VI.11

USING MODEL CHECKING TO HELP DISCOVER CONFUSIONS AND OTHER AUTOMATION SURPRISES

J M Rushby

Rapporteur: V Khomenko

Using Model Checking To Help Discover Mode Confusions And Other Automation Surprises

John Rushby
with
Judy Crow, Denis Javaux (Liège), Ev Palmer (NASA Ames)

Computer Science Laboratory
SRI International
Menlo Park, California, USA

Aviation Background

- Modern passenger aircraft are very reliable
- The dominant cause of incidents and accidents is human error (70% of accidents)
- Modern cockpits are highly automated
  - And highly complicated
  - Can sometimes override the pilot
- Pilots can be surprised by the behavior of the automation
  - Or confused about what "mode" it is in
  - "Why did it do that?"
  - "What is it doing now?"
  - "What will it do next?"
- Can formal methods help?

Postulates (from Human Factors)

- Operators use "mental models" to guide their interaction with automated systems
- Automation surprises arise when the operator's mental model does not acurately reflect the behavior of the actual system
- Mode confusion is a just a special case: the mental model is not an accurate reflection of the actual mode structure
  - Or loses sync with it
- Mental models can be explicitly formulated as state machines
  - And we can "capture" them through observation, interviews, and introspection
  - Or by studying training manuals (which are intended to induce specific models)

Facts (from Computer Science)

- The behavior of automated systems can be formulated in terms of (interacting) state machines
- These state machine descriptions are increasingly being used to document requirements and designs (cf. Statemate, UML)
- A technology called "model checking" can be used to examine the complete behavior of very large state machines
  - Can examine many millions of states
  - Used routinely in h/w design, s/w requirements analysis
  - It is largely automatic
- Can check whether certain properties are always true (e.g., every operator input is eventually acknowledged)
- Or can compare whether two state machines are "consistent"
- Produces counterexample when divergence found

Putting These Together

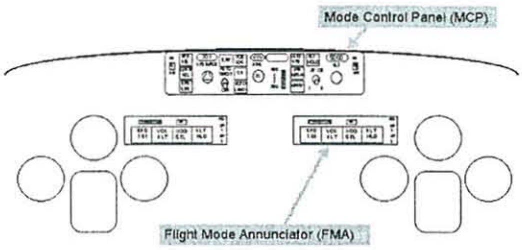
- Take the design of an automated system
  - Represented as a state machine
- And that of a (plausible or actual) mental model
  - Also represented as a state machine

  And check them for consistency
- Any counterexamples will be potential automation surprises

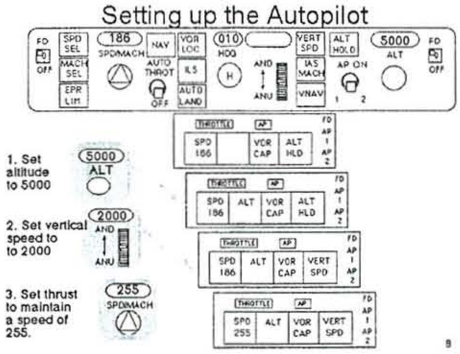
Example: Altitude Bust Scenario

- Scenario describes an automation surprise in the MD-88 autopilot (from Ev Palmer)
- Crew had just made a missed approach
- Climbed and leveled at 2,100 feet

Cockpit Interface: Realistic Picture

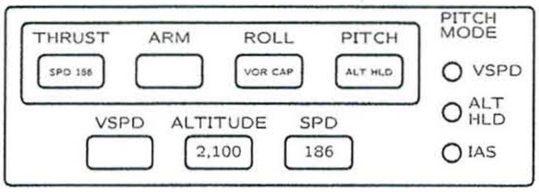


Setting up the Autopilot



Altitude Bust Scenario: Abstracted Picture

- Scenario describes an automation surprise in the MD-88 autopilot
- Crew had just made a missed approach
- Climbed and leveled at 2,100 feet

