REPORT OF THE

# FLIGHT CREW HUMAN FACTORS INVESTIGATION

CONDUCTED FOR THE DUTCH SAFETY BOARD INTO THE ACCIDENT OF TK1951, BOEING 737-800 NEAR AMSTERDAM SCHIPHOL AIRPORT, FEBRUARY 25, 2009.

## LUND UNIVERSITY
### School of Aviation

SIDNEY DEKKER
FINAL REPORT JULY 2, 2009

# TABLE OF CONTENTS

# SUMMARY

On the morning of 25 February 2009, TK1951, a Boeing 737-800 was vectored to the localizer for an ILS approach to runway 18R at AMS at 2000 feet, less than 5,5 nm (nautical miles) from the threshold. This prompted the crew to use vertical speed mode to capture the glideslope from above (a necessity because of the close-in vector while being kept at 2000 feet). Air traffic controller workload had been mounting at the time and the approach sector was to be split shortly after TK1951.

The First Officer (F/O), a newly hired 42-year old pilot (with 4000 hours of Air Force flight experience) undergoing line training, was Pilot Flying (PF). The right autopilot (known as Autopilot B or CMD B) had been selected on, and the right Flight Control Computer (known as FCC B) was giving it all inputs.

Upon the crew's selection of vertical speed mode, and leaving 2000 feet, the 737's autothrottle (A/T) retarded to idle, which was consistent with crew expectations (and, for all they knew) their instructions to the automation. Approaching a new flap setting, the airplane had to slow down and go down simultaneously, something that required idle power at that point. For the next 70 seconds, the automation behaved exactly as the crew would have expected.

The autothrottle, however, had automatically and insidiously retarded in a mode (the so-called retard flare mode) that is not normal in this situation, but that was triggered by an erroneous radar altitude reading by the left Radar Altimeter (RA) and other flight parameters after leaving 2000 feet. There is no autothrottle indication in the cockpit that uniquely marks out the retard flare mode. The RA anomaly had not been reported to the crew, and there was no failure flag, no warning, no light nor any other direct annunciation about it in the cockpit.

Essentially, because of the erroneous radar altimeter input, the autothrottle decided that it was time to land. It no longer tracked the selected speed, nor did it provide so-called flight envelope protection. The autopilot, however, was still

4

flying the airplane, tracking the glideslope to the runway. In other words, the autothrottle was trying to land; the autopilot was trying to fly.

Earlier on, the crew had become aware of problems with the left radio altimeter during their descent towards Amsterdam, and they may have set up the cockpit so as to insulate the automation from those problems. They had Flight Control Computer B (FCC B) selected as the so-called Master FCC, and had selected Autopilot B on. FCC B has its own independent radar altimeter. The training and documentation available to B737 pilots suggests that this would be sufficient for protecting the automation against left radar altimeter anomalies.

But it is not. What is not in Boeing 737 documentation and training available to pilots is that the autothrottle always gets its height information from the left Radio Altimeter (RA), independent of which FCC has been selected as Master and independent of which Autopilot is selected on. The knowledge available through training and pilot documentation is so underspecified that it in fact can create a false or buggy mental model about the interrelationships between the various automated systems and their sensor input. This, on TK1951, produced what the literature has called an "automation surprise." The anatomy of this automation surprise is in the figure below.

### Anatomy of the TK1951 Automation Surprise

| What is trained: | What a crew could have concluded: | What is actually going on: |
|---|---|---|
| "One of the FCC's is specified as the master FCC." | We select FCC B as Master | FCC B controls autopilot B. Autopilot B is following the glideslope |
| "Each FCC continues to calculate thrust, pitch and roll commands" "The autothrottle adjusts the thrust levers with commands from the FCC." | FCC B calculates thrust commands; FCC B is giving the autothrottle its commands | |
| "Two independent radio altimeters provide radio altitude to the respective FCC's." | FCC B has its own independent radio altimeter. An inoperative RA on the left has no impact, as the left RA goes to FCC A | The autothrottle gets its height information directly from the left radio altimeter, independent of which FCC has been selected Master. |
| "FCC A controls autopilot A and FCC B controls autopilot B" | We fly this approach on autopilot B, which gets its input from FCC B. | Flying the approach on FCC B does not protect you from a left RA anomaly. |
| (source: Boeing 737 FCOM and Computer Based Training for pilots) | | (source: Boeing 737 technical documentation that pilots do not have access to) |

*Anatomy of the TK1951 automation surprise. Through training and documentation, the crew (as are all B737 crews) could have been led to believe that a problem with the left RA had no consequences as long as FCC B was*

*selected to control the automation that was flying the airplane. In fact, the left RA was still providing height data to the autothrottle, a fact that cannot be found in B737 pilot training or documentation.*

Because of the tight vector onto final approach, the crew was completing the landing checklist during the thirty seconds when the airspeed decayed below the selected landing speed (as the autothrottles kept the thrust levers in the idle position). On an approach such as this one, the autopilot will keep the nose where it needs to be to stay on glideslope, not where it should go to recover lost speed (because, at this phase of flight, maintaining airspeed is the job of the A/T, after all).

The TK1951 crew was surprised by the automation when it turned out that it had not been keeping the airspeed. When they did notice and tried to intervene, it was too late in this situation, for this crew, to recover.

The analysis conducted here:
- Shows the consequences of ATC not previously announcing or coordinating a short and high turn-in for final approach, so that the flight crew has no chance to properly prepare;
- Exposes design shortcomings in the Boeing 737NG AFDS/A/T systems that can lead to one part of the automation doing one thing (landing) based on corrupted input while the other is doing something else (flying);
- Highlights shortcomings in industry training standards for automated aircraft such as the Boeing 737NG. Such training does not support crews in developing an appropriate mental model of how the automation actually functions and what effect subtle failures have.

From the field work done for this analysis, as well as the DSB investigation into the TK1951 accident, both the length and depth of type training, as well as procedural compliance at THY, appear to at least match industry standard. The Captain (also instructor) on TK1951, moreover, had close to 11,000 hours on the Boeing 737. If such training, procedural standards and line experience are not enough to insulate a flight crew from an automation surprise such as the one that happened on TK1951, then few other airlines today can feel safe that their training and procedures protect their flight crews from a similar event. Post-accident manufacturer recommendations that, in effect, tell flight crews to mistrust the machine, stare harder at it, and intervene earlier, not only mismatch much human factors research, but also leave a single-failure pathway in place.

These are the sections that this report contains, presented here in summary:

## Some considerations for a human factors analysis

This section discusses possible pitfalls associated with a human factors analysis, among them the underdetermination of an explanation by the data we have available (that is, bridging the gap between what happened and why it happened can never be entirely free of contest), the problem of the hindsight bias and counterfactual reasoning, as well as distal versus proximal factors.

## A sequence of events

This section lays out a sequence of events as a basis for the subsequent analysis. The last minutes of the approach are chunked up into portions that hinge on critical events in the approach. Some clarifying remarks are added.

## TK1951 as an automation surprise

This section proposes how TK1951 fits the research on, and previous practical experiences of, automation surprises in modern airliners. The automation did initially what the crew (thought they) instructed it to do. When it finally became obvious that the automation had continued doing this for longer than was appropriate in that context, there was too little margin for this crew to recover.

## Automation training and buggy mental models

This section explores what basis a crew like that of TK1951 may have had in their training to cope with the automation surprise. Concepts such as buggy mental models will be used to trace how training, likely experience and existing B737 documentation leaves a crew fundamentally unprepared for the situation faced on TK1951. This is set against the background of training standards for automation and type-rating training for automated airliners such as the B737NG. A particular focus here is knowledge of FCC's, RA, and A/T, and the conclusion is unavoidable that there is no sufficient training, no sufficient documentation and no likely line experience that can help a B737 crew see through the sequence of events on TK1951

## Speed, mode and the noticing of a non-event

It is important to establish possible crew knowledge, because the mental model built up through training and experience is what people base their expectations, assessments and actions on. This section takes that as a starting point and examines the ability of monitoring humans to discover non-events in a cognitively noisy situation where that knowledge would have had to be brought to bear, and where initial system behavior matched crew expectations and instructions. This examination is set against the background of empirical research on mode monitoring and mode awareness, representational issues around speed tapes and the distributed cognition of speed monitoring and

awareness in a modern cockpit. It also takes into account the issue of interleaving tasks, workload and attentional issues during the execution of the approach of TK1951.

## CRM and the intervention decision

This section uses THY written guidance about cockpit roles, some principles of conversation analysis to the extent possible, and what little there is in terms of theory about intervention decisions in instructional situations and the juggling of flight safety and pedagogical goals, as well as some background on THY culture (to the extent it is available) to map out the coordination of crew interactions in the last ten or so minutes of the flight. It also discusses the eventual intervention (the take-over by the Captain) and the recovery attempt.

## Findings and Conclusions

The final section of this report presents the findings of the Flight Crew Human Factors Investigation into the TK1951 accident. Here is a summary of those findings:

1) During the descent towards Amsterdam, the TK1951 crew discussed and showed awareness of an anomaly of the (left) radio altimeter.
   TK1951 was kept at 2000 feet while being vectored to capture the localizer at less than 5,5nm from the 18R runway threshold at AMS. This put it above the glideslope. The crew appeared aware of the tight vectoring given to them by ATC (the landing gear was already down and the flaps were at 15 even before localizer intercept). No prior warning or coordination from ATC occurred, which would be normal and desirable with such a tight vectoring for approach.

2) Upon the crew's selection of vertical speed mode to capture the glideslope from above (as a result of the tight vectoring received from ATC), the 737's autothrottle (A/T) retarded to idle, consistent with crew expectations. The aircraft had to descend and simultaneously slow down to the next (flap 40) target speed. Upon selecting V/S mode, the A/T window of the Flight Mode Annunciator (FMA) on the Primary Flight Display (PFD) in the cockpit showed "RETARD."

3) The B737 has two RETARD modes that combine autothrottle and autopilot functions: (a) Retard flare and (b) retard descent. Retard descent commands the thrust levers to the aft stop to allow the autopilot to follow a planned descent path. Retard descent mode is normally followed by the ARM mode, in which the A/T protects the flight envelope and maintains a selected speed.

ARM mode also allows crews to manually set the thrust levers forward again.

In contrast, the Retard flare mode is normally activated just prior to touchdown when an automatic landing is performed. The A/T does the retard part, the autopilot the flare part, so as to jointly make a smooth landing. In retard flare mode, the A/T no longer offers flight envelope protection, does not maintain any selected speed, and it will keep the thrust levers at the idle stop (or pull them back there if the crew would push them forward).

The A/T window on the FMA offers no way for a flight crew to distinguish one RETARD mode from the other.

4) While the A/T window would normally have shown "MCP SPD" upon selecting V/S mode (knowledge that a pilot flying his 17th leg on a B737 is unlikely to have ready at hand), the RETARD mode made aircraft behavior insidiously consistent with crew expectations. They needed to go down and slow down so as to capture the glideslope (from above) and get the aircraft's speed down to the speed for the next flap setting. The A/T announced that it did what the crew commanded it to do: it retarded, and aircraft behavior matched crew expectations: the aircraft went down, slowed down and then captured and started tracking the glideslope.

5) As it only showed "RETARD" (and not "FLARE"), the FMA annunciation gave the appearance as if the A/T went into RETARD descent mode. However, the A/T went automatically into the unexpected RETARD flare mode—not because the crew had selected V/S, but because a number of conditions had now been fulfilled, and the A/T was acting according to its own logic: the aircraft was going below 2000 feet RA (in fact, it was at -7 feet RA, according to its only available (and corrupted) input to the A/T system), the flaps were more than 12,5 degrees out and the F/D mode was no longer in ALT HOLD.

While the A/T had, in effect, decided it was time to land, FCC B was still commanding the F/D and Autopilot B to stay on glideslope. One part of the automation was doing one thing (landing), while the other part was doing something else (flying). The part that was landing (the A/T) had control over the airspeed, the part that was flying (Autopilot B) did not; it only tracked the descent path on the glideslope.

6) Based on their training and documentation, the TK1951 crew would have believed that they had protected their aircraft and its flight from any pre-existing problems with the left RA. The right autopilot (known as Autopilot B or CMD B) had been selected on, and the right Flight Control Computer (known as FCC B) was giving it inputs.

Boeing pilot training materials and documentation do not reveal that the autothrottle always gets its height information from the left Radio Altimeter, that, on pre-2005 737NG models, it doesn't cross-check its RA input data with other RA data; and that the right RA does not provide input to the autothrottle—even when FCC B has been selected as Master and Autopilot B is flying (which was the case for TK1951).

7) The crew was completing their landing checklist during the thirty seconds when the airspeed decayed below the selected landing speed as a result of this automation mismatch. Interleaving task demands, speed tape design issues and the erosion of cognitive work surrounding the calculation of final approach speeds in automated airliners, becoming visual with the runway, and landing checklist design could all have interacted with the crew's attention during the 30 seconds of speed decay below approach speed.

8) There is no persuasive basis in the record to conclude that the approach was "rushed." The crew anticipated the late glideslope capture by lowering the gear and selecting flaps 15 even before capturing the localizer, and the only items to be completed after glideslope capture were final flap setting and the landing checklist. Landing clearance had already been obtained.

9) TK1951 fits human factors research on plan continuation. Decisions to go around or intervene in a student pilot's actions involve the assessment and re-assessment of the unfolding situation for its continued do-ability. The dynamic emergence of cues about the doability of the TK1951 approach, suggested to the crew that continuing was possible and not problematic.

10) A breakdown in CRM (Crew Resource Management) cannot be substantiated for TK1951. Other than artifacts of the instructional context of the flight, there is little to no evidence in the primary data source (the CVR) for overlapping talk, for second-pair part silences, or for other-initiated repair—three aspects of conversational interaction that have recently been implicated in CRM breakdowns. The Captain was well-liked, and a popular instructor at THY.

11) The length of B737 type training at THY, as well as procedural compliance at THY, appear to at least match industry standard. The Captain had close to 11,000 hours on the Boeing 737 alone. This combination of training standards and experience is apparently not enough to protect crews from the subtle effects of automation failures during automated, human-monitored flight. The documentation and training available for flight crews of the Boeing 737NG leaves important gaps in the mental model that a crew may

build up about which systems and sensor inputs are responsible for what during an automatically flown approach.

12) TK1951 fits the substantial research base on automation surprises. For 70 seconds, automation and aircraft behavior were consistent with crew expectations (the A/T had insidiously reverted to an unexpected mode that seemed—and was annunciated—as if it followed crew instructions). After that period, the really difficult task for the crew of TK1951 was to discover that the automation was actually not (or no longer) following their instructions. This was discoverable not through a change in aircraft behavior (as it usually is during automation surprises), but through *a lack of change* of aircraft behavior (and a lack of mode change). The aircraft did not stop slowing down, and the automation did not change mode. The crew would have had to discover one or two non-events, in other words. Research shows that discovering non-events is very difficult, particularly when initial system behavior is consistent with expectations, when design does not show the behavior but only the status of the system, and when there is no basis in a crew's mental model to expect the non-event(s).

13) Believing, on the basis of their training, documentation and experience, that they had insulated their cockpit set-up from any problem with the left RA, the TK1951 flight crew was surprised by the automation when it turned out that it had not been keeping the airspeed. When they did notice as a result of the stick shaker, and tried to intervene, it was too late in this situation, for this crew, to recover.

14) Post-accident manufacturer recommendations that, in effect, tell flight crews to mistrust their machine and to stare harder at it not only mismatch decades of human factors and automation research, but also leave a single-failure pathway in place.

In this report, quotes from the Cockpit Voice Recorder (CVR) are presented in their original language/utterance. In case of Turkish cockpit conversation, the Turkish is presented first and then an English translation is offered.

# CONSIDERATIONS FOR A HUMAN FACTORS ANALYSIS

*Human factors* is a scientific field that focuses on the interplay between people, technologies and organizations so that goods or services can be produced safely and effectively. Human factors is an interdisciplinary field that encompasses, among other things, psychology, technology and engineering science, as well as sociology, organizational theory and management. Human factors can concern everything from fundamental ergonomic questions regarding the design of systems and displays, to complex associations between management staffs' ways of expressing their objectives and how this affects everyday work in an organization (Dahlström, Laursen & Bergström, 2008).

