System Crash, Plane Crash

Lessons from Commercial Aviation and Other Engineering Fields

Jon Kuroda, UC Berkeley EECS

<jkuroda@eecs.berkeley.edu>

You might know me from such talks as

- “Life in a Modular Datacenter”
  - **“Well, That Was a Bad Idea”**
- “Adventures in (Small) Datacenter Migration”
  - **“Let’s Not Do That Again”**
- “Catch Fire and Halt: Fire in the Datacenter”
  - **“OMG It Caught Fire!”**

I’m a Sysadmin, not a …

- I'm not pilot, but I do fly a bit as a passenger.
- I’m not a doctor, but I’ve volunteered as a medic.
- I identify as a Sysadmin, not a DevOp or SRE.

There’s something here for everyone.

Do you remember your first flight?

I don’t (6 months old), but I remember my next one at age 8.

Flew in a 747 much like this to Japan.

Back then, they gave kids little plastic model kits of the plane they were on.

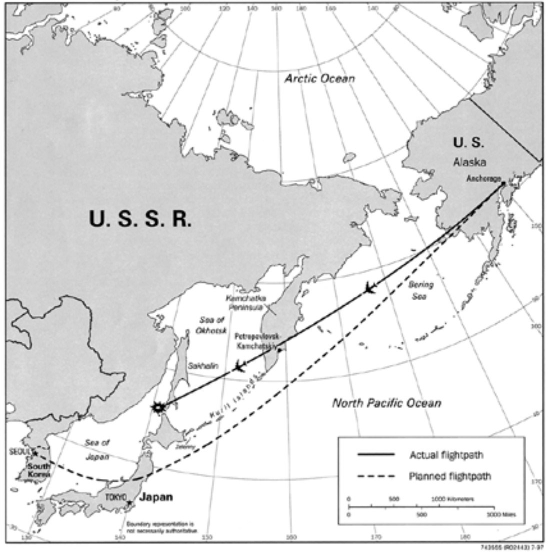
My favorite toy all summer long till it disappeared in September.

Korean Airlines 007 – 1 Sep 1983

Boeing 747-230B, JFK – Seoul via Anchorage

- Navigational system was in incorrect mode
- Flight gradually drifted off-course
- Entered then-Soviet airspace (twice)
- Shot down by Soviet ADF over Sea of Japan
- Major geopolitical impact
- Tons of conspiracy theories

Korean Airlines Flight 007, 1 September 1983



All 269 on board lost.

United Airlines 232 – 19 July 1989

DC-10, Denver to Chicago O'Hare

- Uncontained engine failure severed all 3 “redundant” hydraulic systems
  - Fandisk crack due to latent manufacturing fault went undetected despite use of indicating dye
- Total loss of use of flight control surfaces
- Only control was differential engine thrust - a Huge Hack(tm)
- Crash landed at Sioux Gateway Airport - broke up on impact
- Investigators later concluded "a safe landing was virtually impossible.”

185 of 296 on board survived the crash - over 60%

US Airways 1549 – 15 January 2009

Airbus A320, La Guardia to Seattle-Tacoma

- Bird strike by flock of Canada Geese during initial climbout
- Immediate loss of thrust in both engines
- Ditched in the Hudson River near the USS Intrepid Museum
- "[t]he most successful ditching in aviation history.”

All 155 aboard survived - "Miracle On The Hudson"

Air France 447 – 1 June 2009

Airbus A330, Rio de Janeiro to Paris

- Temporary inconsistency in indicated airspeed
  - Icing in airspeed sensors (pitot tubes) cited as cause
- Led to series of inappropriate flight control inputs by flight crew
- Caused apparently unnoticed and sustained, yet recoverable, stall
- 3min 30s descent while stalled from 38,000 ft to impact with ocean

Loss of all 228 on board. The deadliest crash in history of Air France.

A Tale of Two Outcomes

Two flights with severe mechanical distress

- Lots of people survived

Two flights with temporary and eminently recoverable failure

- Everybody died

What’s going on here?

Why the Contradiction?

- People? No.
- Decisions Made by People? Not quite.
- Decisions Made Given Training and Support? Getting There.

There is no single simple answer.

More often a constellation of factors that align one way or another.

(Oh, and sometimes Luck - good and bad.)

“… Pilot Error is not an answer; it’s only a symptom of some underlying problem.”

-Alan Diehl, Former NTSB Investigator - Human Performance

“Good design removes the defect. It doesn't put people in emergency situations and say, ‘If the pilots react perfectly, everything will be OK.’”

-Donald Nolan, Attorney in Aviation Safety,

regarding Boeing 737 Rudder Actuator Design Fault

KAL 007 – What Happened?