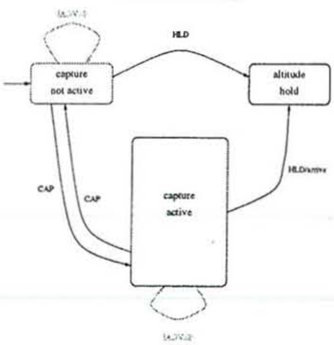


Color code: done by pilot, done by others or by automation

Altitude Bust Scenario: Mental Model

- The pitch modes determine how the plane climbs
  - VSPD: climb at so many feet per minute
  - IAS: climb while maintaining set airspeed
  - ALT HLD: hold current altitude
- The altitude capture mode determines whether there is a limit to the climb
  - If altitude capture is armed
    - Plane will climb to set altitude and hold it
    - There is also an ALT CAP pitch mode that is used to end the climb smoothly
  - Otherwise
    - Plane will keep climbing until pilot stops it

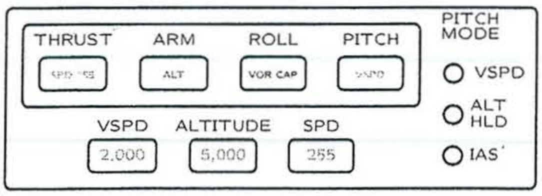
Mental Model



Whether capture is active is independent of the pitch mode

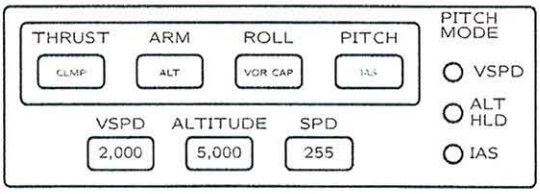
Altitude Bust Scenario—II

- Air traffic Control: "Climb and maintain 5,000 feet"
- Captain set MCP altitude window to 5,000 feet
  - Causes ALT capture to arm
- Also set pitch mode to VSPD with a value of 2,000 fpm
- And autothrottle (thrust) to SPD mode at 255 knots



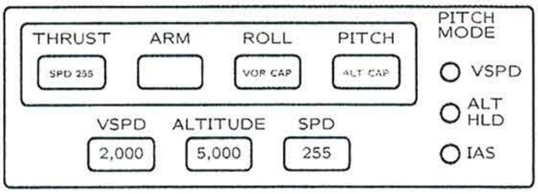
Altitude Bust Scenario—III

- Climbing through 3,500 feet, flaps up, slats retract
- Captain changed pitch mode to IAS
  - Causes autothrottle (thrust) to go to CLMP



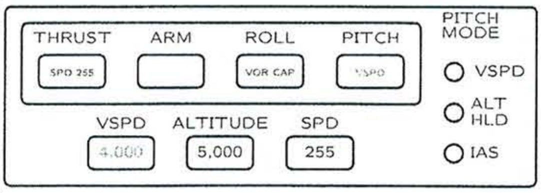
Altitude Bust Scenario—IV

- Three seconds later, nearing 5,000 feet, autopilot automatically changed pitch mode to ALT CAP
  - And disarmed ALT capture



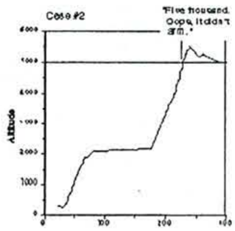
Altitude Bust Scenario—V

- 1/10 second later, Captain changed VSPD dial to 4,000 fpm



Altitude Bust: Outcome

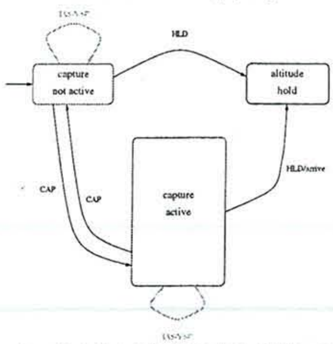
- Plane passed through 5,000 feet at vertical velocity of 4,000 fpm
- "Oops: It didn't arm"
- Captain took manual control, halted climb at 5,500 with the *"altitude—altitude"* voice warning sounding repeatedly



Automated Discovery of the Altitude Bust Scenario

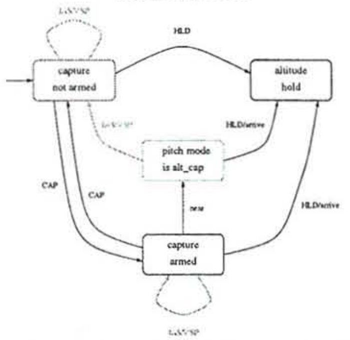
- I did it using a model checker called Murϕ
  - Comes from David Dill's group at Stanford
- But first I'll explain it using diagrams