The aim of a human factors investigation is to try to understand why it made sense for people to do what they did—against the background of their physical and psychological work environment. In a human factors analysis, the data we have available about *what* people did never allows us to entirely close the gap to *why* they did it. The data, in other words, always underdetermine the explanations a human factors analysis can offer. It is not surprising that human factors can often be seen as vague, speculative, indeterminate, or as merely a matter of opinion, and not in great need of the same level of expertise as a technical part of an investigation (after all, everybody has a mind, so we all should have some idea of how a mind works, right?).

What has animated the human factors analysis of the accident of THY 1951 is a dedication by the Dutch Safety Board to explore, at the greatest level of detail and insight available, the possible reasons for why the pilots did what they did. Such a better understanding holds an important key to the accident and the countermeasures and improvement recommendations that flow from it, because the performance of pilots in a situation as encountered here is evidently critical for the recoverability of such a situation.

The results of a human factors analysis may remain infinitely negotiable and never entirely certain. That is, others may always be able to draw different conclusions about human performance from the data as they are offered. This is not to say that some explanations are not more plausible than others. Some explanations account for more of the data, or account for the data better and at a greater level of detail, or make better use of the theoretical, methodological and research bases for contemporary human factors work. The account offered here aims for a maximally plausible explanation of why the pilots did what they did. It uses a number of analytical and theoretical perspectives to lay out the broadest possible basis for its conclusions about what went on in that cockpit, particularly during the last minutes of the flight.

## A word on hindsight

When confronted with the wreckage of an accident such as THY 1951, it is always tempting to ask why people didn't do something that, with hindsight and knowledge of outcome, would have made much more sense. This is a natural response that we as people use not only to express our surprise and incomprehension and perhaps even indignation about what went on in a particular workplace, but also because we want to extract maximum learning from what we see and read and hear about. Questions or remarks such as "But why didn't they…?" or "If only they had…!" have a role to fulfill because they allow us to prepare ourselves and others for similar situations, and perhaps we may then do what we now wish others had done.

Such remarks or questions have no role to fulfill, however, in explaining why people did what they did. Saying what people could or should have done does not explain the reasons behind what they in fact did. Such questions and remarks are, literally, counterfactual—counter the known facts of the accident. What people didn't do (but could or should have done) is not as pressing as finding out why they in fact did what they did, which is where a human factors investigation needs to direct its resources.

A standard response after a mishaps is to point to the data that would have revealed the true nature of the situation. But knowledge of the "critical" data comes only with the privilege of hindsight. If such critical data can be shown to have been physically available, it is automatically assumed that it should have been picked up by the operators in the situation. Pointing out, however, that it should have been does not explain why it was perhaps not, or why it was interpreted differently back then. There is a difference between:

• **Data availability**: what can be shown to have been physically available

somewhere in the situation;
- **Data observability**: what would have been observable given the features of the interface and the multiple interleaving tasks, goals, interests, knowledge of the people looking at it.

The mystery, as far as an investigation is concerned, is not why people could have been so unmotivated or stupid not to pick up the things that you can decide were critical in hindsight. The mystery is to find out what *was* important to them, and why.

The most important factor is that the people involved in the accident didn't know that the accident was going to happen. Which we *do* know. The people involved did not enjoy the hindsight we now have. If they had known, they would almost certainly have done something other than what they factually did, because we have to assume that it is not the outcome they desired. This is why counterfactuals, which talk about something the crew didn't factually do (for example: "if only they had…") are not very helpful for understanding why the crew did what they factually did. This also goes for reference to procedural guidance that would have applied during the flight and the approach. While it is easy to point out where such procedural guidance may not have been followed, the real point is to find out why it would have made sense to the crew not to precisely follow this guidance in the context of their flight and their unfolding understanding of the circumstances.

Knowing the outcome of a sequence of events can easily bias an investigation toward those data points that we now know were significant; that showed the real nature of the situation. But the world surrounding the crew at the time consisted of much more than just those few data points that we now know, with the outcome in our hands, were critical. In what was likely a cognitively noisy and pressurized situation, with time constraints, multiple interleaving and overlapping activities and tasks, and a host of indications that needed attention, the crew had to accomplish a meaning integration that we, in hindsight, have no trouble accomplishing. But that is because we know the outcome. Without knowledge of the outcome certain tasks and data points become glaringly obvious. But that is only with hindsight.

Knowledge of outcome tends to simplify a situation down to a choice to notice the critical data or not notice the critical data, to have the accident or not have the accident. That, of course, is not how the world looked to the crew at the time. This conversion into simple, linear causality is called the hindsight bias and it significantly distorts our ability to understand why it made sense for people to do what they did. The hindsight bias occurs because *we* know, and *we* can start from, the outcome. We can trace back from it, into the assessments

and decisions that led up to it. Tracing back up the causal flow, it is easy to identify the junctures where people could have done something to avoid the outcome; where they could have zigged instead of zagged. Again, this makes us no wiser in explaining why the crew did what they actually did. The question is not what the crew should have done in order to avoid the outcome we now know (i.e. given *our* understanding of the situation). The question is why the crew did what they did given *their* understanding of the situation (which did not include the outcome).

One effect of the hindsight bias is depicted below.
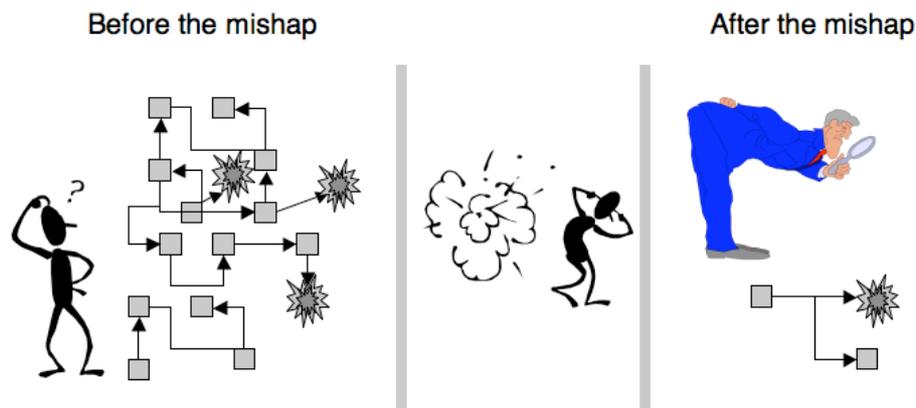


*The hindsight bias: before an accident the crew confronted the world in its full complexity, with a large number of choices to make and potential risks to manage. After the accident, that complexity can get boiled down to a single binary decision to see or not to see a particular piece of data, which of course significantly oversimplifies the situation as the crew faced it at the time (The idea for this image comes from Richard Cook, M.D.).*

That is why the analysis offered here (as do many other human factors analyses) tries to reconstruct the unfolding situation from the point of view of the flight crew as faithfully as possible. If we want to understand why people did what they did, we have to reconstruct the situation as it unfolded around them at the time, not as it looks to us now, in hindsight. If we can't comprehend what the crew did or why they did it, it is not because the crew's behavior was incomprehensible, but because *we* have taken the wrong perspective, the wrong point of view. The crew, after all, acted on *their* understanding of the situation as it rapidly developed around them at the time. What was the pace, the tempo, the order in which cues and indications about the situation emerged and confronted the crew at the time? This is likely different from *our* understanding that we have gained over months of study after the fact. The aim of any such reconstruction is plausibility—knowing all the time that no amount

of dedication, theory, or analysis is enough to ever entirely close the gap to full certainty about why the crew did what they did.

## The scope of this human factors analysis

The starting point for the detailed portion of this human factors analysis is chosen as the moment when the crew receives clearance to descend to 4000 feet and told by the approach controller to expect vectors for an ILS to runway 18R at Schiphol. Portions of crew interaction that are possibly significant (such as the approach briefing) and that occur before that moment, are used in the analysis in greater detail too. The remainder of the preceding (almost) two hours of voice recorder and (many more hours of) flight data reveal no significant details that could somehow, in any plausible way, have altered the dynamics of crew interaction and human performance in the last minutes of the flight. That said, it is worth noting that in its communications with each other, the cockpit crew is dedicated exclusively to the execution of the flight (a line training flight, after all). There is hardly any social or extraneous discussion between the three crewmembers present in the cockpit, and none at all in the latter portions of the recording.

It has become custom in human factors analyses to spread out from the so-called "sharp end" (for example, the cockpit) and up into the operation and organization and regulatory or industry environment that surrounded it. In other words, human factors analysis should not just focus on the proximal factors (those close in space and time to the actual accident), but also on the distal factors that contributed and could again express themselves (but then in a different incident) if not addressed. The aim of a distal reach is to explore how aspects of training, documentation, operational or managerial decision making, regulatory influences or industry standards may all have influenced what people at the sharp end saw as making sense at the time. The value of doing this is as great as its difficulty. The further away in space-time we move from the accident, the more difficult it becomes to couple particular actions and assessments (by people far removed from that cockpit) to what went on at the sharp end on any particular occasion.

But the attention to distal factors, factors from further away in space and time, is justified (Maurino *et al.*, 1995). People at the sharp end do not work in a vacuum. Their performance is in great part the product, or the expression, of the environment in which they work, the environment from which they come, the environment which prepared them for the job. This includes, for example, the operational air traffic environment in which pilots executed their flight; how an entire airline or industry has trained and educated its pilots, how the industry

has designed and maintained its equipment, how it has managed earlier incidents that bear a relationship to what happened on the accident day, and so on. If it made sense for these particular people to do what they did, it may make sense to others as well who find themselves in a similar situation. Thus, attention to distal factors is necessary (where possible) also from a preventive perspective.

## A word on time

There is something hugely artificial about any *post-hoc* analysis as conducted and the format in which it is delivered. It has taken months to excavate and study the data remnants from a few minutes of human performance. And it has taken scores of pages to represent the result of this study. In other words, we have had the luxury to consider and reconsider the assessments and decisions and actions of a crew, not only with the benefit of hindsight, but with the benefit of ample time and resources, able to draw on expertise from wide and far. This contrasts sharply with the few minutes the crew had to consider their assessments and decisions and actions in real time. The time from the crew's first contact with the approach controller at Schiphol to the crash is roughly equal to the time it takes to read this page. The time in which things finally went sour during flight THY 1951 was less than the time it takes to read this paragraph. The time available for recovery was roughly what it takes to read its last few sentences.

In the sequence of events in the next section, the time for the particular chunk of assessments and actions is depicted on a bar stretching from 09:15 UTC to 09:26 UTC. As the sequence of events progresses, more things happen, and time available typically becomes less. The same effect is depicted on the maps that display the progression of the sequence of events. As the sequence progresses, those segments become shorter and shorter, while the number of tasks to be accomplished in them becomes larger and larger. These are two visual attempts to convey the compression of work and increasing task load faced by the crew of TK1951 (and, indeed, most crews on approach to a major airport).

# A SEQUENCE OF EVENTS

The beginning of each sequence of events is of course arbitrary. It can always be debated at what point the events that led up to the accident were set in motion. This can, in some sense, be said to have been triggered long before the day of the accident. For the purposes of this analysis, however, the sequence of events is started with the first time the landing gear warning horn sounds at 09:15:06 (all times in the sequence of events are UTC, and expressed as hh:mm:ss), when the TK1951 is still at 8000 feet and northeast of the airfield, vectored towards an approach for runway 18R.

It is difficult to convey, in the quietly linear format of a paper report, the activities and cognitive noise of the situation in the cockpit (or, for that matter, any cockpit) on approach to a busy international airport. The sequence of events between the first landing gear warning horn and the crash played out in just over ten minutes. In that time, the crew had to interact with each other about the conduct of the approach, talk with ATC multiple times, as well as talk to ground personnel, change headings twice on the basis of ATC clearances to be vectored for the ILS localizer, then figure out that they were vectored high onto the ILS and do extra work to capture the glide slope, slow the airplane down, configure the airplane for landing with flaps in multiple stages and landing gear, check the preparation of the cabin for landing, complete the landing checklist, and descend from 8000 feet altitude to the level of the runway. The sequence of events is broken up into analytic chunks, where it matches phases of flight or cockpit work.

The landing gear warning horn intruded loudly into a large initial portion of those ten minutes, and during almost the entire time of the approach a busy ATC approach frequency (talking mostly to other airplanes, sometimes to TK1951) is transmitted through the loudspeakers and headsets in the cockpit. Again, this is almost impossible to represent faithfully through any post-accident format, whether written in prose, in table form or pictorially. The sequence of events presented below is written in present tense, the times are all UTC (local time was UTC+1).

TK1951 had a cockpit crew of three onboard on the 25th of February 2009. There were two pilots (in the right and left seat) and one so-called "Safety Pilot" who was seated in the center cockpit observer (jump) seat. The Safety Pilot was a fully rated and experienced THY (Turkish Airlines) F/O himself. His presence was required in the cockpit according to THY's Operations Manual because the F/O, in the right seat of the cockpit and in the role of pilot flying (PF) was a newly rated B737 pilot at THY and being trained according to what is called *Line Flying Under Supervision* (see more in the section "CRM and the Intervention Decision). The Captain, in the left seat of the cockpit, had the role of Commander of the aircraft as well as Line Instructor.

*09:15:06 — 09:18:00*
*This sequence starts with the first time the landing gear warning horn sounds at FL080 and ends with a possible conclusion by the Captain about it.*



09:15                                                                 09:26



The aircraft is flying on an westerly heading. The crew is slowing the aircraft down from over 300 KIAS to 250 KIAS during this time, while descending from FL080 through FL055. At 09:15:58, amidst a flood of clearances to other aircraft on the approach frequency, TK1951 gets a clearance from the approach controller to fly to SPY (Spijkerboor VOR) and expect the ILS for runway 18R. It is around 35 track miles away from the touch-down point on runway 18R.

During this clearance, the landing gear warning horn sounds. It starts at 09:15:06, and continues to sound (with one small interruption) until 19:16:35 (meaning it sounds continually for almost a minute and a half). The horn is then silent for 37 seconds, only to start again at 09:17:12. It continues to sound about another half minute, until 09:17:46.

At 09:17:55, the Captain (PM and instructor) says "Landing Gear." About a minute earlier, the Captain had announced "Radio Altimeter."

The discussion among the flight crew is a result of the warning horn. The horn would likely be accompanied by an RA reading on the left Primary Flight Display (PFD) that was not consistent with actual aircraft altitude at that moment. This in part may have prompted the Captain to mention the radio altimeter (and again not much later announce "Radio Altimeter") during this discussion. In the context of the flight at that moment, a landing configuration warning horn coming on and off, makes no sense. There is little guidance available in the Flight Crew Operations Manual Volume 2 (FCOM2) about the logic of the warning horn and nowhere does it say that it could sound with a faulty sensor in one of the RA's. The information available to crews is that the horn sounds "to alert the flight crew whenever a landing is attempted and any gear is not down and locked" (FCOM2, chapter 15). No landing is being attempted yet (the aircraft was at 8000 feet).

*09:18:00 — 09:19:43*
*This sequence starts with the F/O announcing he has pre-set the courses for the ILS frequency for runway 18R, and concludes when TK1951 gets a new heading (265°) which puts them on a base leg.*

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|

09:15                                                                                    09:26

At the beginning of this sequence, the airplane is flying towards the SPY VOR on a westerly heading. The track miles to the threshold are approximately 25 nm, and the aircraft is flying 250 knots and descending to FL040, where it levels off for about half a minute.

At 09:18:59, the landing gear warning horn sounds for just two seconds. Three seconds after the warning goes silent, at 09:19:04, the F/O says he is reducing

to 220 KIAS, and almost at the same time (09:19:06), TK1951 is cleared to descend down to 2000 feet, while continuing to the SPY VOR beacon (as per the clearance).

At 09:19:19, about 21 nm to go to the threshold (according to the plan to fly via SPY), the F/O asks the Captain "Vereyim mi, hocam," whether he may use LEVEL CHANGE, a mode that idles the thrust levers and uses aircraft pitch to maintain the airspeed set on the mode control panel. Immediately after the question, the A/T goes into retard mode (consistent with LVL CHG mode), suggesting the F/O has executed what he asks the Captain about. The Captain replies with "OK" a few seconds after this.



At 09:19:27, with the airspeed at 240 knots, the Captain asks the F/O "Surati dusurecen mi," whether he's going to slow down yet. The aircraft is now 20 nm from the threshold, according to the planned track via SPY. The F/O replies that he is going to reduce (the 220 target setting had already been selected on the Mode Control Panel (MCP) almost half a minute earlier), and that they're not 13 miles (away from the field) yet (they are about 20 track miles from the runway).