- Pilots didn't notice navigation system in wrong mode (TLDR: Bad UI)
- Poor US Civilian radar coverage beyond Alaskan coast
- Lack of USAF/Civilian radar coordination
- Inability to maintain VHF communication with ATC/FSS
  - Had to resort to HF (less clarity) or relay via other flights
- Weather inconsistent with reports from another flight on “same” route
- Soviet ADF predisposed to view anything as US aggression.
- Soviet ADF early-warning radar site non-operational
  - Could have had 2 more hours to properly ID KAL 007 as civilian

UI. Complacency. Monitoring. Confirmation Bias. Incident Response.

UAL 232 – What Happened?

- Redundant: That word. I do not think it means what you think it means
- Impossible: That word. I do not think it means what you think it means
- Quick Diagnosis ... but no Checklist or Manual for this …
- Presence of UAL Flight Instructor who had studied JAL 123
- Ad Hoc Solutions: Differential Engine Throttle Control
- Crew Resource Management: Flightdeck crew was a textbook example
- Iowa ANG on duty at the airport.
- Shift Change at nearby Trauma and Burn centers

Fault (In)tolerance, Monitoring, TeamOps, and Luck

US Air 1549 – What Happened

- Timely, accurate evaluation of impact of bird strike - No Engines
- Immediate activation of APU and RAT (Wind-driven electrical generator)
- Evaluated options based on overall "Expectation Values"
- A320 rated for over-water flight - lifevests and extra rafts
- Crew Resource Management - Excellent Team Work
- A320 Fly-By-Wire flight control system
- First time flying together for Sully and Skiles
- First time on an A320 after IOE for Skiles

Correlated Failure, IR, TeamOps, UI, and Luck

AF 447 – What Happened

- Airspeed Indicator (ASI) Sensor Flaw - Prone to Transient Icing
- Lack of Training in forseeable conditions: loss of ASI, Hi-Altitude manual flight
  - If no airspeed sensor data, possible to stall an Airbus w/ no stall warnings
- Pilot flying induced a stall
  - fell into "closest fit pattern" - Take-off Go-Around (TOGA)
- Attempts to correct led to restoration of anti-stall system: alarms
- Flight Systems: no ASI failure alarms, AoA, other critical flight info
- Crew Failure to recognize stall or even possibility of stall
- Input-averaging hid conflicting flight inputs from both Co-Pilots
- Captain not in control of plane or cockpit during event (was on rest break)

UI, Training, Human Performance Factors, Alarm Saturation, TeamOps

Starting to Sound Like Our World

- UI/HCI/HSI: That button/cmd. It did not do what I thought it did.
- Training: I may never truly understand Paxos or RAFT
- TeamOps: MeatOps, DevOps, CRM: People working with People.
- Correlated Failures: Entire batch of HDDs dying all at once.
- Monitoring: If a log entry is made, but nobody sees it, did it matter?
- Fault Tolerance: 2 name servers, in same room, on same circuit ...

Other General Similarities

Along with Nuclear Power, EMS, and Spaceflight

- Multi-Billion Dollar Industries Borne out of WWII development
- Human interactions with complex human-designed systems
- Complex Human-Human Interactions
- Customer Support Challenges
- Varying Operator Understanding of Underlying Systems
- Sometimes contentious relationships with vendors
- Significant ethnic/gender imbalances

But There Are Differences

Commercial Aviation Has

- Externally Regulation with Strong Union Presence
- Physical Constraints on Redundancy and On-call support
- Staff/Equipment considered more or less interchangeable
- Customer-facing Systems and Operators moving in space/time
- Stricter Time Demands - OODA/Situational Awareness
- Waterfall vs Agile approaches
- Limited Operator Ability to Investigate During Events or Undo
- Formal Ops Training
- Limited number of platforms to support

But The Big “Difference” …

When AWS, Git*, Yelp, or Slack go down,

¯_(ツ)_/¯

When planes crash,

People die.

Some Post-WWII History

Industries that saw massive WWII-era development

- Commercial Air Travel - Radar, ATC, Jets, Aluminum
- Nuclear Power - Fermi/U. Chicago, Manhattan Project
- EMS - Air Transport, Radio, Combat medics
- Space Flight - Project Paperclip
- Electronics Industry/Silicon Valley - Terman
- Computing/Sysadmin - Bletchley Park/WRNS

We’re Not That Young of a Field

Yet, maybe the least mature?

“The Wild West” / “YOLO”

– Jamesha Fisher on SecOps, LISA 2017

Is This a Fair Comparison?

Aviation does set some high standards

- At 100K flights per day, 5 “9”s reliability avgs to 1 crash per day ☹
- 0 passenger fatalities on US-flag airline scheduled flights since 2009
- That's a lot of proverbial 9s no matter how you run the numbers.

But is this the right way to think about things?

(And not just whether “9”s are actually a useful metric)

Maybe Comparison is the Wrong Mindset?

- True, we may not have to maintain life-safety standards[now].
- True, we don't have to maintain 8+ 9s reliability [right now]
- But is this a difference of kind or degree?