Mental Model (again)



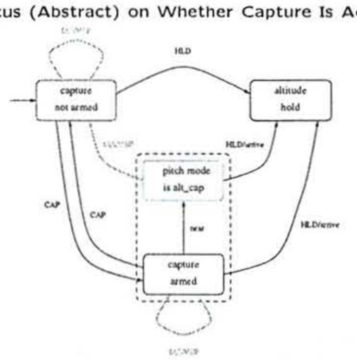
Whether capture is active is independent of the pitch mode

Actual System



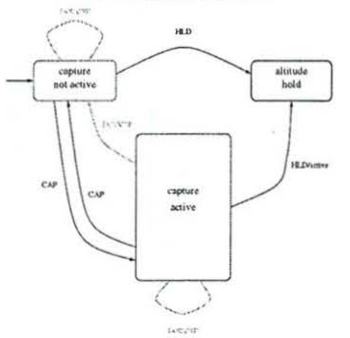
There is an alt_cap pitch mode that flies the final capture

Focus (Abstract) on Whether Capture Is Active



Capture is active if it is armed or if pitch mode is alt_cap

Abstracted System



Can compare this description directly with the mental model

Altitude Bust: Murϕ Specification

```
type
  pitch_modes: enum{vert_speed, ias, alt_cap, alt_hold};
var
  pitch_mode: pitch_modes;
  capture_armed: boolean;
  ideal_capture: boolean;

startstate
begin
  clear pitch_mode;
  capture_armed := false;
  ideal_capture := false;
end;

rule "ALT CAPTURE"
begin
  capture_armed := !capture_armed;
  ideal_capture := !ideal_capture;
end;

rule "HLD"
begin
  pitch_mode := alt_hold;
end;

rule "IAS"
begin
  pitch_mode := ias;
end;

rule "VSPD"
begin
  pitch_mode := vert_speed;
end;

rule "near"
begin
  if capture_armed then
    capture_armed := false;
    pitch_mode := alt_cap;
  endif;
end;

rule "arrived"
begin
  if pitch_mode = alt_cap then
    pitch_mode := alt_hold;
  endif;
  if capture_armed then
    pitch_mode := alt_hold;
    capture_armed := false;
  endif;
  if ideal_capture then
    ideal_capture := false;
  endif;
end;

invariant ideal_capture = (capture_armed | pitch_mode = alt_cap)
```


Altitude Bust: Murϕ Analysis

```
Invariant "Invariant 0" failed.
Startstate Startstate 0 fired.
pitch_mode:vert_speed
capture_armed:false
ideal_capture:false
----------
Rule ALT CAPTURE fired.
capture_armed:true
ideal_capture:true
----------
Rule near fired.
pitch_mode:alt_cap
capture_armed:false
----------
Rule VSPD fired.
The last state of the trace (in full) is:
pitch_mode:vert_speed
capture_armed:false
ideal_capture:true
```


Altitude Bust: Results

- Found the "surprise" scenario (in 0.24 seconds)
- So did Leveson and Palmer
  - By looking for "indirect mode changes"
- They suggested a fix (see HESSD paper)
- I incorporated it in my model
- And found that it caused another surprise
- I fixed that
- And found yet another surprise
  (also present, in a different form, in original specification)
- I fixed that, and the system and the mental model now align

Altitude Bust: Additional Experiment

- Mode confusions can arise even with consistent models if operator loses sync
- I introduced a rule to model a forgetful operator (nondeterministically flips the mental state)
- Obviously this introduces mode confusions
- I then modified the mental model to "reload" its state from a display that indicates whether altitude capture is armed
- This works (no surprises), even with a forgetful operator
- Can be used to validate cues provided by displays

Observations

- Once the initial model was constructed, these experiments required negligible effort (and only seconds of machine time)
- Provides complete demonstration of consistent behavior
  - Relative to the models used
  - General experience with model checking is that you learn more by examining all possibilities of a simplified model than by probing some of the possibilities of the full thing (cf. simulation or testing)
- Approach does not supplant the contributions of those working in human factors and aviation psychology
  - Provides a tool to examine properties of their models using automated calculation