The F/O and Captain then coordinate their altimeter settings (which is the procedure when changing from flight level to altimeter altitude) on the Captain's initiative, to a setting of 1027. During this entire time, clearances to other aircraft are given by the approach controller and responded to, and all this can be heard on the cockpit speaker.

In descent in aircraft such as the B737-800, there is always a tussle between slowing down and going down. The aircraft needs to reduce 30 knots airspeed *while* losing 2000 feet of altitude. The result is that it does both slowly.

*09:19:44 — 09:22:40*
*This sequence begins with the turn onto base leg (heading 265° by radar vector) and ends with a turn onto heading 210° to intercept the localizer for the ILS of 18R.*
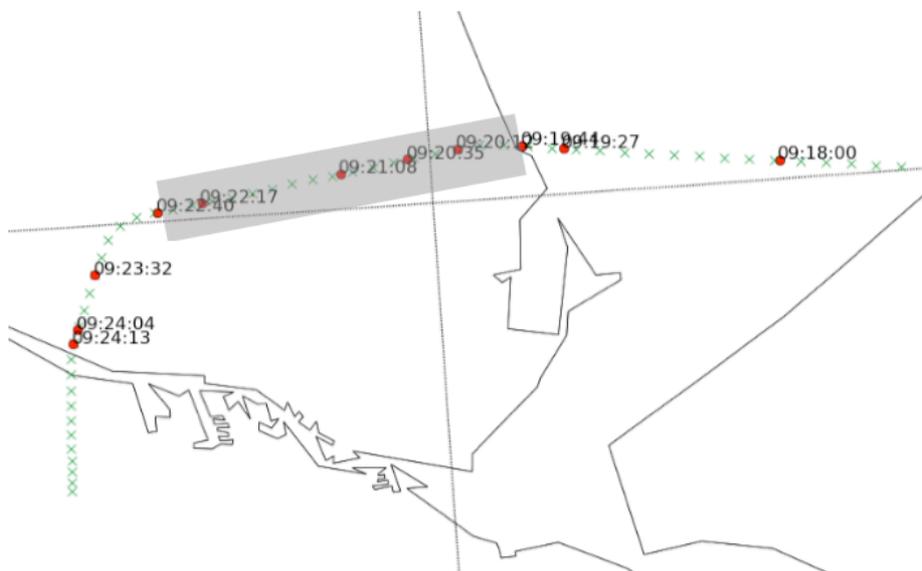


At 09:19:44, TK1951 is cleared onto a heading of 265°, which puts it on a baseleg for runway 18R. The Captain reads back 265° to the approach controller, and then repeats it to the F/O in the next second, who then reads back 2-6-5 to the Captain again a second or so later, presumably as he selects a heading of 265° on the MCP. This heading means TK1951 is no longer going all the way out to the SPY VOR, but rather cuts west onto a base leg towards the extended centerline of runway 18R, shortening the track miles for the approach slightly, to about 18 nm at this point. The aircraft is slowing down to the earlier selected 220 knots now, and is descending through 3900 feet, for 2000 feet when this clearance comes.

At 09:20:12, the Captain says "FMS in" ("your FMS") to the F/O. At 09:20:20, the Captain says to the F/O "Telsiz sende" ("you have the radio,"); the F/O asks him to repeat this but the Captain does not, a second later he is on the radio

with Ground Ops at AMS to coordinate passenger numbers and parking position. About 20 seconds after announcing that the F/O has the radio (meaning the F/O is responsible for contact with ATC, of which none was necessary during this time), the Captain is back from his brief coordination with Ground Ops and announces that they will have the parking position (and presumably taxi route towards it) as they had briefed before.

By 09:20:35, or less than minute after being turned onto 265°, the aircraft has slowed down to the target speed of 220 knots. At 09:21:08, there is a sound of a person briefly whistling in the cockpit, most likely the Captain. At 09:21:20, the aircraft has about 15 track miles to go to the touch-down point. The aircraft levels off at 2000 feet around 09:22:00, while still on the base leg.

Less than 10 seconds later, while still on base leg, at 09:22:17, the F/O asks for Flaps 1, and a speed check. Flaps 1 are selected and a new speed of 195 knots is set on the MCP in the seconds following. The aircraft slows down towards 195 knots. It is flying in ALT HOLD mode and HDG SEL mode during this time, remaining level at 2000 feet.

The approach controller is still coordinating other traffic during this entire time, which is audible through the cockpit loudspeakers.

The aircraft is flying in HDG SEL mode on the 265° base, so one possibility of the "Your FMS" remark is that the Captain wants to point out to the F/O that he can or should select the initial approach fix to the ILS of 18R on the FMS so as to get a better reading of the track miles to go. Clicks are heard that could be the yoke chart holder (possibly for pulling up the ILS plate in front). The F/O offers a short response to the Captain at 09:20:15 that cannot be made intelligible from the CVR recording.

*09:22:40 − 09:23:35*
*This sequence begins with TK1951 being turned onto an intercept heading of 210° and ends with a selection of flaps 5 and concomitant MCP speed reduction.*

09:15                                                                      09:26

At 09:22:40, TK1951 is turned onto heading of 210° and cleared for the approach runway 18R. They are over 10 nm from the threshold of 18R at that point. The Captain reads back the clearance and the F/O announces that 210 is now set, at 09:22:49. This represents an intercept heading of 26° to the final approach course.
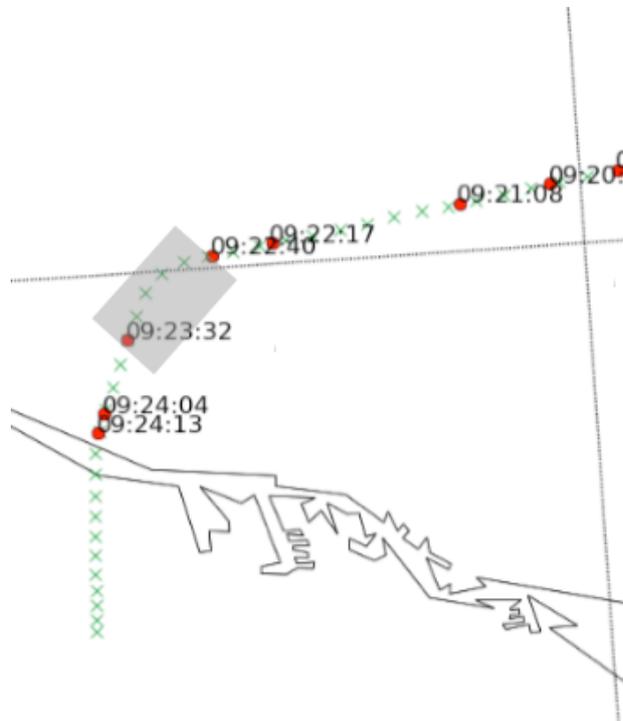


Ten seconds later, still on the intercept heading, at 09:22:59, there is an attempt to engage the second autopilot (CMD A). The second autopilot, however, does not engage—instead, the right autopilot (CMD B), flying the aircraft up to that point, disengages, accompanied by the autopilot disconnect warning. The right autopilot is re-engaged after 3 seconds and the aural warning ceases. A few seconds later, at 09:23:06, the F/O says "Korslar aktif, hocam" ("Courses active") then announces "Second autopilot engaged" (likely meaning CMD B, rather than the second of two autopilots), and the Captain replies with "Tamaam" ("Okay"). CMD B then remains on (CMD A remains off). The aircraft is still level at 2000 feet and flying the target speed of 195 knots.

At 09:23:32, about 8 nm from the runway threshold, the F/O calls for flaps 5, which are then set, and a new target speed of 170 knots is dialed into the MCP. The aircraft starts slowing down to 170 knots from 195 knots.

Whereas the written guidance of LVNL (ATC in the Netherlands) says that aircraft should be turned onto the localizer between 8 and 5 nm, it is also common to coordinate earlier turn-ins with pilots themselves. Either the

controller will ask the pilots whether they can accept a short line-up, or the pilot him- or herself will ask ATC if such a line-up is possible.

The crew of TK1951 was not asked whether they could accept a short line-up onto the localizer, they did not themselves ask for it, and TK1951 was not given a descent clearance to a lower altitude so as to be in a better position to capture the glideslope. Indeed, a line-up such as the one given to TK1951, without coordination, or a question, or a descent clearance to a lower altitude, is inconsistent with such practice (and the written guidance). It could, however, be a natural result of increasing controller workload at that time (the sector that this controller was responsible for was about to be split in two shortly after the approach of TK1951, so as to better handle the traffic volume that was rising).

It is standard THY procedure to fly approaches on dual autopilots down to becoming visual with the runway or down to approach minima. The reason for this standard operating procedure (which is actually recommended by Boeing in the B737 Flight Crew Training Manual, or FCTM) is that it offers automatic go-around capability. If two autopilots are engaged, then all a crew needs to do when they do not see the runway or approach lights at minima is push the so-called TOGA buttons (Take Off Go Around) and (if a feature called automated LNAV engage is installed) the aircraft will itself fly the missed approach procedure (provided it's been programmed into the FMS). All the crew would need to do is clean up the aircraft (landing gear and flaps).

The F/O's reference to the "Courses active" is the beginning of the sequence "CMAPS" used by B737 pilots to set up the cockpit for an approach (Courses, Manual (this appellation applies to the selection of the ILS frequencies on the second-generation 737s only, it's called "Active" on the 737NG but the CMAPS mnemonic survives), Approach, Second Autopilot, and Standby Horizon). The F/O then repeats "Engaged." His remark "engaged" most likely refers to CMD B (fitting the "Second Autopilot" part of the CMAPS mnemonic).

*09:23:35 — 09:24:10*
*This sequence begins with the selection of the landing gear down, and contains the capturing of the localizer.*

09:15                                                                 09:26

At 09:23:43, the landing gear warning horn begins to sound and lasts 5 seconds. As it is silenced, the F/O offers a garbled "flaps" and then asks for gear

25

down and flaps 15 and the MCP target speed is changed down to 160 knots (at 09:23:49; this is about 17 seconds after selecting flaps 5 and *its* target speed of 170 knots). The aircraft is level at 2000 feet. When 160 knots is dialed on the MCP, the airspeed is still over 180 knots (still slowing down to the previous target speed of 170 knots).

The aircraft is about 7 nm from the threshold of 18R at this point (consistent with a typical ILS approach that may be flown at 180 knots to 6 nm and 160 knots to 4 nm).
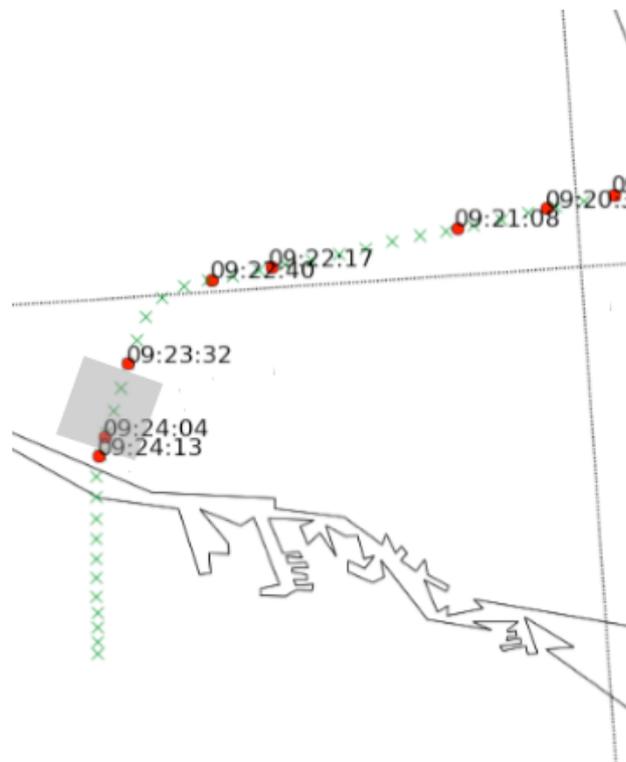


The order by the F/O of gear down and flaps 15 even before localizer capture indicates that the crew is aware of how the turn-in is going to put them on the final approach course and that they need to take action to slow the airplane down (which the landing gear does most effectively).

At 09:24:04, the Captain announces that the localizer is alive, the F/O confirms and at 09:24:09 the Captain announces that the localizer is captured. The aircraft is turning towards the runway heading at that time, and is still at 2000 feet. The speed is about 170 knots at localizer capture and slowing down (the MCP target is still set at 160 knots for flaps 15). The localizer is captured at less than 5,5 nm from the threshold.

*This sequence begins with a mode change associated with capturing the glideslope from above, contains the hand-over to the tower controller and concludes with TK1951 capturing the glideslope and receiving landing clearance.*
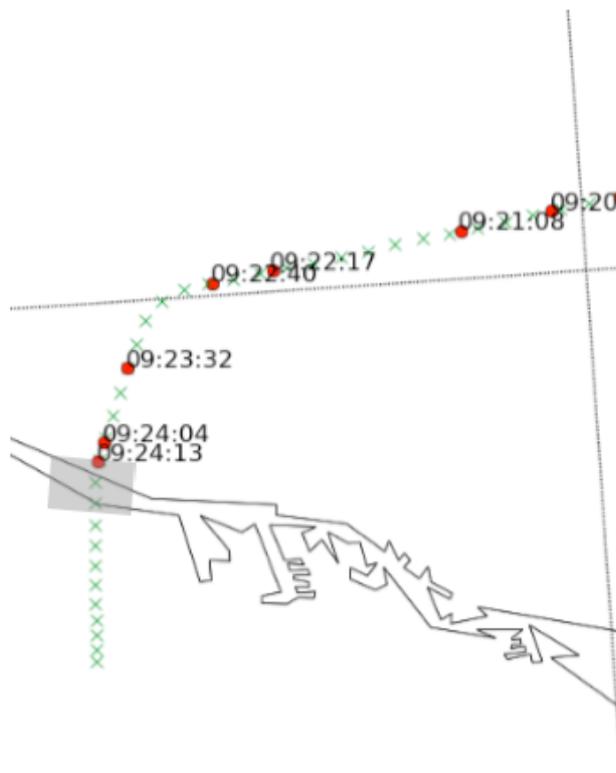


At 09:24:13, V/S mode is engaged (ALT HOLD mode is thereby cancelled), a second later the A/T goes into RETARD mode, and the aircraft starts descending from 2000 feet. Seven seconds after selecting V/S mode, the airspeed has decreased enough to attain the target of 160 knots (in fact, it briefly dips just below 160), only to increase again to 170 knots at around 09:24:45, consistent with a descent in V/S mode. The rate of descent, with throttles in idle and speed increasing to about 170 knots, goes up to about 1350 feet/minute, consistent with an attempt to capture the glideslope from above. Around the same time the F/O announces, "Speed 1-4-0, setting set."  The MCP target speed is still 160 (and flaps are 15). At 09:24:18, there is the sound of a cabin chime.

With the reason for this unbeknownst to the crew, the autothrottle reverts to RETARD mode when V/S is selected (rather than MCP SPD). The throttles are

pulled to their idle stop and kept there by this mode. For the next 70 seconds, throttles at the idle stop is exactly what the crew expected and (thought they) had instructed the automation to do. The aircraft had to go down and slow down at the same time.

At 09:24:24, TK1951 is handed off to the tower controller, which the Captain reads back to the approach controller at 09:24:27. Before the Captain contacts the tower on the new frequency, the Safety Pilot (on the center cockpit observer's seat) announces at 09:24:36 to the Captain, "Hocam, radio altimetre arizamir var, hocam" ("We have radio altimeter failure").

The Captain replies to the Safety Pilot with a drawn-out "Tamaam" ("Okay"), and then contacts the tower 5 seconds later. While he is still talking, TK1951 captures the glideslope at 09:24:46 and the aircraft's vertical speed becomes consistent with an ILS glideslope. The altitude is approximately 1330 feet at glideslope capture, the speed is 169 knots and starting to come down again, flaps are 15 and the landing gear is down.

The showing of -8 on the Captain's PFD, seen by the Safety Pilot, may be one possibility that prompted him to say "We have radio altimeter failure," the RETARD mode on the FMA may be another (but it is impossible to connect this mode annunciation to a left RA failure on the basis of training and documentation available to B737 pilots). The remark may also have been a reminder about the earlier discussion when TK1951 was still at 8000 feet and the crew talked about the Landing gear warning horn. The Captain's response "Okay" makes sense in the light of the selection of CMD B with FCC B selected as Master. This would, according to the knowledge available from B737 training and documentation, insulate the approach from any left RA problem (see under the section "TK1951 as automation surprise").