Maybe "Spectrum of Choice" is a better way way to think of it?

“You’re Doing It Wrong!”

I'm not here to simply say “We're doing it wrong!”

(We might be, but that’s not quite my point)

I'm here to:

- Find useful ideas from others that we don't know [enough] about yet.
- Formalize things others do that we may do or reinvent without knowing it

But why?

Why Indeed

Do a better job? Sorta. **Do a job more easily**? Sorta

Do a better job more easily? Getting there

How about:

Better tools and more informed perspectives in order to make better decisions about tradeoffs between budget, customer impact, time, staff impact, professional pride, and and and …

With Great Power …

Technology already enables one person's work to impact many.

Google. Facebook. Twitter. AWS.
HealthCare.gov. Apple Maps.
Equifax. Heartland. Target. Verizon.

We already have tremendous power to affect people's lives.
For better and for worse.

But "There's a storm coming, Mr. Wayne."

The buffer between us and Life-Safety Critical is shrinking.

"Machine Learning in Real Time Air Traffic Control Applications"

Scariest Research Talk Title I've Ever Seen.

- Self-Driving Cars
- Statistical Machine Learning / AI / Deep Neural Nets
- Google/Apple/Microsoft/Amazon Health?
- 911 handled by VOIP Systems.

Buffer between Us and Life-Safety Critical?

In some areas, it may be gone already.

In others, it may be there forever.

In others, it may be shrinking – quickly.

Do we even know?

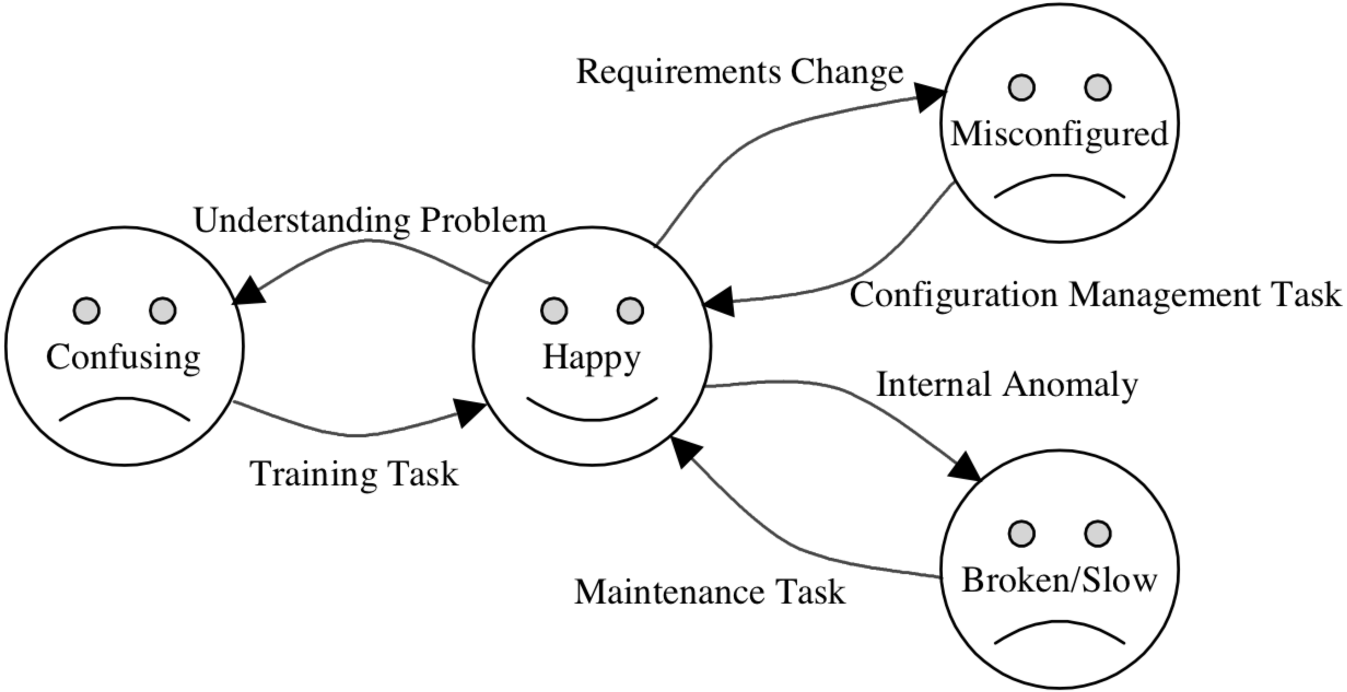
So, hey, let's go borrow some questions (and answers) from other people.

