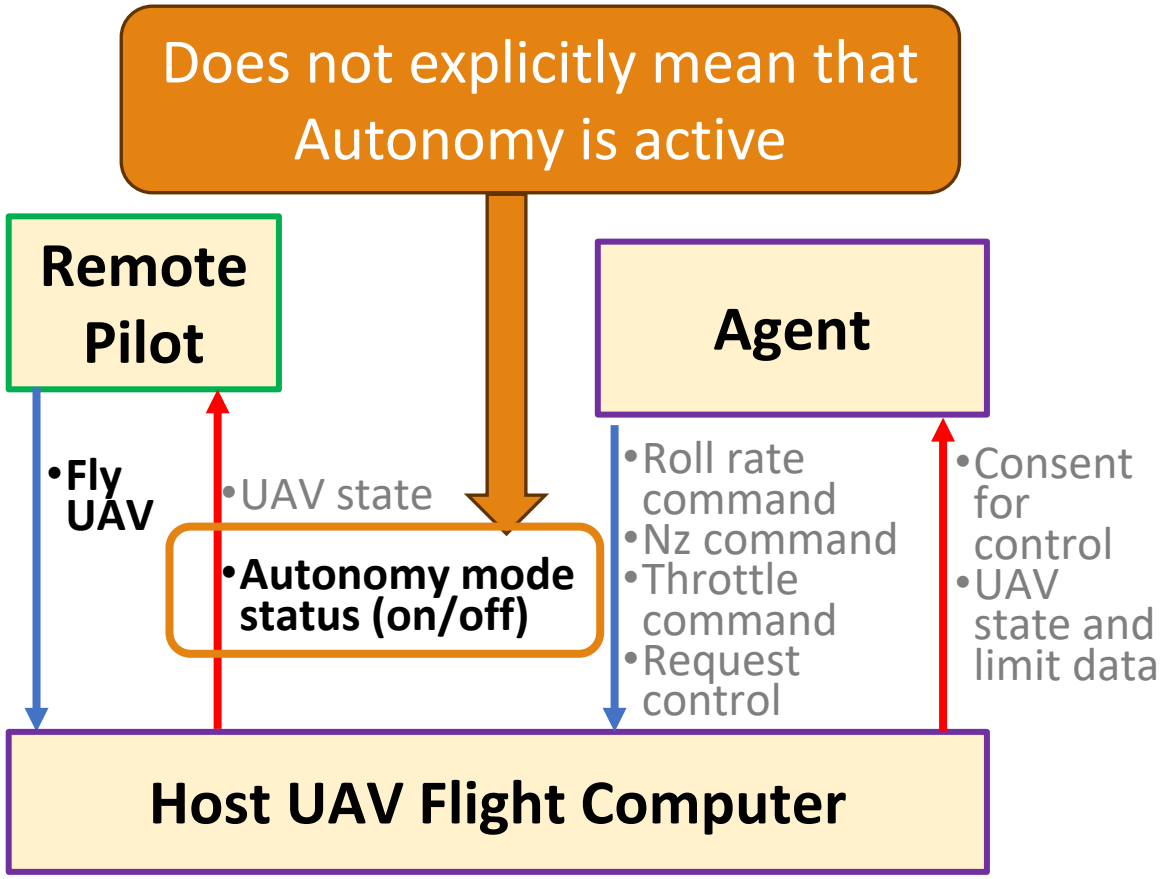
Finding 2: Inflexible UAV Auto-Disable Mechanism

- Auto-Disable altitude/airspeed bounds could not be easily modified
- Could not test Auto-Disable mechanism without assaulting the real limits
- Recommendations:
  - Make limit enforcement mechanisms flexible
  - Early tester involvement in system design



Finding 3: Incomplete Feedback from Autonomy to Remote Pilot

- Remote pilot had no direct indication of agent's status or actions
- "Autonomy mode" did not always mean the agent was in control.
- Recommendation: Provide unambiguous indication of agent status to the remote pilot.



Conclusions – Autonomy Safety Sandbox

- Three-pronged safety framework was effective but imperfect
- UAV safety mechanisms would not prevent all likely concerns
- Can mitigate those concerns by adding/modifying test procedures, but that tends to be heavy handed
- Some issues required band-aids because system design was fixed – recommend STPA during design

Conclusions – Use of STPA

- STPA was effective in identifying new test hazards and gaps
- Does not need to be the only method – use it as it makes sense
- Requires resources – time, personnel availability
  - Recommend 5+ days for detailed analysis
  - Invite the test team, operators, system SMEs
  - Bring in STPA experts if possible
  - In-person participation highly recommended

Questions?