TK1951 is given landing clearance at 09:24:48, almost simultaneous with the capturing of the glideslope.


*09:24:50 — 09:25:20*
*This sequence begins with the read-back of the landing clearance and ends with the setting of flaps 40 and a concomitant reduction in target speed.*

09:15                                                                           09:26

The Captain reads back the landing clearance at 09:24:52. Immediately after that, the F/O announces "Established altitude set," likely referring to the missed approach altitude on the MCP. The airplane is descending through 1200 feet, the speed is coming down toward the target (for flaps 15) of 160 knots, at this point it is still a few knots above 160.
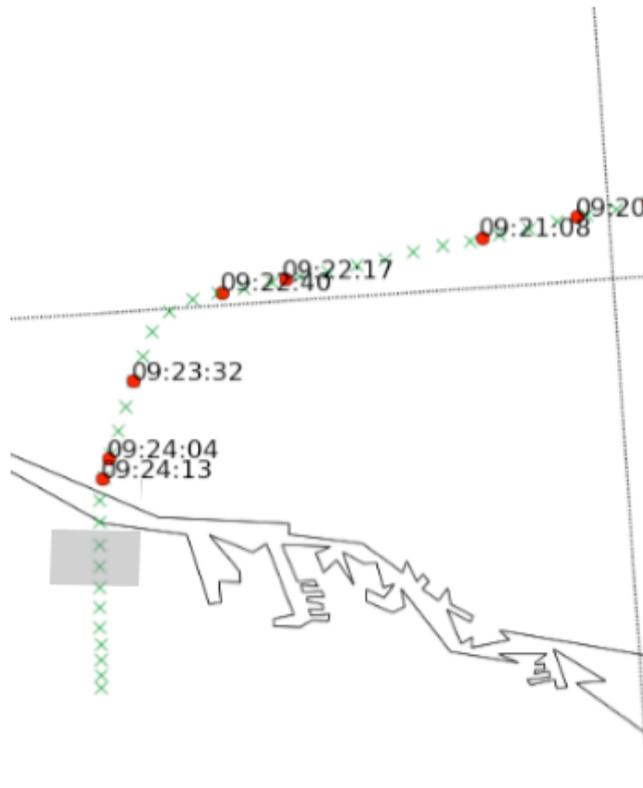


The crew would likely not have seen the runway or approach lights at this point yet. From 09:21:39, the weather reported was wind 210/8, visibility 4500 meters with a temporary visibility of 2500 meters, mist scattered at 700 feet, broken at 800 feet and overcast at 1000 feet (temperature 4, dewpoint 3).

At 09:25:04, almost exactly the moment that the aircraft passes through 1000 feet, the Captain announces "One thousand feet." There is no automatic call-out of "one-thousand." The F/O says "check." The speed at this point is 160 knots, the target for flaps 15.

Five seconds after the previous exchange is completed, the Captain calls for flaps 40, flaps 40 are selected around 900 feet, and the F/O then says "Speed set," most likely referring to the new target speed the F/O would have then dialed into the MCP: 144 knots (VREF+5). Then there is the sound of the speedbrakes being armed. Then, at 09:25:19, the Captain says, "Yes, not in checklist completed—Speedbrake." This prompts the F/O to complete the

landing checklist (see next sequence). Up to this time, the airplane is descending through 800 feet and the speed is coming down from the previous target of 160 knots toward the new target of 144 knots.

*09:25:20 — 09:25:48*
*This sequence covers the completion of the landing checklist and speed decay below MCP target speed.*
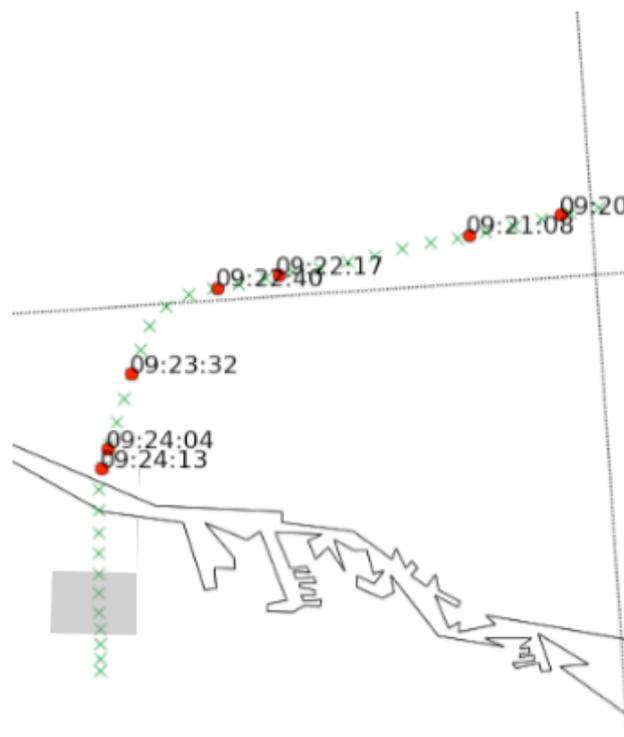


When this sequence starts, the airspeed is still a few knots above the flap 40 target speed of 144 knots, and coming down. The F/O immediately follows the Captain's request of "Speedbrake" with "Speedbrake armed, green light." At about this time, the airspeed goes below the target of 144 knots (the A/T remains in RETARD mode) and the AFDS starts trimming the nose up. The aircraft passes through 750 feet. The Captain then says "One, one, one," and the F/O repeats, "Speedbrake armed, green light." The Captain almost simultaneously says "Landing gear," and the F/O responds "Gear down, please, three green." (the landing gear has been down since before glideslope capture).

The Captain says "Flaps" and the F/O responds "Flaps 40, green light," the Safety Pilot announces "Cabin report confirmed," and the F/O almost simultaneously says "Missed approach altitude set." The Captain then says "Bes yüz feet" ("Five hundred feet"), the F/O says "All lights on." The Captain then asks the Safety Pilot to "Kabine de bi kas etsene" ("Please warn the cabin crew"), which the Safety Pilot confirms by saying "uh-huh" and then does by saying "Kabin ekibi yerlerinize" ("Cabin crew take your seats please") over the PA. When he has completed his announcement, the airspeed is about 110 knots and the aircraft passes through 500 feet. Nose-up trimming by the AFDS then stops.

During this time, visual attention of the crew is directed to the checklist, the flap handle and its indicator on the instrument panel, the speed brake handle, the landing gear handle and its associated lights on the instrument panel, the light switches above the Captain's windshield. Also, at the same time, the pilot may have been looking for the runway lights in addition to directing his attention to the various items inside the cockpit associated with the landing checklist. TK1951 is now about 2,5 nm from the runway threshold, which is consistent with the minimum reported visibility at that time (4500 meters). There is no verbal annunciation by any of the crew members of lights or runway in sight, but it is plausible that the runway or approach lights would have become visible at this point.

*09:25:48 — 09:26:00*
*This sequence covers the recovery attempt.*

09:15

09:26
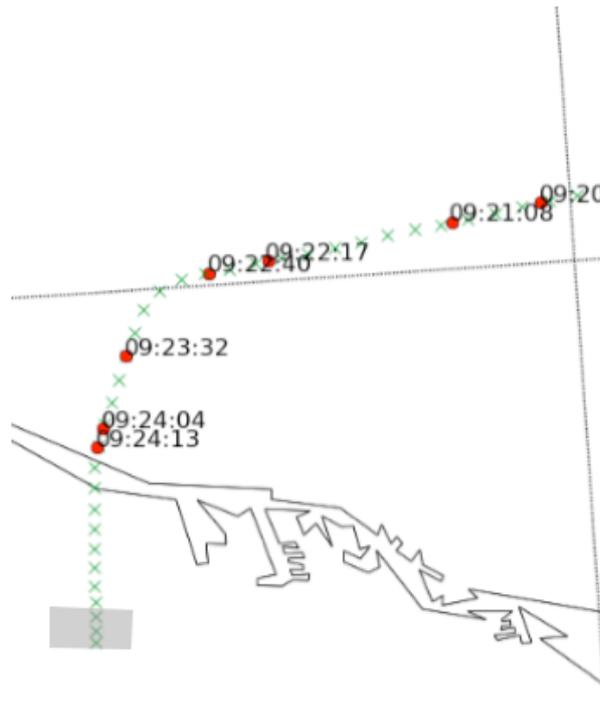
At 09:25:48, the stick shaker starts. The aircraft is at just over 400 feet. During the next three seconds, the Safety Pilot says "Speed," the Captain says "I have," and power is applied to the thrust levers. The thrust levers, however, only go up halfway and then come back to idle. This is because the A/T is still engaged (in RETARD mode).

The stick shaker consists of two eccentric weight motors, one on each control column. They are designed to alert the pilots before a stall develops. The warning is given by vibrating both control columns.

The fact that the thrust levers are pushed forward without disengaging the A/T may be a possible indication of the automation surprise, with the crew not expecting or realizing that the A/T were engaged in a mode that will pull the thrust levers back to the idle stop quickly if they are moved manually.

The Safety Pilot emphasizes "Yüz knot, hocam" ("100 knots") and again says "Sürat, hocam") ("Speed"). At 09:25:51, the A/T is disconnected and the throttles are set to full power. The autopilot is disconnected at 09:25:53 and at 09:25:57, the stick shaker stops, but starts again two seconds later. The aircraft nose is pitched to 8° below the horizon as part of the recovery attempt. The recordings end at 09:26:02 as the aircraft impacts the ground short of the runway, at a pitch attitude of 22° nose up and 10° left wing down.

# TK1951: An Automation Surprise

The case of TK1951 fits the phenomenon of an "automation surprise," and adds a new twist to it. Automation surprises have become defined as those cases where the automation does something without immediately preceding crew input related to the automation's action, and in which that automation action is inconsistent with crew expectations (Wiener & Curry, 1980; Sarter, 1991; Sarter & Woods, 1992; Billings, 1996). In TK1951, the automation did to the aircraft what the crew expected (and thought it) instructed it to do, but in a different mode than the crew had expected, or could have known from their training or experience.

The story of the TK1951 automation surprise starts with an earlier version of the B737, the second generation (-300, -400, -500). Some of these series were equipped with a single Flight Control Computer (FCC), a box manufactured either by Collins or Honeywell, which contains the central computer responsible for taking information from the environment and giving commands to the autopilot, flight director and autothrottle systems of the aircraft. In order to improve navigational performance in certain approaches, all of the next generation (NG) of the Boeing 737 (-600, -700, -800, -900) were equipped with two FCC's instead of one. This offered a number of advantages. The Boeing 737NG FCOM (Flight Crew Operations Manual, Volume 2, chapter 4) explains:

> **"Autopilot Flight Director System (AFDS)**
> The AFDS is a dual system consisting of two individual flight control computers (FCCs) and a single mode control panel.
> The two FCCs are identified as A and B. For A/P operation, they send control commands to their respective pitch and roll hydraulic servos, which operate the flight controls through two separate hydraulic systems. For Flight Director (F/D) operation, each FCC positions the F/D command bars on the respective attitude indicator.
> Normally, FCC A drives the Captain's command bars and FCC B drives the first officer's command bars. With both F/D switches ON, the logic for both pilots' F/D modes is controlled by the master FCC…"

These FCC's get their information from a number of sensors around the aircraft (the Computer-based training module for B737 pilots says these include the alpha vane and pitot tube), and also, says the FCOM (chapter 4):

> "Two independent radio altimeters provide radio altitude to the respective FCCs."

The FCOM2 (chapter 4) then further explains:

> "If a F/D switch is ON, the light indicates which FCC is controlling the F/D modes.
> • illuminated – related FCC is controlling F/D modes.
> • extinguished – F/D modes are controlled from opposite FCC
> • both lights illuminated – each FCC is controlling modes for related F/D."

In the case of the approach by TK1951, FCC B had been selected as Master. FCC B was therefore controlling the Flight Director (F/D) modes (and, by extension, directing the autopilot what to do). Autopilot B was selected on. The pilots would have assumed that FCC B was also controlling the autothrottle, using, according to the FCOM, radio altitude from its own "independent radio altimeter." (Nowhere in the FCOM does it say that the related FCC is NOT controlling A/T modes, nor that the A/T always gets its height data from the left radio altimeter.)

The B737 Computer-based training program, a simplified run-through of the systems that make up the airplane, offers and confirms the following about the FCC:

> "The mode control panel sends data to the two flight control computers, or FCCs. The FCCs calculate thrust pitch and roll commands for the other autoflight system components."

> "One of the FCCs is specified as the master FCC. Autopilot and flight director status control which FCC is the master."

> "While each FCC continues to calculate the thrust, pitch and roll commands, the master FCC usually positions command bars for the Captain's and first officer's flight directors."

> "The FCC's are called FCC-A and FCC-B. For Autopilot functions, FCC-A controls autopilot-A and FCC-B controls autopilot B."

"The autothrottle adjusts the thrust levers with commands from the FCC."

An example is then offered on the Computer-based training (CBT) program that shows the left F/D having been selected on as Master (meaning FCC A is in control) and a connection between FCC A and the autothrottle. There is no contrasting picture or example of what the connections would be if FCC B were selected as Master.

The conclusions that the TK1951 crew could reasonably have drawn from this training and documentation would be the following:

- *FCC B is selected as Master, and we have CMD B (Autopilot B) selected on. We will fly the approach on Autopilot B, which gets its commands via FCC B.*
- *FCC B has its own independent radio altimeter, so a strange indication on the left radio altimeter has no consequences for this approach.*
- *The FCC is sending commands to the thrust levers (which must be FCC B because that's the one that is selected as Master).*

The crew, like any other normally trained B737 crew, could not have known that this mental model was false. What really was going on between the FCC's and the autothrottle was not in any of the training materials or documentation that is normally available to pilots undergoing type- or line training on the B737. This, which can be gleaned only from the sort of technical documentation that pilots have no access to, is what was actually going on (and is in part a legacy from the earlier generations of B737's):

- FCC B is selected as Master and CMD B is on, FCC B is sending commands to the right AFDS.
- FCC B may have its own independent radio altimeter, but that has nothing to with the autothrottle.
- The autothrottle **always** gets its height information from the left RA, independent of which FCC is selected as the Master, and independent of which autopilot has been selected on.
- In pre-2005 New Generation model B737's (which the TK1951 one was), there is no comparison or cross-validation of RA data from the left and right RA.

The divergence between the mental model that training and documentation allows pilots to build up and the actual working (which would have been impossible for a crew to know from what was available to them, let alone piece together on the fly), is shown in the figure below.

## Anatomy of the TK1951 Automation Surprise

| What is trained: | What a crew could have concluded: | What is actually going on: |
|---|---|---|
| "One of the FCC's is specified as the master FCC." | We select FCC B as Master | FCC B controls autopilot B. Autopilot B is following the glideslope |
| "Each FCC continues to calculate thrust, pitch and roll commands" "The autothrottle adjusts the thrust levers with commands from the FCC." | FCC B calculates thrust commands; FCC B is giving the autothrottle its commands | |
| "Two independent radio altimeters provide radio altitude to the respective FCC's." | FCC B has its own independent radio altimeter. An inoperative RA on the left has no impact, as the left RA goes to FCC A | The autothrottle gets its height information directly from the left radio altimeter, independent of which FCC has been selected Master. |
| "FCC A controls autopilot A and FCC B controls autopilot B" | We fly this approach on autopilot B, which gets its input from FCC B. | Flying the approach on FCC B does not protect you from a left RA anomaly. |
| (source: Boeing 737 FCOM and Computer Based Training for pilots) | | (source: Boeing 737 technical documentation that pilots do not have access to) |

*Anatomy of the TK1951 automation surprise. Through training and documentation, the crew (as are all B737 crews) could have been led to believe that a problem with the left RA had no consequences as long as FCC B was selected to control the automation that was flying the airplane. In fact, the left RA was still providing height data to the autothrottle, a fact that cannot be found in B737 pilot training or documentation.*

When the TK1951 crew selected V/S to capture the glideslope, they would have expected the A/T to retard the throttles to the idle stop. Speed still had to come down by about 10 knots to the flap 15 MCP speed, and the airplane was now going to increase its rate of descent to catch up with the glideslope. The combination would have left no choice but to retard the throttles. The behavior of the automation, in other words, was expected and consistent with intentions and (for all the crew knew, their) instructions.