~~Stealing~~ Borrowing Ideas from Others

- Rote Repetition
- Understand the idea in the old environment
- Apply the idea in a new environment
- Understand the broader abstract principles
- Come up with a new(er) Idea based upon the borrowed idea.
- Evaluate, Share, Repeat

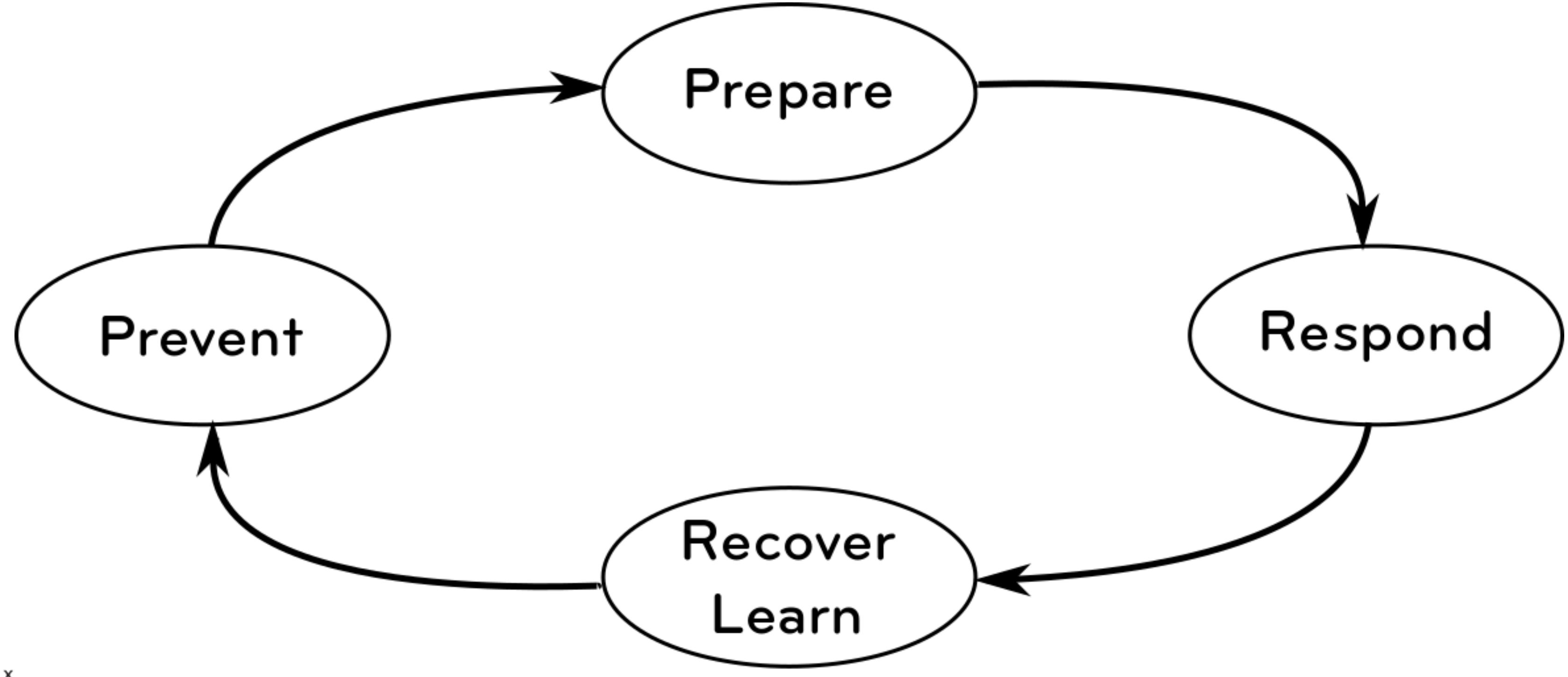
Basically, Bloom’s Cognitive Taxonomy

A 1999 View of Our Workcycle



Anderson, Patterson – LISA 1999

An Updated Cycle From Emergency Response



People and the Decisions They Make

Let’s go all the way back to High School Literature.

The 4 classic types of narrative conflict (Now 5, 6, 7 ...)

- Person vs Environment
- Person vs Other Person(s)
- Person vs Society/Culture
- Person vs Self

Not everything fits neatly into this, but it’s a helpful metaphor.