Comparisons

- Leveson enumerates error-prone design elements (e.g. indirect mode transitions)
  - And examines system design to locate them
    - ⋆ Must then determine whether those found are real problems in their specific context
  - Examination is not automated
  - Tension between examining too much and too little
- Butler (NASA Langley), Miller (Collins) and colleagues use mechanized formal methods (theorem proving and model checking) to examine specification of autopilot for safety invariants (e.g., no mode change without pilot input)
  - Similar to my approach
  - But mental model is richer specification than an invariant

Other Examples

- Have also used this approach to examine a surprise related to speed protection in A320
- And a known surprise in the pitch modes of the 737 autopilot
- Need to try it out on large, realistic examples
  - NASA Ames has done this with MD-11 FMS

Further Work (TBD)

- Denis Javaux (psychologist from University of Liège in Belgium) has proposed two processes that give mental models their "shape"
  - Forget rarely taken transitions
  - Forget preconditions

  Could take the model implied by training manual, then apply these two simplification processes, to generate plausible mental models "automatically"
- Could also take mental model from one airplane and compare it to the automation from another as a way of predicting training difficulties

Speculation

- Can also do design exploration on effects of
  - Simpler design
  - New operating instructions
  - Improved displays
  - Faulty operator
- The mental model could also be interpreted as a requirements specification
  - Describes desired rather than observed operator interface
- Lack of an accurate and simple mental model then suggests overly-complex design
  - How many states are needed?
  - Any complex data structures (e.g., a stack)?

  Minimal safe model assesses cognitive load

Technical Challenges: Methodological

Can only go so far modeling just the mode behavior
And abstracting everything else away

- Need to investigate incorporating limited models of the environment and of the control behavior
  - E.g., to distinguish climbing from descending, up from down
  - Qualitative physics may prove adequate
    - ⋆ Reasons about signs of quantities and rates of change
    - ⋆ E.g., climb means height increases (derivative is +)
    - ⋆ 737 example uses (some) of this
  - May need hybrid automata (and model checkers for these)
- Also need to look at real time issues (e.g., delay between reading display and taking action)

Deeper Models of Cognition

- Mental models deal with only part of the cognitive processes involved in operating a complex system
- People use different mental models for different situations, so may also need to examine issues like "how quickly can an operator load the right model?"
- Deeper models of cognition can allow some of this to be explored scientifically
  - E.g., ICS (interacting cognitive subsystems) from Cambridge
  - Being explored by Howard Bowman at Kent, David Duce at Oxford Brookes
- In general, modern models of human cognition are built on a computational interpretation, so combine well with formal computer science

To Learn More

- Our papers and technical reports are at `http://www.csl.sri.com/programs/formalmethods`
  - `http://www.csl.sri.com/~rushby/abstracts/hessd99` describes this work and provides the Murϕ code
    - ⋆ Links to Murϕ there also
  - `http://www.csl.sri.com/~rushby/abstracts/safecomp01`,
  - `http://www.csl.sri.com/~rushby/abstracts/dasc99`,
  - `http://www.csl.sri.com/~rushby/abstracts/hci-aero00`, and
  - `http://www.csl.sri.com/~rushby/abstracts/fm-elsewhere00` are other papers on this topic
- Information about our verification system, PVS, and the system itself are available from `http://pvs.csl.sri.com`
  - Runs under SunOS, Solaris, or RH (X86) Linux
  - Freely available under license to SRI

DISCUSSION

Rapporteur: V Khomenko

Lecture Two

Mr Newman mentioned that a pilot may treat a plane not as a single object (state machine) but as multiple objects. Dr Rushby replied that he modelled not a pilot, but rather the interaction between the pilot and the system. Mr Newman enquired if it is right to say that a pilot's mental model is in fact several mini-models. Dr Rushby agreed that typically for any system, people tend to have a number of mental mini-models, e.g. one for each mode of the system.

Professor Schneider mentioned that there are two schools of program specification: prescriptive, where one writes down axioms and they define a set of behaviours, and descriptive, where a system is described as a state machine. He stated that Dr Rushby was doing everything in the descriptive style and wondered if, from the psychologists' point of view, people in fact use both patterns for understanding things. Dr Rushby replied that there is a discussion among psychologists what a mental model is: is it a state machine, or is it goal-oriented? And there were several experiments conducted, e.g. one concerning the Mac interface, which appeared to be goal-oriented. In his opinion, people use both patterns.