*The Boeing 737 and its two RETARD modes*

The A/T mode in which this happened, however, was one of two RETARD modes. The B737 has two RETARD modes:
1. RETARD flare
2. RETARD descent

The RETARD descent is activated when the autopilot commands the thrust levers to the aft stop during a descent condition (this happened, for example, when the co-pilot selected LVL CHG mode earlier on the approach and was entirely consistent with intentions and expectations). The autopilot may command this reduction in power to follow a planned descent path.

In contrast, the RETARD flare mode is activated by the autopilot when a dual channel auto land is performed just before touchdown (and when there is a combination of a particular flap setting and a radio altitude of less than 27 feet).

The RETARD mode will be displayed when this mode is active, there is no visual distinction between RETARD flare or RETARD descent on the PFD. In a normal descent the RETARD descent will be followed by the ARM mode when throttles are at idle. In the ARM mode flight (alpha floor) envelope protection is still active and speed is controlled by the A/T.

While just showing "RETARD" (and not "FLARE") the FMA annunciation gave the appearance as if the A/T went into RETARD descent mode (though V/S would not have triggered RETARD mode on the A/T but rather MCP SPD— knowledge that a pilot who is flying his 17th leg on a B737 is unlikely to have ready at hand). However, the A/T went into the unexpected RETARD flare mode not because the crew had selected V/S mode on the MCP, but because a number of conditions had now been fulfilled, and the A/T was acting according to its own logic: the aircraft was going below 2000 feet RA (in fact, it was at -7 feet RA, according to its only available input to the A/T system), the flaps were more than 12,5 degrees out and the F/D mode was no longer in ALT HOLD.

The RETARD mode that became active during the sequence of events onboard TK1951 is an A/T mode designed for automatic landings in the B737. It is normally combined with the F/D FLARE mode (which pulls the nose up to flare for a landing), but FLARE mode didn't come up in this case. The reason is that the RETARD mode gets its data via the left RA because it is an A/T mode. The FLARE mode is an F/D mode (or Flight Director mode), and FCC B was responsible for giving inputs to *it*. FCC B's own independent RA was supplying correct altitude information, according to which a FLARE would not make any sense (it would be way too far off the ground). In effect, the A/T had decided it was time to land, but FCC B was still flying the approach. The A/T commanded the thrust levers to the idle stop, while FCC B was commanding the F/D and Autopilot B to stay on glideslope (which required more and more nose-up trimming as speed was bleeding away because of commands by the A/T).

Put more simply, the autothrottle was trying to land, the autopilot was trying to fly. Given that the A/T was trying to land, and using a mode specifically

designed for it, there was no minimum speed protection (why have that when you've just landed the airplane?) and no reversion to ARM mode that would allow pilots to take over the throttles without disconnecting them first. As for the autopilot, when flying an approach automatically in the G/S vertical mode, speed control is in the hands of the autothrottle. The autopilot will keep the nose where it needs to be to stay on glideslope, not where it should go to recover lost speed (because maintaining speed is the job of the A/T).

Insidiously, the RETARD mode made aircraft behavior consistent with crew expectations. They needed to go down and slow down so as to capture the glideslope (from above) and get the aircraft's speed down to the speed for the next flap setting. The A/T announced that it did what the crew commanded it to do: it retarded, and aircraft behavior matched crew expectations: it went down, slowed down and then captured and started tracking the glideslope.

Research (see the section "Speed, mode monitoring and the noticing of a non-event") has shown that such propositional annunciations of system state are severely underspecified when it comes to helping users understand system behavior or future behavior. Annunciations of system state (through a propositional abbreviation of the mode the system is in) are critical contributors to automation surprises and similar representations of the system have contributed to failures and accidents in other domains too.

For the crew of TK1951 to discover that the automation was not doing (or no longer going to do) what they (thought they) instructed it to do, they would have had to discover one of two non-events: the reduction in speed did not cease, and the A/T mode did not change. Research has shown that both are exceedingly difficult.

## TK1951 and the research on automation surprises

TK1951 may represent a special kind of automation surprise, but it is far from the first one. In the early 1990's automated systems with considerable operational capability and authority, but limited observability of their behavior, were spreading into cockpits and other operational worlds (e.g. Moll van Charante *et al.*, 1992; Perry, Wears & Cook, 2005). The industry and researchers started to notice a peculiar pattern of incidents in accidents, in which the automation did something that the human operators didn't tell it to do (Eldredge, Dood & Mangold, 1991; Dornheim, 1995). The human operators became, quite literally, surprised by the automation.

One of the major contributors to automation surprises, particularly on flight decks, was found to be the low observability of the automated system's mode behavior (Sarter & Woods, 1992; Corwin, 1995). While mode status on modern flight decks is annunciated (on the so-called FMA, or Flight Mode Annunciator), mode changes are highlighted briefly (not as a change, but simply as a boxed new mode) which, research showed, is often imperceptible in the context of other flight deck work (Sarter & Woods, 1994; Corwin, 1995; Huettig et al., 1999; Mumaw, Sarter, & Wickens, 2001). As a consequence, mode awareness can often be low. In 1995, the journal *Human Factors* published a research paper entitled "How in the world did we get into that mode?" (Sarter & Woods, 1995), which summed up the problem central to automation surprises. "Mode errors" became (and still are) a frequent and logical consequence, in which the crew does something that would have been appropriate in the mode they believed the system was in, while the system was actually in a different mode (Norman, 1988; 1990; Sarter & Woods, 1994). It seemed that packing a strong automated system behind a relatively weak representation of its behavior was a perfect recipe for the creation of automation surprises (Sarter & Woods, 1992; 1994; Sarter, 1997; Woods & Sarter, 2000; Winograd & Flores, 1986).

In TK1951, the "mode error," to the extent that it fits the literature, is not that the crew did something that the aircraft's automation interpreted differently because it was not in the mode the crew thought it was. Rather, the automation seemed to do to the aircraft what the crew expected it and instructed it to do, but the automation did so in a mode that eventually turned out to be the wrong one for that context. The annunciated word, "RETARD," as displayed on the FMA, which, if seen at all, happened to be a word that fit the context and the crew's instruction: Retard the thrust levers so we can capture the glideslope from above because we've been vectored tightly onto final here. There was no accompanying "FLARE" annunciation that could have disambiguated the "RETARD" annunciation. (Remember: This annunciation, after all, does double duty: it stands either for the RETARD descent mode or for the RETARD flare mode). If the FMA doesn't also show FLARE, then any crew may reasonably expect that the RETARD mode seen is the RETARD descent mode, which would be reasonable for that phase of flight and their intentions.

Research has shown that automation surprises are more likely during certain operational circumstances than others. The likelihood of an automation surprise goes up when (1) the automated system does something on its own, without immediately preceding crew input related to the automation's action; (2) training in the underlying automation logic has been buggy or incomplete; (3) representation of automation behavior and intentions in the cockpit is weak, and when (4) circumstances push the flight beyond the routine, particularly when they induce high taskloads. TK1951 fits these conditions. The reversion of

the autothrottle system into RETARD mode was not commanded by the crew (but the concomitant aircraft behavior fit their expectations of what was going to happen); training in the understanding of the underlying logic is buggy and incomplete (across the entire industry); the behavior and intentions of the automation were displayed only as a state (RETARD) on the FMA and the tight turn-in onto final approach for runway 18R of a training flight may have pushed the circumstances beyond the routine and created high taskload.

Another key finding of the automation surprise research is that crews do not notice a "mode error" from the annunciations available in the cockpit. A critical reason, discussed in more detail in the section "Speed, mode monitoring and the noticing of a non-event," is the underspecified nature of mode status annunciations in the cockpit as representations of actual and future automation behavior. As a result, every incident that involved a "mode error" reveals that the "error" is discovered only on the basis of aircraft behavior, not on the basis of the FMA. The divergence between what the crew expected the automation to do with the aircraft and what it factually did may take some time to develop. In other words, it may take some time for the effects of the "mode error" to become apparent in how the aircraft behaves. This is particularly true for TK1951: the initial aircraft behavior (for a full 70 seconds) was consistent with crew expectations and, for all they knew, with their instructions.

The really difficult task for the crew of TK1951 was to then discover that the automation was actually not (or no longer) following their instruction. This was discoverable not through a change in aircraft behavior (as it usually is during automation surprises), but through *a lack of change* of aircraft behavior (and a lack of mode change). The aircraft did not stop slowing down, and the automation did not change mode. In other words, the crew would have had to discover one or two non-events. Research shows that discovering non-events is very difficult (both for human and machine monitors, in fact). In addition, a host of other possible reasons that are discussed below (e.g. task load, checklist and the distribution of visual attention, training context, speed tape design) could have conspired against making the full extent of the divergence between crew expectations and automation behavior noticeable before the stick shaker went off. Only when the stick shaker went off was it obvious that the automation had continued doing something that the crew (thought it) had instructed it to do for longer than was appropriate in that context.

# AUTOMATION TRAINING AND BUGGY MENTAL MODELS

A central irony with the introduction of automated aircraft (such as the Boeing 737 New Generation) has been that greater automation has not reduced the demands imposed on pilots' knowledge and skills. Instead, automation has changed knowledge and skill demands and in certain ways even increased them:

> "One of the myths about the impact of automation on human performance is – as investment in automation increases, less investment is needed in human expertise. In fact, many sources have shown how increased automation creates new knowledge and skill requirements." (FAA, 1996, p. 86)

The airline pilot training footprint, however, has been shrinking across the industry ever since the spread of automated airliners picked up pace (Billings, 1996; FAA, 1996). Regulatory requirements now enable the conversion of a pilot (even one who has never flown transport category airplanes) onto an automated airliner in as few as five weeks.

In response, airline training departments have had to select and address training demands so that they can be fit into a small and shrinking footprint (FAA, 1996). This has led to various strategies, for example training a basic set of modes (or a "cook-book" approach that offers recipes for how to operate a subset of the automation) and leave alternative methods or modes to be addressed during line training or subsequent line operations. This, of course, can lead to the deferral of more complicated automation functions, and of functions that carry a higher vulnerability or risk. Another strategy is to focus teaching on step-by-step procedures that emphasize input-output relationships ("if you push this, that will happen") glossing over any deeper understanding of the underlying system logic or mechanics that make it so.

The result, as Nadine Sarter once put it (e.g. Sarter, 1991) is that pilots have increasingly been taught *how to work the system*, rather than *how the system works*. The basis for this gradual but accelerating and critical shift in pilot training can be found in the early 1960's, when Trans World Airlines (TWA) took delivery of its first DC-9 aircraft (Billings, 1996). This was its first jet with a two-crew cockpit (as opposed to a three- or four-crew cockpit). The airline decided to fundamentally revise its training philosophy for the new aircraft, emphasizing what it called specific behavior objectives (SBO's). This departed from the previous pedagogy of essentially teaching pilots how to take apart and rebuild the entire airplane. The earlier pedagogy emphasized detailed knowledge of how airplane systems were constructed, often offering pilots hands-on experience of the various sub-systems. It taught pilots how the various parts contributed to the whole, and then, on the basis of this, how to operate the whole assembly to produce safe and effective flights.

The novel SBO approach offered significant savings in training time. It taught pilots what they "needed to know" without (seemingly) unnecessary excursions into deeper systems territory. The approach spread across the industry, for example from TWA to United Airlines, and soon followed by others, both in the US and the rest of the world. The move towards "need to know" for pilots was mimicked also by the airplane operating manuals which started to feature diagrams that were significantly simplified simulacra relative to the actual engineered systems or software logic underlying them. Diagrams, as well as textual information, are often incomplete and superficial relative to that base, as the industry's judgment has increasingly become that pilots don't need to know. The shift from the old view that told pilots "you should be able to build the airplane" to the new view "If you can't see it, touch it, or affect it, you don't need to know about it" has been wide and deep, both in range and effect. It has of course been accompanied by the wide and deep spread of microprocessors and software into virtually all airplane systems. You cannot really offer "hands-on" experience with software. Not only have systems become more invisible and untouchable (truly "black boxes" that remain closed for pilots), but they have also become increasingly complex, powerful, and more intricately interconnected.

The industry did take notice of some of the problems associated with this approach. In 1989, the Air Transport Association wrote that although many benefits of aircraft automation had been realized, big gaps between how these systems were designed and how people were prepared to deal with them had opened up: "serious questions have arisen and incidents and accidents have occurred which question [whether] we understand how to design automated systems so that they are fully compatible with the capabilities and limitations of the humans in the system." (ATA, 1989, p. 4). During the preceding decade,

NASA (the US National Aeronautics and Space Administration) had already sponsored several studies into the effects of flight deck automation on crew performance and safety (e.g. Wiener & Curry, 1980; Wiener, 1989) and started its own research initiative to examine human-machine interactions in aviation and future aircraft automation options, including how to train for their operation, in 1990.

In the mid-1990's, partly in response to concerns raised by a series of incidents and accidents with automated airliners, the US Federal Aviation Administration (FAA), together with the European Joint Aviation Authorities (JAA) commissioned a broad and in-depth study into the interfaces between flight crews and automated airliners that resulted in a searching and urgent report (FAA, 1996). The study, conducted by test pilots, human factors researchers, regulators and manufacturers from both Europe and the US, concentrated on design, training and flight crew qualification, and operation of automated systems that dealt with flight path management. At the top of its list of concerns, the FAA report listed the need for "Investments in people (designers, users, evaluators, and researchers)…flight crew training investments should be re-balanced to ensure appropriate coverage of automation issues" (FAA, 1996, p. 4). The authors wrote,

> "In our investigations, we heard from operators that the subtle nature and complexity of automated flight decks result in the flight crews needing additional knowledge about how the different automated subsystems and modes function. Industry investigations have shown that the complexities of the automated flight decks make it easy for pilots to develop oversimplified or erroneous mental models of system operation. Training departments tasked with developing and teaching flight crews how to manage the automated systems in differing flight situations confirm this finding." (FAA, 1996, p. 87)

In a note particularly pertinent to the events leading up to the TK1951 accident (if not prescient), they continued:

> "We heard how the new knowledge and skill demands are most needed in unusual situations where different or extraordinary factors push the chain of events beyond the routine. It is just those circumstances that are most vulnerable to a breakdown in reliable human-automation performance through a progression of miss-assessments and miscommunications." (FAA, 1996, p. 87)

The combination of a training flight with a relatively new first officer, a radio altimeter indicator showing incorrect values on the Captain's side and giving

false height inputs to the autothrottle, and a tight turn-in onto the ILS during a workload increase for air traffic control could count as those circumstances that "push the chain of events beyond the routine." In this case, too, it may have turned out to be one of those situations "most vulnerable to a breakdown in reliability human-automation performance." The FAA report pointed to the importance of preparation through training even for these kinds of unusual situations, and expressed concern about how some qualification programs push this onto the next phase, line training, or even beyond, to normal line flying:

> "Contrary to the content of some qualification programs, the HF Team believes it is important for flight crews to be prepared by their training (as opposed to "picking it up on the line"), so that they will be prepared to successfully cope with probable, but unusual situations." (FAA, 1996, p. 87)

The authors of the report concluded how they were "very concerned about both the quality and the quantity of automation training flight crews receive" (FAA, 1996, p. 33). This concern was confirmed as recently as 2004 in a UK CAA report (p. v):

> "Pilots lack the right type of knowledge to deal with control of the flight path using automation in normal and non-normal situations. This may be due to … a lack of emphasis within the current requirements to highlight the particular challenges of the use of automation for flight path control."

A 1998 Australian study also highlighted how the content and standard of instruction was not considered to provide adequate knowledge to operate advanced-automated aircraft in abnormal situations. Traditional airline training and checking systems, which were developed to maintain flight standards on earlier generations of aircraft, did not necessarily cover all issues relevant to the operation of advanced aircraft (BASI, 1998; CAA, 2004).