Person vs Environment

Nature is out to get us

- You can't beat physics (or chemistry)
  - Fuel Line Icing (BA 38)
  - Drive Failure due to Power Line Proximity (unnamed cloud provider)
- Nature is out to get us
  - Volcanic Plumes (BA 009, KLM 867, Eyjafjallajökull)
  - Canada Geese (US Air 1549)
  - Windshear / Microbursts / Clear Air Turbulence

Person vs Environment, Part Deux

- Poorly Understood or Unknown Phenomena
    - Aeroelastic Flutter (Tacoma Narrows)
    - Aluminum Metal Fatigue (BOAC DeHavilland Comets)
- Nature exceeds our attempt to withstand it.
    - Lightning and Composite materials - Bristow Flight 56C
    - One storm obscured a larger storm on radar - TACA 110

“Humans are intelligent” vs “Nature is persistent”

Person vs Society/Culture

Shared customs, history, language, art, technology, values, etc

- Societal/Ethnic/Regional/Academic/Familial Background
- Professional Culture - Training Training Training
- Corporate Culture - Customer Customer Customer or $$ $$ $$
- Group Think
- Personal Culture

Person - System

Person vs Expression of Designer/Engineer culture

- HCI, UI, UX
- Ergonomics - No Average Person
- Automation: TCAS/GPWS, Auto Pilot, Auto Throttle.
- Insufficient Testing, Design Flaws

Person vs Others

Interpersonal Interactions

- Fellow Team Members, Other Teams, Other Orgs
- ATC / Ground Crew / Gate Agents / Customers
- Company/Vendor Reps
- Communication skills
- Documentation and Training is Communication (and Culture).

Person vs Self

Human Performance Factors or "We are our own worst enemy"

- Stress Management / Furious Pattern Matchers / OODA
- Physical Incapacitation / Fatigue
- Distraction / Lack of Focus / Complacency
- Loss of Situational Awareness
- Poorly maintained skills
- Cognitive Biases

"Humans Are Terrible"

-Tanya Reilly, LISA 2017

Are we doing any of this already?

Post Mortem/Event Retrospective/RCA/etc

What, How, Why, Future Prevention/Response

- Comes from Medicine, Flight Operations, Military AAR, NRC
- “Blameless” “Transparent” – objective, factual, actionable
- More like AWS S3 and GitLab outage reports
- Less like UC Berkeley 9/15 Fire Talk from LISA16 [**REDACTED**]
- Not like 2011 Berkeley E-Mail outage (What Post Mortem?)

We’ve improved at this, but room for improvement

What Kind of Improvement?

Organized collective reports

- NASA’s ASRS – (Anonymous) Aviation Safety Reporting System
  - One could say we use Reddit for this, but it’s not organized
- American Alpine Club’s Accidents in North American Climbing
  - Costs $$, but REI provides a “5 Takeaways” summary
- FAA’s Lessons Learned Aviation Safety Website
- NTSB’s own Investigations, Chemical Safety Board’s

What if I could find outage reports, big and small, in one place?
Would this require an external/3rd party organization?

Testing …

Does it do what we said it should (and only that)?

- Code Review
- Test Driven Development, CI/CD
- Code Coverage
- Writing All-The-Testing-Requirements remains hard

We're on the Right Path … challenges remain.

Speaking of Test Coverage …

Design / Testing (and Monitoring)

- “That's Impossible!” - Put in a test / monitor anyway.
- Test over entire envelope and then some.
  - TACA 110, BA 38
- “Passed Unit Tests - Failed Integration Tests”
- “Redundant” – Is it really? Against what?

Testing after Change

Failure when small changes cause outsize effects

- MGM Grand Fire, Las Vegas 1980
  - 24/7 Restaurant not required to have sprinklers by code
  - Restaurant was closed for renovation
  - Aluminum conduit came in contact with Copper Refrigerant pipe
  - Fire had time to develop unnoticed
  - 85 dead, 650 injured

Good testing protocol “should” mitigate this, right?

Checklists

Scripting Human Behavior

- Another idea from Medicine and Flight Operations.
- Human Automation – mitigate Human Fallibility.
- Consistent Use Reinforces Adherence to Procedure.
- Airline checklists even have checkpoints now (Save Points?).

I think we buy into the "Checklist Manifesto".

I think. Keep it up. Keep improving.

But, speaking of Humans (are Terrible)…

(and things we might not do or address yet)

Humans are Bad in Some Situations

Too Many Inputs

- Sensory Overload. Alarm Saturation. “Helmet Fire”
- Can be overcome to some extent with training.

Too Few Inputs

- Monotony. Repetition. Concentration w/o engagement
- Can be alleviated by better automation to an extent

Humans Are Also Bad at …

Cognitive Load under Stress

“Every healthcare provider is fallible. Under conditions of time pressure and stress, complex human responses are particularly prone to error. These include impaired perception of elapsed time and decreasing performance in situations associated with excessing cognitive load and rapid task switching.”

-Emergency Airway Management, 2015, Cambridge University Press.

Humans under Stress

- “Furious Pattern Matchers” by nature.
- Under stress, “Best Fit” against recent patterns.
- We can acquire new “Patterns”.
- But it takes time, practice, training.
- Small changes to patterns, sometimes much harder.

We’re highly evolved primates, but highly evolved for what?

Critical Operations

“Sterile Cockpit Rule” During Critical Operations

- Example: Early Phase of Critical Incident Response
- Example: Upgrade of major system (OMG, Jenkins)
- Essential staff only in the room (or channel)
- Minimize off-topic discussion till after critical period passes
- Incident Command System from EMS/Disaster Response

Cognitive Biases

Humans Get Smarter.