*CBT and systems knowledge*

Around the same time that the FAA and BASI reports were published, a new trend was becoming visible in the industry: the accelerating use of CBT or Computer-Based Training for teaching pilots about airplane systems, supplanting (more instructor-intensive) classroom teaching. Whereas this was certainly seen as an improvement in some ways, particularly the opportunity for "free-play" on a laptop or desktop version of the Flight Management System (Dornheim, 1996), it may have helped the trend towards superficiality and cook-book style learning of the operation of automated systems, as CBT's (apart

from the free-play FMS device, that is) typically focus on only a subset of typical or rote ways of how to operate the systems in often routine flight situations. The introduction of more computer-based training has not persuasively been accompanied by more teaching of the underlying mechanics, interactions, principles and assumptions of the system's design (Billings, 1996). The 1996 FAA report warned that:

> "In the absence of this understanding, flight crews are likely to substitute their own model of how the automation works, based on their observations and assumptions of automation behavior. In some instances, the flight crew's model will be incomplete or incorrect, leading to confusion and increasing the potential for error. In critical circumstances, such confusion can lead to a hazardous situation or at least make it difficult for the flight crew to respond in an appropriate manner." (FAA, 1996, p. 34)

This is backed up by much psychological research (e.g. Reason, 1990; Woods *et al.*, 1994). Teaching about technological systems on a "need-to-know" basis carries a risk. There is no such thing as a cognitive vacuum. If something is not taught, that doesn't mean that people don't form their own models and ideas about it. In the absence of adequate guidance (through documentation, training, observability of the system-in-action), people will use their experiences with the system to substitute their own understanding of how the system works for how it actually works. In the case of TK1951, as will be shown below, there really is no adequate training, no adequate documentation and no line experience that can make crews comprehend precisely what is going on; that can help them grasp the significance of this particular sequence of events. This, to many in the industry, cannot come as a surprise. In 1996, the FAA report recommended that the industry:

> "reassess the requirements that determine the content, length, and type of initial and recurrent flight crew training. Ensure that the content appropriately includes:
> · Management and use of automation, including mental models of the automation and moving between levels of automation;
> · Flight crew situation awareness, including mode and automation awareness;
> · Basic airmanship;
> · Crew Resource Management;
> · Decision making, including unanticipated event training;
> · Examples of specific difficulties encountered either in service or in training;
> · Workload management (task management)." (FAA, 1996, p. 34)

Automation has now gradually been introduced into CRM (Crew Resource Management) training and become a greater focus of LOFT (Line Oriented Flight Training) scenarios during for example recurrent training or half-yearly pilot proficiency checks. This may not, however, fully deal with any remaining concerns about the superficiality, brevity and incompleteness of automation training throughout the industry.

*Automation training and its relationship to buggy pilot mental models*

A likely result of the trend towards superficial and less training, is what the human factors literature has called "buggy mental models," where somebody's knowledge of the world and its operation is incomplete or inaccurate. The notion of a mental model, or mental representation of how (a relevant part of) the world works, was developed in the late 1980's and early 1990's. Even if researchers do not agree on how such a model is developed or maintained, it is clear that the function of such models is to order the knowledge of work so as to allow the practitioner to make useful inferences about what is happening, what will happen next, and what can happen (Woods *et al.*, 1994). Accordingly, if the models are buggy (i.e. partial, imperfect, incomplete, inaccurate), inferences about what is happening, what will happen or what can happen can be wrong too. This can be particularly hazardous in cases when subtle failures make that systems do not behave according to expectations, again pertinent to (if not prescient for) the TK1951 accident:

> "If a pilot does not have an adequate internal model about how the computer works when it is functioning properly, it will be far more difficult for the pilot to detect a subtle failure. We cannot always predict failure modes in these complex digital systems, so we must provide pilots with adequate understanding of how and why aircraft automation functions as it does" (Billings, 1996, p. 191).

The training given to B737 pilots today (such as those onboard TK1951) may not offer this "adequate understanding of how and why aircraft automation functions as it does" particularly to ensure pilots' ability "to detect a subtle failure." It is important to see how training for the Boeing 737NG today may relate to the mental model a pilot may bring to a situation such as that faced by the crew of TK1951. Particularly, the question is whether it is likely that a newly type-rated pilot on the Boeing 737 would be able to relate exactly how an autoland works, which indications from the environment are used by precisely which of the aircraft's computer systems and how the behavior of the automation is affected if one of those indications is not working properly. It is

also important to explore whether this understanding is more likely to build as operational experience on the airplane is accrued, and if it does, whether such knowledge can be brought to bear in a meaningful way from the role of instructor pilot (which was the role of the Captain on flight TK1951). This is not academic: in 1996, the FAA (p. 87) noted that:

> "Investigations have shown that the complexities of the automated flight decks make it easy for pilots to develop oversimplified or erroneous mental models of system operation."

Already in the early 1990's, Sarter (1992; 1993) found that buggy mental models contributed to problems with cockpit automation. Presaging the FAA's 1996 conclusions, she found how a detailed understanding of the various modes of flight deck automation was a demanding new knowledge requirement for pilots in highly automated cockpits. And, confirming Wiener's earlier suspicions (1989), training was clearly not up to the task. She was able to trace how buggy mental models played a role in a range of what were called "automation surprises," cases where pilots were surprised by the automation's behavior (see the next section). Buggy knowledge had left significant flaws in pilots' understanding the automatic system's behavior and made it hard for pilots to determine what the automation was doing, why it was doing it, and what it would do next. Research reports from the end of the decade (e.g. Wiener, Chute & Moses, 1999) were able to confirm how very little had changed, despite the FAA's recommendations in 1996. The wash-out rate of pilots transitioning to glass-cockpit (or automated airliners) may have improved over the same time, but it is hard to determine whether that is the result of training improvements or of changes in standards and pass criteria (FAA, 1996; Dornheim, 1996).

The idea of mental models is that people can only act on the knowledge they have. Research has shown that training (and the way in which people are trained for specific tasks) has an impact on the availability and usability of knowledge when people are confronted with operational situations that call for that knowledge (Woods *et al.*, 1994). In studying the acquisition and representation of complex concepts Feltovich *et al.* (1989) found that students and (and even experienced practitioners) are liable to apply knowledge to certain problems in ways that amount to oversimplification if they get taught about these problems in a certain way. If training relies significantly on analytic decomposition (chunking up the problem into several building blocks without regard for the steps necessary to sew these together into a coherent operational picture), students will likely miss or miscomprehend interactions among critical system variables; they may understand some aspects of a complex phenomenon but miss other crucial aspects and be misguided or misled on yet others. In fact,

students may acquire a false sense of understanding and even resist learning a more complex model once the simpler one has proven its apparent usefulness to them. Thus, "...bits and pieces of knowledge, in themselves sometimes correct, sometimes partly wrong in aspects, or sometimes absent in critical places, interact with each other to create large-scale and robust misconceptions" (Feltovich et al., 1989, p. 162).

## Autothrottle in the Airplane manuals

The learning and retention of technical information about how systems work (whether engineered or biological systems—indeed, this appears true for medical education too), there is a tendency on part of the learner to think about the system more in a top-down than a bottom-up fashion. The starting point for a learner aiming to create a stable conceptual mental model of the system is often the overall goals that the system wishes or needs to achieve, rather than the components that make up the overall system. Hence, in part, the problems identified by Feltovich *et al*. (1989) with the mental models that result from a bottom-up, decompositional approach to teaching systems.

Airplane manuals, the basis for training and proficiency checking, typically resort to a bottom-up fashion for presenting their information about the various subsystems that make up the airplane and its functions. Information about how the autothrottle system works, and how to conduct a dual-autopilot approach is distributed across a number of documents, namely:

- The Boeing 737 Flight Crew Operations Manual (FCOM), Volumes 1 and 2;
- The Boeing 737 Flight Crew Training Manual (FCTM);
- The airline's Operations Manual (OM);
- Various materials associated with the airline's low visibility (CATII/III) training.

This echoes a concern offered by the authors of the FAA report, namely that:

> "Most of this information may be available in current training and operating manuals; however, it is typically scattered throughout several volumes and may not be emphasized to the extent necessary for flight crews to grasp its practical significance. Current qualification programs may cover this material to some extent, but it is generally not emphasized to the extent the HF Team considers necessary, nor is it integrated with training, simulator, or Line Oriented Flight Training (LOFT) scenarios." (FAA, 1996, p. 38)

The basis for both concerns expressed here can be recognized in the organization and content of knowledge as offered in the Flight Crew Operations Manuals and other materials: (1) the information is scattered and maybe not emphasized for crews to grasp its practical significance, and (2) this information lacks integration with actual training scenarios. In addition, the information as presented in these manuals is shallow in a technical sense, a natural by-product of both the increasing complexity and computerization of systems inside of modern airliners. Only a superficial characterization of systems and their functions can be found: natural-language algorithms of mode changes or reversions are not available, nor are the linkages between the various computer systems and their sensor parameterization in various phases of flight made available.

If we take, however, and as much as possible, a top-down approach as would likely be followed in a learner's build-up of a mental model, the starting point could be FCOM2, chapter 4, which describes the overall flight management system (chapter 4):

> "The automatic flight system (AFS) consists of the autopilot flight director system (AFDS) and the autothrottle (A/T). The flight management computer (FMC) provides N1 limits and target N1 for the A/T and command airspeeds for the A/T and AFDS… AFS mode status is displayed on the flight mode annunciation on each pilot's primary display."

This lays out how the automatic flight system consists of two different systems (the AFDS and A/T). It does not say how exactly these two systems collaborate to achieve flight path goals during different phases of flight or to what extent these two different systems rely on different sensor inputs to achieve goals in their various modes. The AFDS itself is explained as a dual system that consists of two individual flight control computers (FCC's) which send control commands to their respective pitch and roll servos which operate the flight controls through two separate hydraulic control systems.

In particular, the FCOM2 does not mention specifically how the two systems (AFDS and A/T) collaborate in accomplishing an autoland, or whether they in fact always do so. Under the "Approach (APP) switch" description, the AFDS is responsible for localizer and glideslope capture and tracking (FCOM2 chapter 4). In that same place, there is no mention of what the A/T is responsible for during an approach in APP mode and where the A/T gets its sensor data, or what may happen if the two systems that make up the AFS are pursuing incompatible goals (the AFDS trying to remain on glide slope, the A/T reducing

speed, as in TK1951). A later description of an approach in APP mode (with dual autopilots as was the intention of the TK1951 pilots) then says:

> "**Approach (APP) Mode Dual A/Ps**
> Approach mode allows both A/Ps to be engaged at the same time. Dual A/P operation provides fail–passive operation through landing flare and touchdown or an automatic go–around. During fail passive operation, the flight controls respond to the A/P commanding the lesser control movement. If a failure occurs in one A/P, the failed channel is counteracted by the second channel such that both A/Ps disconnect with minimal airplane maneuvering and with aural and visual warnings to the pilot.
> One VHF NAV receiver must be tuned to an ILS frequency before the approach mode can be selected. For a dual A/P approach, the second VHF NAV receiver must be tuned to the ILS frequency and the corresponding A/P engaged in CMD prior to 800 feet RA." (FCOM2, chapter 4)

The manual makes clear that the APP mode fail passive operation ensures that both autopilots will disconnect when one of them fails during the approach, handing the airplane back to the pilots with both "aural and visual warnings." There is no mention of what the A/T does in such a case, or whether it is similarly protected though fail-passive logic.

Significantly, during the approach of TK1951, the failure of the left RA was not protected through a cross-validation with the right RA, the A/T did not disconnect as a result, but rather remained engaged and then stayed in RETARD mode, and there was no aural or visual warning of the RA failure (a failed RA is indicated through the yellow letters "RA" presented steady on the pilot's PFD where otherwise the altitude numbers would be—there are no aural warnings or other visual indications of an RA failure. Onboard TK1951, there was no yellow "RA" on the left side, only a normally colored digit that was inconsistent with the phase of flight; the right side PFD (of the co-pilot, who was flying) was showing the actual radio altitude of the airplane above the ground).

*Autothrottle mode changes*

FCOM2, chapter 4, offers the manuals' description of the autothrottle system:

> "**Autothrottle System**
> The A/T system provides automatic thrust control from the start of takeoff through climb, cruise, descent, approach and go–around or landing. In

normal operation, the FMC provides the A/T system with N1 limit values. The A/T moves the thrust levers with a separate servo motor on each thrust lever. Following manual positioning, the A/T may reposition the thrust levers to comply with computed thrust requirements except while in the THR HLD and ARM modes.

The A/T system operates properly with the EECs ON or in ALTN. In either case, the A/T uses the FMC N1 limits. During A/T operation, it is recommended that both EECs be ON or both be in ALTN, as this produces minimum thrust lever separation."

During the approach of TK1951, the autothrottle reverted to RETARD mode shortly after the crew selected V/S mode. The aircraft behavior that this triggered was entirely consistent with their intentions and expectations. The possible autothrottle modes are listed in FCOM2 (chapter 4):

"**Autothrottle Modes**
• N1 – the autothrottle maintains thrust at the selected N1 limit displayed on the thrust mode display, including full go-around N1 limit
• GA – the autothrottle maintains thrust at reduced go–around setting
• RETARD – displayed while autothrottle moves thrust levers to the aft stop. RETARD mode is followed by ARM mode
• FMC SPD – the autothrottle maintains speed commanded by the FMC. The autothrottle is limited to the N1 value shown on the thrust mode display
• MCP SPD – the autothrottle maintains speed set in the MCP IAS/MACH display. The autothrottle is limited to the N1 value shown on the thrust mode display
• THR HLD – the thrust lever autothrottle servos are inhibited; the pilot can set the thrust levers manually
• ARM – no autothrottle mode engaged. The thrust lever autothrottle servos are inhibited; the pilot can set thrust levers manually. Minimum speed protection is provided."

According to this listing, RETARD mode is followed by ARM mode, and in ARM mode, minimum speed protection is provided. In TK1951, RETARD mode was not followed by ARM mode (since the flight was completed as far as the autothrottle algorithm was concerned), and no minimum speed protection was provided. The sequence of mode events during TK1951 thus matched roughly what a crew would expect after selecting a lower speed and lower altitude on the MCP—the behavior matched enough and the displays showed little enough (about the automation's real intentions) so that no suspicion whatsoever was raised about anything being awry. The description of A/T modes (and the

RETARD mode in particular) in the FCOM2 hardly offers a basis for understanding, let alone recognizing in context, the subtle difference that confronted the crew during the approach of TK1951.

Note that there is no mention in the Flight Crew Operations Manual of the sensor inputs that the autothrottle uses for its mode changes during an approach to landing, or on which FCC it depends. In particular, there is no mention of the role of RA in providing inputs for the A/T algorithm in certain modes. And no mention that this happens via FCC A only.

FCOM2, chapter 4, says how the autothrottle (A/T) holds speed when the AFDS is engaged in two modes relevant for the final moments of TK1951 (V/S and G/S capture):

> "Speed Mode: Autothrottle holds speed in IAS/MACH display or a performance or limit speed… Speed mode engages automatically when:
> …
> • V/S engages
> • G/S capture occurs."

On the basis of this, a pilot can expect the authothrottle aiming to hold the speed displayed on the Mode Control Panel when the autopilot is flying in V/S mode or G/S capture mode. It is not uncommon, however, for the AFS to have some trouble achieving or maintaining the target speed, particularly in V/S in a descent. This is evident, for example, from the TK1951 flight data readout which shows an increase in speed (relative to the target set in the Mode Control Panel) as V/S mode is engaged and the airplane noses over to descend (in an aim to capture the glide slope for which the crew had been vectored late). There is no commonly known parallel of the autothrottle having difficulty achieving or maintaining target speed with the autopilot in G/S capture mode.

It is, of course, an open question whether this sort of guidance is sufficiently "clear and concise" to live up to the FAA's 1996 recommendation to:

> "require operators' manuals and initial/recurrent qualification programs to provide clear and concise guidance on:
> …
> The conditions under which the autopilot or autothrottle will or will not engage, will disengage, or will revert to another mode." (FAA, 1996, p. 38).

In any case, there is no documentation or training that says that the A/T does *not* revert to another mode (e.g. MCP SPEED mode after having been in

RETARD mode) as a result of an RA problem (in fact, having an RA problem on only one side determine what the A/T will or will not do is counterintuitive and contradicts the presence and common awareness of multiple layers of redundancy in most other aircraft systems (including the AFDS)). In its post-accident bulletin, Boeing offers that:

> "On some 737s, the autothrottle logic uses left radio altimeter data regardless of the autopilot selected" (Boeing Flight Operations Technical Bulletin, #737-09-2, March 19, 2009).

This information was not readily available before the accident, and is not in the FCOM or other materials available in initial or recurrent pilot training. Even if it was, pilots can impossibly know whether they happen to be flying a 737 on which this is or is not the case (which it is only on "some" 737's). As a result, from how they were prepared for their jobs, the crew could not possibly have made the link between a left side radar altimeter problem (that was *not* annunciated as a failure) and A/T going and staying in RETARD mode.


## Alerts and indications associated with RA failure

Even if it would be possible for pilots to establish a relationship between a left RA problem and the A/T remaining in RETARD mode, the A/T problem was not annunciated as a problem or a failure that could have alerted the crew in that direction, nor did the A/T disengage as a result of the faulty sensor input. As a result, there was no salient, primary trigger that alerted the crew to the nature of the problem or offered any suggestions of what to do about it.

The 737 Flight Crew Operations Manual Volume 2 states the following about Radio Altitude as it relates to the events onboard TK1951 chapter 10):

> "Radio Altitude
> The current radio altitude is displayed in the bottom center of the attitude indication area when radio altitude is below 2,500 feet AGL. The display turns amber when the radio altimeter is below the radio altitude mimimums."

The Boeing 737 has an ALT DISAGREE alert but that only works for the barometric altimeters, not the radio altimeters (RA). An RA disagreement (i.e. the left and the right side RA getting different sensor readings) is not annunciated as an alert, nor does it generate any other primary consequences (e.g. no fault indications and no disengagement of the A/T).

In FCOM2 chapter 4, it says how the A/T will disconnect if a fault is detected:

> "Any of the following conditions or actions disengages the A/T:
> ...
> • an A/T system fault is detected" (FCOM2, chapter 4).

In the case of TK1951, however, the A/T did not disengage, thus not communicating to the flight crew that there may have been a failure of some sort underlying or misinforming the A/T. Also, this continued engagement of the A/T set the crew up for problems with the recovery maneuver later on.

In TK1951, the landing gear warning horn during various times before the crew began with the approach. At that time, the landing gear was still up and no landing was attempted (as TK1951 was still far from the field and thousands of feet in the air), so the landing gear warning horn would not have made much sense. In fact, the FCOM2 does not relate the landing gear warning horn to the RA (or, for that matter, the left RA). The warning could not have made much sense (The manuals says about it that "a steady warning horn is provided to alert the flight crew whenever a landing is attempted and any gear is not down and locked" (FCOM2, chapter 15) in the context in which it sounded (as no landing was being attempted). That the warning is a secondary and spurious by-product of a faulty sensor input of another system (the RA) is not described in the manuals and becomes very difficult to hold up as a persuasive indication that something is wrong with the RA. The FCOM2 says the following about the landing gear warning horn:

> "Aural Indications
> A steady warning horn is provided to alert the flight crew whenever a landing is attempted and any gear is not down and locked. The landing gear warning horn is activated by forward thrust lever and flap position as follows:
> Flaps up through 10 –
> • altitude below 800 feet RA, when either forward thrust lever set between idle and approximately 20 degrees thrust lever angle or an engine not operating and the other thrust lever less than 34 degrees. The landing gear warning horn can be silenced (reset) with the landing gear warning HORN CUTOUT switch
> • if the airplane descends below 200 feet RA, the warning horn cannot be silenced by the warning HORN CUTOUT switch.
> Flaps 15 through 25 –
> • either forward thrust lever set below approximately 20 degrees or an engine not running, and the other thrust lever less than 34 degrees; the

landing gear warning horn cannot be silenced with the landing gear warning HORN CUTOUT switch.

Flaps greater than 25 –

• regardless of forward thrust lever position; the landing gear horn cannot be silenced with the landing gear warning HORN CUTOUT switch. The warning indication is cancelled when the configuration error is corrected." (FCOM2, chapter 15)

Aural warnings are designed with a particular context in mind (hence the various considerations of flap settings above). A spurious warning that sounds in a context in which it makes no sense relative to its original purpose (the landing gear warning horn, after all, sounds when a landing is attempted), cannot be assumed to help a normally trained crew (whose manuals do not suggest a link either) in diagnosing a subtle problem that has occurred in another system. In fact, such diagnostic displacement (a failure in one system being channeled through a spurious alert of another) has been a perennial problem in the design of warning systems, and one of the drivers behind the call for more intelligent approaches to fault diagnosis in transport aircraft (see Billings, 1996).

To the extent that the landing gear warning horn could be taken as an indication of other problems, like those with the RA (which, again, is a link that is virtually impossible to make on the basis of training and available written guidance), the same written guidance would have made concern about that problem disappear once the landing gear was selected down and locked, since the "landing gear warning horn is deactivated with all gear down and locked," (FCOM2, chapter 14) Once the landing gear was down and locked on TK1951 and the horn no longer sounded, the crew can only have assumed that whatever the horn was trying to tell them (that they were attempting a landing with an unsafe landing gear) had no more relevance or had been resolved from that point on.

Under "Flight Director Display" (FCOM2, Chapter 4) it does not say that the F/D bars will disappear from the PFD as a result of an RA failure. There is nothing in the manual, in other words, that links the disappearance of the F/D bars with RA problems; one cannot be inferred from the other on the basis of training or available written guidance. In fact, the manual says that the F/D can be used with or without A/P and/or A/T, which confirms (and indeed this is usually shown in training, either in the simulator or on the line) that the disappearance of the F/D bars has no influence on the functioning of either A/T or A/P. Both parts of the AFS can and will operate normally without the F/D bars showing on the PFD. In the mental model that a pilot will likely build up, then, the disappearance of the F/D bars says nothing about the functioning of malfunctioning of the A/T or the quality of inputs it receives.

Finally, the stabilizer on the B737 is automatically trimmed an additional amount nose-up at 400 feet Radio Altitude during an automatic approach to ensure adequate missed approach performance should it be needed. This information is available in the manuals and typically shown and rehearsed in pilot low visibility training. There was a significant amount of nose-up trimming during the latter stages of the final approach of TK1951. While this trimming turned out to be related to the airspeed decay due to A/T RETARD mode, nose-up trimming could have been seen as not anomalous given the crew's earlier experiences of dual autopilot approaches down to approach minima. That such trimming occurred may also have served as an indication to the crew that the RA was working properly.

To be sure, TK1951 was not a case of RA "failure," as far as the indications available to the flight crew were concerned. No yellow "RA" annunciation was made, and the RA readings on either side, while different, were both coded as if they were valid readings. No A/T disengagement occurred. No altitude disagreement flags or alerts were presented. This leaves no human factors basis to conclude that the indications available to the crew amounted to a persuasive rendering of the exact nature of the problem they faced. This is confirmed by the responses of the flight crews of the earlier flights on this aircraft where the RA problem appeared. While those crews responded differently (or, for that matter, earlier e.g. in taking over manually) there is no evidence of a confirmed diagnosis of what was wrong by them either and no anomaly reporting or other follow-up done by those crews or THY from which such a confirmed diagnosis could be inferred.

Indeed, the pilot of one of those prior flights in which the same RA problem showed up and was responded to by disengaging both A/T and A/P, decided not to write up or report the problem because there was never a warning flag or other anomaly indication associated with the RA problem. The RA was indicating an altitude without any flags or warnings, thus not suggesting a failure or prompting a report or write-up.


## Comparison with other TRTO

The lack, dispersal and disintegration of information about the functioning or malfunctioning of critical systems prompted the FAA team in 1996 to recommend:

> "…consolidating this information into clear and concise guidance to promote better flight crew understanding of the capabilities and

limitations of the automation, and, to the extent necessary, incorporating practical demonstrations of its use into training and checking scenarios or events." (FAA, 1996, p. 38)

The events onboard TK1951 suggest that this has gone unheeded to date. To calibrate this finding against industry standards and practices, a JAA-approved Type Rating Training Organization (TRTO) program for the Boeing 737NG, that was active in training pilots during the same time that the TK1951 co-pilot was type-trained on the 737, was studied for comparison. The focus of this study was on the preparation of pilots for the use of the 737NG's automation, particularly in approach and landing. The type rating training for the 737NG conducted by this TRTO had also been approved for training CATII/III approaches (for which dual autopilots must be used).

The CATII/III training handout to type-rating students in this TRTO stated that an autoland requires, among other systems, two dual channel autopilots engaged as well as a low range radio altimeter and -display for each pilot (this information can, under certain circumstances and regulatory regimes, be found in the minimum equipment list, or MEL, for the aircraft). The training handout also stated that decision height for CATII/III approaches is defined as the distance between the wheels of the airplane on the glideslope and the highest elevation of the runway in the touch down zone, a height that is read on the RADIO ALTIMETER (capitals in original). Importantly, the handout stated nothing about the advisability or wisdom of keeping hands on the thrust levers and the control wheel during an automatically flown approach.

The following are sample questions taken from the systems exam that follows the so-called "ground school," a class-room training schedule that takes ten days and consists mostly of self-guided study with the help of a CBT (Computer-based training) program. Students undergoing CBT typically receive a laptop from the TRTO (or find banks of computers in a lecture room) with a partially interactive program that takes the student on his or her own pace through the systems of the Boeing 737. An instructor is available for any questions that may arise. The total number of questions on the systems exam, which concludes these ten days, is sixty. All questions are multiple-choice. The entire question bank is given to students in advance (at the beginning of ground school). The questions offer a choice of two to five answer alternatives of which one is correct (the correct answers are given to the students in advance as well). The answer alternatives are not listed below.

Under the section "Flight instruments, displays," the following question was found:

- During an ILS approach, the Captain observes that his radio altitude display turns yellow and begins flashing. What does this mean?

Interestingly, it points pilots to a failure of the RA, which prompts the flashing and yellow indication. There was no such failure in the case of TK1951 and no such indications.

Under the section "Automatic Flight," the following four questions related to the autoland function were found:

- What is the minimum altitude (above ground level) for selecting the second autopilot during an ILS approach?
- During a dual autopilot ILS approach, at which altitude (above ground level) should the FLARE mode engage?
- Which mode must be armed before the second autopilot can be selected?
- What is the maximum and minimum glide slope angle for autoland?

A question relevant to the sequence of events of TK1951 could be found under the section "Flight management":

- What happens when the first officer's radio altimeter is inoperative?
a) Autopilot channel A should not be used for the approach
b) Autopilot channel B should not be used for the approach
c) All modes of the GPWS are inoperative
d) Autothrottle automatic retard during landing flare is inoperative

Alternative b) is the correct answer, but, interestingly, it uses the word "should" rather than something like "cannot" or "is not available." The word "should" leaves the factual technical status of Autopilot channel B during a right side RA failure ambiguous: perhaps it can be used, but should not (because of some underlying unreliability maybe), or perhaps it really cannot be used, so therefore it should not be used. Interestingly too, the question is inclusive when it comes to the non-functioning of the radio altimeter, as it uses the phrase "inoperative" rather than "failed." An inoperative RA could be one that has not failed (and is not indicating as such) but has a sensor problem, as was the case with TK1951 and is indicating a number like it would any other. On the other hand, inoperative could mean "failed," and indicated as such on the F/O's PFD (through a yellow, boxed RA annunciation). That the question does not disambiguate this means that a newly type-rated B737 pilot does not have, on the basis of the question and the documentation underlying it, the preparation to see through the nuance of the problem that faced the crew of TK1951.

Under the section "Warning systems," the following question related to the landing gear warning horn was found:

- The landing gear warning horn will activate anytime a gear is not down and locked with the flaps set to 15, one thrust lever at idle and the other at a high power setting (above 34 degrees): True/False.

The approved training program examined for comparison here would not have been able to adequately prepare a 737 pilot with the knowledge, understanding and skills necessary to effectively diagnose and recover from the problem faced onboard TK1951. A sample of two TRTO/Airline training programs does not necessarily represent the industry. But given that both organizations meet international regulatory demands and standards, there is compelling reason to believe that they are in fact representative.

Another aspect stands out from this comparison of THY type training for the Boeing 737 with this TRTO here. The length of type training on the B737 at THY significantly exceeds that of the comparison TRTO, and would seem to exceed industry standards. More on this can be found in the section "CRM and the Intervention decision" in this report.

## Experience on aircraft type and buggy mental models

Whereas the issues relevant to a type rating and the associated systems exam may have had a direct bearing on what the F/O of TK1951 (and any F/O like him) may have known about the automation, what about the Captain of TK1951? The Captain had flown the B737 since 1996 and had accumulated 10,885 flight hours on the type at the time of the accident, of which more than 3,000 as commander. Paradoxically, experience with a particular airplane type may not necessarily increase the sophistication of the mental model that pilots use to operate that airplane. In fact, research has shown that any misconceptions in that mental model, if shown to have been effective for most operational circumstances, can become even more robust and resistant to change (Feltovich *et al.*, 1989).

NASA researchers have identified two mechanisms in particular that may not have enabled a Captain like the one on TK1951 to have developed a mental model that accommodated the subtle failure his crew faced (Berman & Reed, 2009). The first is that having a lot of uninterrupted experience on one type tends to pull a pilot further and further away from the technical training that went into preparing him or her for that type. That is, it will be longer and longer ago that the pilot in fact underwent a technical or systems exam for the airplane.

Other than the twice yearly proficiency checks and a yearly line-check, there are few opportunities to become up to date (again) with the complexities and subtleties of the aircraft's systems (to the extent that the available documentation allows a pilot to become up to date, which it doesn't—see above). The result, NASA research suggests, is that there are particular risks associated with what could be called "homesteading" on one particular aircraft type only (Berman & Reed, 2009).

The second mechanism that may inhibit the development of an appropriate model for anomaly response with experience, is that anomaly response on the Boeing 737 (the sort of anomaly response typically practiced during the proficiency checks, and, if ever, the ones that are typically experienced on the line) is largely, what NASA researchers call, "light-driven" (Berman & Reed, 2009). The various warning lights distributed through the cockpit (another hold-over from the pressure to keep as much commonality across various generations of B737 cockpits) are all named in a book, the Quick Reference Handbook (QRH). This book, handily available on the flight deck, is the source to go to when a light comes on, and it will spell out the right steps of the anomaly response (how to deal with whatever the light may be suggesting). The light-driven anomaly response model of the airplane, according to the NASA research, may make it much more difficult for pilots with a lot of experience on the type, to appreciate the possible existence of (subtle) failure sources that are not represented by lights in the cockpit. The RA malfunction that occurred on TK1951 is one such subtle malfunction. It has no light, and no corresponding response in the QRH.

Finally, the scenario as it played out on board of TK1951, and especially the automation surprise it produced, is rare. It would have been unlikely for this Captain to have experienced this subtle malfunction or sequence of events before. And indeed, as was the response of his colleagues on flights on this B737 prior to TK1951, this particular malfunction may not have been interpreted as a malfunction or failure, as there were no warning- or failure flags associated with the behavior of the RA.

# SPEED, MODE MONITORING AND THE NOTICING OF A NON-EVENT

A key question in the last part of the sequence of events concerns the apparent difficulty of recognizing that the automation was not doing what it was supposed to do, including producing an airspeed decay below the target set on the MCP. In pursuing answers here, it is important to remember that the A/T going to idle and speed decaying was expected by the crew because of the late turn-in for final and, as a result, their having to capture the glideslope from above. Of 100 seconds of speed decay, the first 70 seconds were consistent with this expectation. And of these 70 seconds of speed decay, the latter 20 furthermore match the intention of the crew to have the aircraft fly the new flap 40 target speed, which indeed involves a further (expected) deceleration. Only the last 30 seconds of the 100 seconds would have required the crew to notice two things that were *not* happening (a stopping of the slowing down that didn't happen, and a mode change that didn't happen). Research has shown that noticing non-events is very difficult. The two phases of the 100 seconds of speed decay are detailed under the heading "How to see that something doesn't happen" below.

This should be understood against the background of considerable human factors research about the inadequacy of having humans monitor automated systems. Naturally, Boeing prescribes that, as part of the Normal Procedures associated with using the Autopilot Flight Director System (AFDS), the crew must always monitor: airplane course, vertical path and speed. This requirement for crew monitoring of course rows against the stream of human performance research. Collectively, this research strongly points to the human ineffectuality in monitoring automated processes (e.g. Moray & Rotenberg, 1989; Mathews, Davies. & Holley, 1993; Sheridan, 1995; Billings, 1996; FAA, 1996; Mayer, 2000). Indeed, research (from all kinds of fields) points to the problems of maintaining the same level of focus when automation is in control of the process (e.g. Byrd, Adams & Ntuen, 2002). Humans are not good monitors of automated processes (Bainbridge, 1983).