==

Humans Get Better at Self-Deception.

(There's a lot of these – Humans are creative.)

(And Terrible.)

Cognitive Biases: **Normalization of Deviance**

Eventual acceptance of deviance from organization’s accepted and documented standards due to absence of poor outcomes from previous repeated deviations from org’s accepted practices.

THIS IS FINE.
I'M OKAY WITH THE EVENTS THAT ARE UNFOLDING CURRENTLY.
THAT'S OKAY, THINGS ARE GOING TO BE OKAY.

Cognitive Biases

“Confirmation Bias” and “Look-Elsewhere”

Defend existing beliefs by mix of:

- Interpreting information favorably to one’s beliefs
- Straining to find explanation for contradictory data

“It is easy to turn a civilian type of plane into one for military use.”

– Major Genadi Osipovich, pilot who shot down KAL 007.

Cognitive Biases Writ Large

Near Misses

Repeat “Near Misses” can lead to false confidence in one’s ability to extract oneself from bad outcomes.

Challenger and Columbia Disasters

Engineering concerns downplayed by management due to lack of bad outcomes from previous deviations from standards.

Organizational Culture: Ethics

Ethics Fails

- Tay Bridge
- Kansas City Disaster

Ethics Non-Fails

- Citigroup Building in NYC
- University of California bulk traffic monitoring

Communication: Commercial Aviation

Birth of Crew Resource Management

- Air Incidents due to poor communication/task management
- Tenerife Airport Disaster, 1977; United Air 173, 1978
- 1979 NASA Workshop on “Flightdeck Resource Management”
- Resulted in CRM: Crew (then Cockpit) Resource Management
- Training required for all flightcrews flying in US/EU
- Has been applied to Maintenance, Fire Fighting, Healthcare

Communication: Commercial Aviation

Results of CRM

- Flightcrew trained on how to communicate with each other
- Less dependence on extensive interpersonal relationships
- Crew more able to provide and receive input on flightdeck
- More effective authority delegation and task management
- Dramatically Reduced Incident Rate

How Dramatically?

Operator Culture vs Designer/Engineer Culture

Korean Air suffered a streak of incidents through the 1990s.

KAL 8509 / KAL 801 - failure of communication in the cockpit

- Societal / Corporate reinforcement of Senior/Junior hierarchy
- Boeing cockpit designed around Flight Deck cooperation

Began CRM Training with consideration of Korean Culture.

No fatal crashes since 1999.

Lesson: Communication Training

Communication training is often ad hoc, reactive for us.

- Make it pro-active.
- Make it part of on-boarding.
- Make it part of the culture in the work environment.

Why?

Promote communication, clarify responsibilities, while respecting personal culture in the work environment.

Lesson: Match Training and Design

Example:

System depends on cooperation / coordination?

Then training can’t just assume it will happen.

It must actively promote it.

Lesson: Culture Matters

(Personal) Cultural Awareness

- People come from all sorts of backgrounds.
- Acknowledge this and get people onto the same page.
- Address this during system design.
- Address cultural differences in training.

Where we come from affects how we get somewhere together.

Human/UI interactions

KAL 007 / AF 447

- Poor indication Navigation System was still in HEADING mode
- Could have been set to INS mode but failed to “capture” track
- System continued to check-off navigational waypoints

Three Mile Island

- Monitoring system indicated a vital valve had been activated
- Reality: Indicated noted that **Power** to valve had been applied

Lesson for Us: UI Matters

UI on the Tools We Use

- AWS S3 Outage of 28 Feb – typo in command
- Removal of wrong device (on wrong host) from array
- Setting up HDFS by hand from scratch
- Debian Installer PreSeed File vs Kickstart ks.cfg
- UI may be good during normal operation, but during failure?

Lesson for Us: Telemetry

The Ship's or Captain's Log – A Record of What Occurred.

- Monitoring aka Telemetry
- Commercial flight has FDR, QAR, ATC radar recordings, ADS-B
  - Established and Standardized Formats – can basically replay events
- We have /var/log/*, Ganglia, ~~BRO~~ Corelight, etc etc
  - Some (/var/log) is free format. Others have their own format. Compatible?
- Can you reconstruct events given enough logfiles?

I think we can do better, but how?

Structured Distributed Tracing?

Lesson: Training.

OMG Checklists!

- But how often do you train/practice using those checklists?
- What are our "sandboxed" Flight Simulators?
- Check-captains? Periodic refresher training?

Pair Training

- 1st Officer often pilot-flying for take-off, landing, etc
- Captain provides feedback, shares experience – trains 1st Officer
- Ready to take over when needed – "My plane."

Lessons/Skills We Can Share Back

What lessons can we pass back?

DevOps, Agile, CI, Unit Testing?

Go see the talks by Jason Victor and Peter Lega.

No, really, just go see their talk from last year or Peter Lega's talk from this year. It's 10x better than anything I was going to say about this.

Open Questions

External Regulation. Licensing. More?

- Treat like Classified Work? Only when needed?
- Is this inevitable?

I have no easy answers for you.

Only (hopefully) better questions.

There's Hope

Yes, we have challenges ahead of us.

But, we have people we can ask for help.

We're not doomed.

(But I have *lots* of paperclips)

Thank you.

Questions?

Tomatoes?

KAL 007 UI

Inertial Navigation System (INS)

- Fancy Physics and Math™ - gyroscopes, sensors, and calculus
- Only Satellite, Landbased Beacons, Astro/Stellar more accurate
- Needs no external reference once calibrated/zeroed out.
- Can keep plane within a few miles at most of the flight path.

On 747 flown for KAL 007, the INS is engaged by “**arming**” it to “**capture**” the programmed flight path, but plane must be within 7.5 miles of that path and going in the proper direction to capture.

KAL 007 UI

- But what if the plane is not within 7.5 miles of that path?
- What if INS was engaged when plane was more than far off?
- Say due to magnetic heading (at high latitude) …

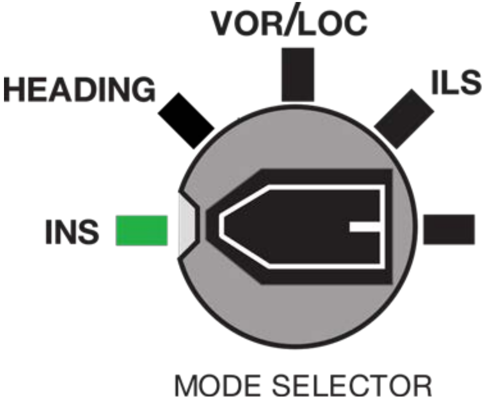
Then INS will never “capture” the flight path.

Plane will stay on that magnetic heading.

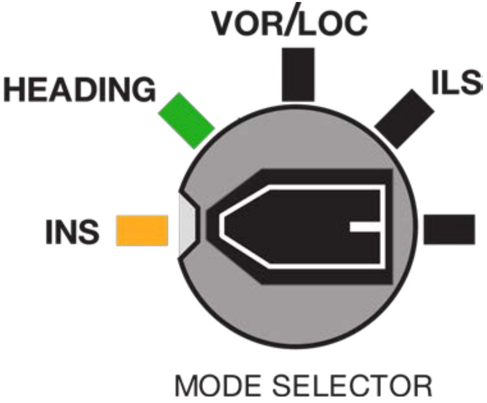
Unless the pilots realize INS is not properly engaged.

Would they?

KAL 007 UI

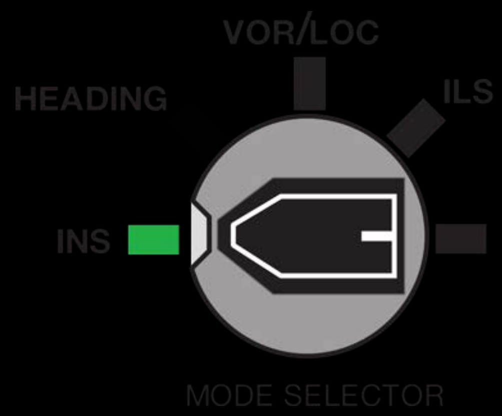


Track Captured

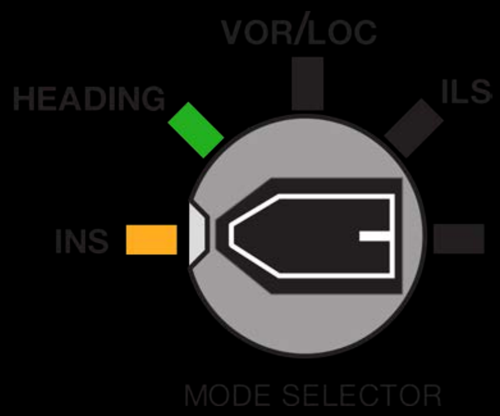


Armed to Capture Track

KAL 007 UI



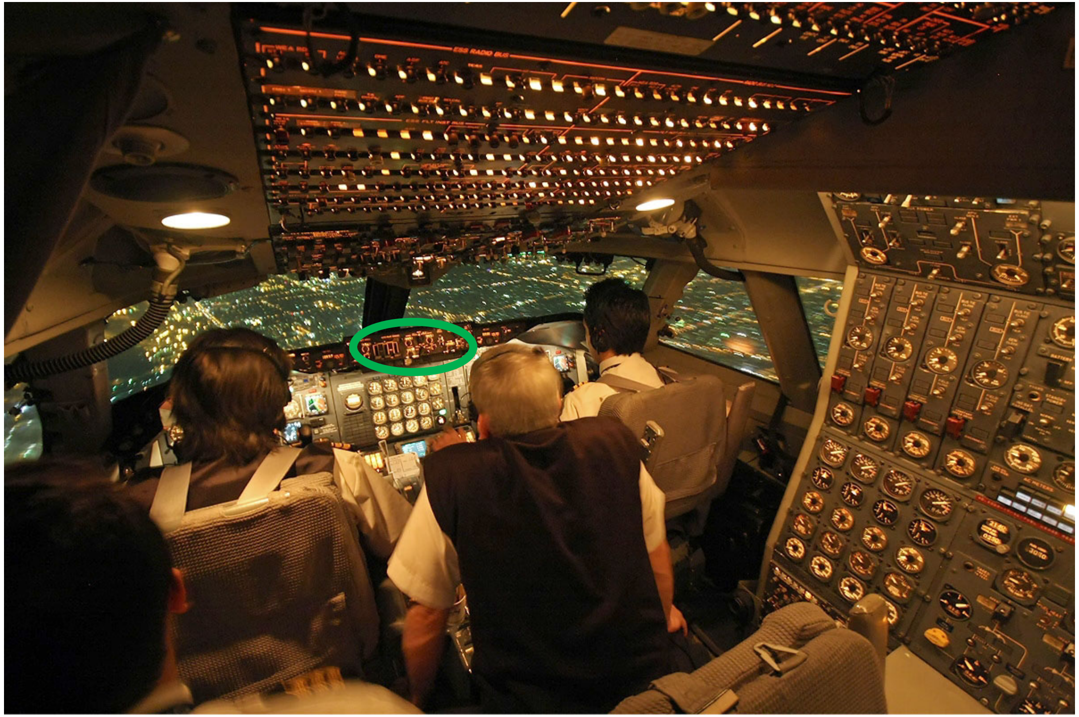
Track Captured



Armed to Capture Track

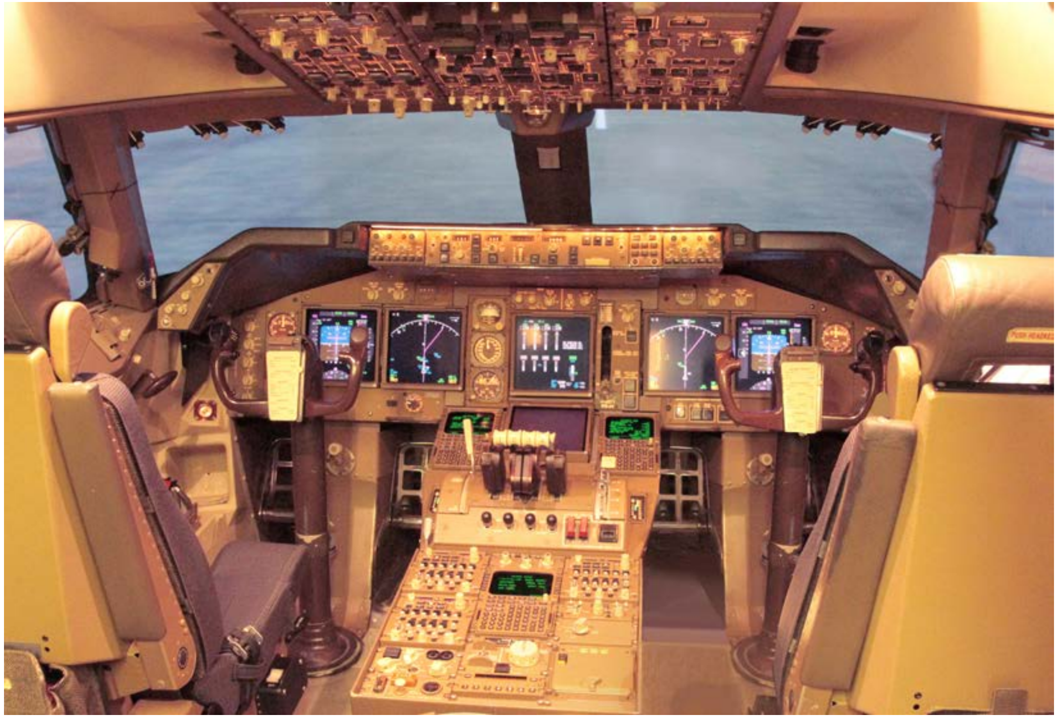
Kinda hard to tell, isn’t it?

747 Cockpit



This is a 747-200 cockpit of similar vintage to that flown for KAL 007 (Wikipedia says it's a 747-200 flown by Iran Air)

747 Cockpit - updated



This is a cockpit from a 747-400 flown by JAL.

Fewer dials.
More LCD panels.
Less clutter.