Despite manufacturer reminders to continually look at the machine (and, essentially, mistrust it), "it has become evident that humans, when put in the role of monitor, supervisor, and automation backup in case of failure, may not perform well" (Sheridan, 1995, p. 823). This is particularly the case when the machine, over long periods of operating with it, shows itself to be trustworthy. It is hard to ask a human to mistrust a machine (and keep staring at it intently) that has so far given no or little cause for mistrust (Matthews et al., 1993; Billings, 1996). This is very much true for more modern systems that are normally very reliable. A research effort triggered by the UK CAA (now driven in part by EASA concluded that newer automated system that are normally very reliable lead to poor detection of automation failures—poorer than with older, less reliable systems, and there is no real effect that training has on this:

> Detection of automation failures was poor under constant reliability automation, even following a catastrophic failure. However, monitoring was efficient under variable reliability automation. These effects do not significantly alter following training (CAA, 2004, p. 1).

## How to see that something doesn't happen

*The first 70 seconds of speed decay*

For the first full 70 seconds of those 100 seconds of speed decay, the behavior of the aircraft was entirely consistent with crew instructions and expectations. The crew would have expected their speed to decay since that is what they instructed their aircraft to do (because of the new flap 40 target speed set on the MCP). The crew would also have fully expected that thrust levers were going to idle because they had been forced to use V/S mode to capture the glideslope from above due to the tight turn-in for final. Using V/S to help capture a glideslope from above almost always leads, and also in this case led, to a momentary increase in airspeed.

Aircraft such as the 737-800 are known to have difficulty "going down and slowing down" simultaneously. While trying to descend on a glideslope, a speed reduction from the target of 160, itself overshot by 14 knots due to the need to use V/S mode (so in reality the speed had to come down from 174), to the new target of 143, can only be achieved by idle power. The A/T going into a mode that would idle the throttle handles was thus fully expected. It was consistent with the path of flight given to TK1951 by ATC and the concomitant crew intentions. During the last 20 seconds of these first 70 seconds, the crew

selected a new target speed associated with flaps 40. The required (and produced) a further (expected) deceleration.

*The remaining 30 seconds of speed decay*

For the remaining 30 seconds of those 100 seconds of speed decay, the crew would have had to notice two non-events, or, in other words, notice that two things were *not* happening. The first thing that was not happening was the stopping of slowing down. The second thing that was not happening was a mode reversion of the A/T back to ARM. The speed kept doing what it had been doing for the first 70 seconds of the 100 seconds, as did the A/T: its mode remained the same (and the A/T remained in idle, also what they had been doing since the beginning of those 100 seconds in which idling was consistent with initial expectations).

Studies on the monitoring of dynamic processes have shown that it is very difficult for people (as well as machines) to notice non-events (Chow *et al.*, 2000; Patterson & Woods, 2001; Christoffersen *et al.*, 2003). Things that do not happen are not meaningful or informative phenomena in the monitoring of dynamic processes. Something that is *not* happening is not a good trigger for human (or even machine) intervention. These studies show that non-events, that is, the lack of change over time (or even the lack of change in the derivative, i.e. lack of change in a change) are associated with difficulty for practitioners to detect meaningful phenomena in their monitored process. Evolutionarily, change is information; lack of change has little information or adaptive value. The sensory and neural systems of most organisms, including mammals (which includes humans) are highly attuned to pick up and respond to change, while reserving cognitive resources for other tasks when there is no noticeable change in the environment (e.g. Gibson, 1979; Mathews, Davies & Holley, 1993; Heft, 2001).

The loss of a sensor or a sensor failure in particular presents problems to monitoring because the automated system typically does not send persuasive signals about the failure to the human monitor, or may even send misleading signals (Rohloff, 2005). Such research findings apply to TK1951 which involved a sensor failure that was not accompanied by a persuasive signal to the pilots. Indeed, this research has called for the design of sensor failure-tolerant supervisory control, something that, in case of RA sensor problems, the B737 does not provide.

From their training, written guidance, and the design of the FMA (Flight Mode Annunciator) the crew could not have surmised that the throttles remained in

idle because of an RA problem. The A/T had gone into idle in order to achieve a speed reduction during descent onto the glideslope, but there was no basis or indication for the crew to know that the A/T *stayed* at idle because of a reason different from the one that *got* it into idle (and, again, a reason they would have had no chance to know on the basis of training, written guidance or design). The visible effects, i.e. an A/T mode annunciation (RETARD), thrust lever position, and a deceleration stayed the same throughout the 100 seconds.

## Not noticing a mode change

It would be consistent with the considerable human factors literature on the topic if the TK1951 crew did not notice the mode annunciation on the FMA at the time V/S was selected, or did not notice it on the FMA during the seconds after, or would not have understood the implications of the annunciated A/T mode if they had (Degani, Shafto, & Kirlik, 1995; Dekker, 2000; Dornheim, 1995; Funk, Lyall, & Niemczyk, 1997; Lyall, 1998; Nikolic & Sarter, 1999; Sarter, 1995; 1997; Vakil, Hansman, Midkiff, & Vaneck, 1995; Woods & Sarter, 2000). A number of sources who studied this problem in detail (Huettig *et al.*, 1999; Mumaw, Sarter, & Wickens, 2001) discovered that pilots actually do not look at the FMA very often, which could mean that the FMA does not present information convincingly to pilots, even though manufacturers stress that it is the only reliable source for current and expected automation mode information (Huettig *et al.*, 1999).

Even if mode events are annunciated (through the box around the mode label), research results indicate that the design of the current generation of FMA panels support human supervisory control very poorly (Sarter & Woods, 1994; Corwin, 1995). A study by Mumaw *et al.* (2001) showed that pilots did not look at the FMA during 53% of the manually induced transitions, 45% of automation-induced transitions that were expected by the pilots, and 62% of automation-induced transitions that were not expected by them. Up to 10 sec after the box had disappeared (i.e., 20 sec after the mode transition), 32% (and 29% and 40%, respectively) of the mode transitions announced on the FMA had still not been looked at (Mumaw *et al.*, 2001).

Another study (Björklund *et al.*, 2006), conducted during 12 flights with eye-point-of-gaze trackers on both crew members concluded that for all 512 mode transitions, the official airline procedure was followed only 29 times. In other words, the study recorded 483 "procedure violations" with respect to the automation alone, and that within 12 hr of flying, or more than 40 automation-related procedure violations per hour. Rather than following official protocol, crews may look at the FMA and not say anything, or they may say something

but not look at the FMA. Doing both, and in the right order, with a visual verification before a call-out, appeared rare.
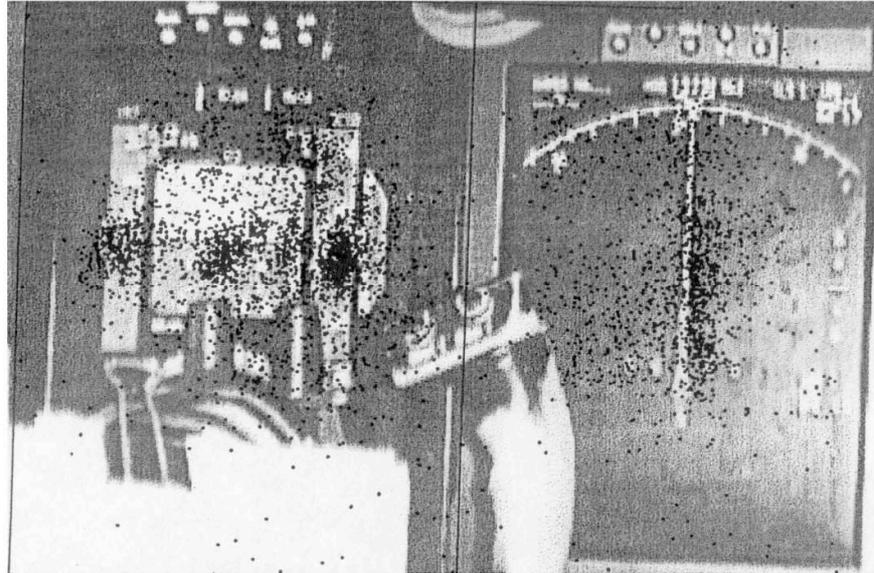


*Eye-point-of-gaze registration superimposed on the primary flight display (the display to the left) and of the navigation display for the duration of a one-hour flight, including the approach. The FMA is located on top of the primary flight display and shows virtually no eye fixes (from: Björklund et al., 2006).*

This study too, confirmed that the FMA as designed is not a very useful basis for building or maintaining mode awareness. Two out of five mode transitions on the FMA were never "seen" by the flight crews in this study. In contrast to instrument monitoring in non-glass-cockpit aircraft, monitoring for mode transitions is likely to be based more on a pilot's mental model of the automation that drives expectations of where and when to look. Such models are often incomplete and buggy (e.g., Sarter, 1995). Therefore, it may not be surprising that many mode transitions in this study were neither visually nor verbally verified by flight crews, and that the FMA triggered only 4% of call-outs in this study, of which one out of four was not even the official call-out. The FMA did not get consulted for 40% of all mode transitions.

Crucial for the understanding of TK1951, the flight crews in this study did not seem to consistently check the FMA when changing modes manually. Indeed, the mode control panel, where changes are actually made and selected settings (altitude, speed, etc.) can be seen, seemed to be used as a more dominant visual resource for knowledge about what the aircraft was going to do (despite manufacturer cautions against this) (see Fox *et al.*, 1996).

Another finding of this study, even though its call-out rates were too low for statistical analysis, is on the effect of workload on call-out frequency. Of all crews and flights studied, only one copilot made a mode call-out during the missed approach procedure. All others and all captains remained silent with respect to mode status, not even using informal communication to guide each other's attention to current or future mode status. This would confirm that in higher workload situations, automation call-outs could be among the tasks to be put aside, another factor of relevance for TK1951 at the moment of realization that they had been vectored too high for the glideslope and had to intervene to capture it.

During the approach of TK1951 from 4000 feet on down, eight mode changes (including the arming rather than engagement of certain modes) occurred that would have been annunciated on the FMA on both sides. Three of those changes were automatic, five were the result of immediately preceding crew inputs (made on the MCP). One of those mode changes was called out by saying "localizer capture." Three others were accompanied by verbalizations or call-outs associated with the mode selection or change (e.g. "courses active" when arming VOR/LOC mode, or "approach selected" when arming G/S, or verbalizing and double-checking the actual heading selected ("2-6-5") when going into HDG SEL mode). This is consistent with previous research (Björklund *et al.*, 2006): mode changes are accompanied by a variety of call-outs and verbalizations that support crew coordination around them. At THY, the standard operating procedure for mode callouts actually does not require the callout, but, consistent with Boeing guidance on the topic, requires crews to "check" the modes. Such checking may involve a verbal announcement, but that does not have to be the case.

When it comes to the relationship between workload and mode callouts, another recent study (Goteman & Dekker, 2007) showed how callouts relating to aircraft automation, such as FMA call-outs, are shed in higher workload situations before other callouts. During the 19 flights of this study, the average number of mode callouts that was shed fell between four and five (15%) per flight got shed. In higher workload situations, as rated by observing domain experts, callout shedding increases dramatically. In those situations (like on approach), 21 out of 40 (53%) of the occurring mode changes were not called out. The shedding rate was 17% for vertical mode callouts, 19% for lateral mode callouts and 6% for autothrottle mode callouts. Callouts that were shed were not recalled or saved for later verbal announcement. During such periods of high task load, other verbal coordination, especially that with Air Traffic Control, does not suffer.

In the case of THY, at the moment that the RETARD mode was annunciated on the FMA, (most likely) the Captain had just selected V/S mode and was dialing in a new altitude on the MCP, then dialing the desired rate of descent into the MCP and very likely following this up with a visual check of the VSI (Vertical Speed Indicator) on the PFD (which is on the opposite side of the screen from the left top where the A/T mode announcement is made), as well as the purple diamond representing the glideslope and matching and adjusting the vertical speed with the up-movement of that glideslope indicator so as to capture the glideslope from above. The F/O may likely have followed the Captain's rapid hand movements across the MCP at this time, so as to try to keep up with what was going on and understand the Captain's intentions (no verbal announcement of those intentions was made, there usually is no time or cognitive resource for that in a situation like this).

Research into the FMA, mode callouts and mode awareness suggests that mode callouts about the automation are perhaps seen as a secondary task, while verbal coordination with, or about, other human partners in the system (e.g. the controller, or coordinating something about the flight path or automation setting with the other pilot without specifically mentioning the mode annunciation by label) are deemed central, or primary to the conduct of the flight. The results from that study suggest that FMA callouts are not used primarily as a tool to detect or remember automation mode changes, but rather as an opportunity for coordinating aspects of the flight between the pilots themselves.

## Automation surprises and representations of the future

The problematic nature of visual FMA verification and mode change announcements is due in part to the work that people are doing when mode changes occur. It is related to where automation or other system selections are made, who makes them (or whether they are done automatically), and what other workload demands surround the moment of mode change (e.g. monitoring other flight parameters or instruments) (see Nikolic & Sarter, 1999). For TK1951, at the moment of selecting V/S mode, so as to capture the glideslope, it may not have been likely that either crewmember would have seen the RETARD mode annunciation on the left top of the FMA because of workload and visual attention demands as well as the coordination of the rate of V/S against the up-movement of the glideslope diamond on the PFD (see above). However, if either or both of the crewmembers had seen the RETARD annunciation, had they understood its implications for what was going to happen next?

For more than the next minute, the aircraft was doing what the crew (thought it) had told it to do: retard the throttles, slow down and descend. As explained in the section on automation surprises above, crews usually judge an aircraft by its behavior to see whether their intentions or instructions are being followed—not the FMA or other annunciations. Here, aircraft behavior matched instructions, for over a minute. The only indications, during those 70 seconds that were, in hindsight, pointing in another direction were:

- The left RA indication that did not match the right RA indication (the left one being inconsistent with actual flight height at that moment. This may have been judged, on the basis of knowledge available to the crew and Boeing training materials, as non-problematic since the right FCC was giving all auto-flight commands.
- The annunciation of RETARD on the FMA, instead of another mode.

The latter representation of automation mode status (the word "RETARD" on an FMA), is known to be the least developed form of visual reference, and associated with a host of problems in human monitoring and human performance. It is called a "propositional reference" (Woods, 1994). In this type of reference, there is an arbitrary relationship between token and referent (e.g., words). The token is the thing or word presented on the screen ("RETARD"), the referent is that which that word or token refers to (in this case an autothrottle mode; in reality a mode that in turn represents or stands for a range of possible behaviors too, depending on circumstances). This technique, of propositional referencing, relies on description. The token functions referentially to "tell" the observer something about the state of the domain referent (I am in RETARD mode). It is a mere descriptor (Woods, 1994).

There is one advantage of propositional reference, which is that it takes up very little cockpit real estate. But the problems with propositional reference are severe. First, the relationship between token and referent can only be understood only if there is prior stored knowledge about that relationship, and if that knowledge can be brought to bear in context so as to make sense of the annunciation there. In the case of TK1951, that knowledge (of what the RETARD mode meant at that stage of a flight) was likely not available to the flight crew through previous training or experience or Boeing materials or documentation. Even if it had been, bringing such (incomplete and buggy) knowledge to bear in a rapidly unfolding sequence of events as a result of the tight turn-in onto final approach would have been all but impossible (see Woods *et al.*, 2009).

Second, propositional references require reading, and reading requires focal attention. People won't see what the referent in the automated system is, or is

doing, without focusing conscious attention on the descriptor or token, reading it out, and matching it with prior knowledge (again, not directly existent in this case) to make sense of what the referent may be or may be doing. If attentional resources are demanded by other tasks, propositional references will not likely get consulted. Given that attention is a limited resource and (what we think of as) multi-tasking capabilities are largely confined to highly practiced and deeply familiar situations, it is unlikely that a propositional reference gets read and processed extensively in a situation such as that faced by the crew of TK1951 during the selection of V/S and the intention to capture the glideslope (see Loukopoulos, Dismukes & Barshi, 2009).

The third problem of propositional reference is that it can only refer to a state, not to behavior. It has no predictive value and cannot represent the dynamics of the underlying referent process. Propositions, in other words, are ill-suited to support event recognition: they don't help human monitors see change, events, behaviors. The single label (RETARD) in the case of TK1951 severely underdetermines the range of behaviors that the word may imply, and how those behaviors may be sensitive to the influence of context (Woods, 1994; Nikolic, Orr & Sarter, 2001). Indeed, it is such an underspecification that it could logically and lexically easily include the sort of retardation that the crew had expected and (thought it) had instructed the aircraft to conduct. The very word (or token) "RETARD" semantically matches the action that is expected of the autothrottle system. Also, the token "RETARD" cannot disambiguate what the referent is (a "hard" retard behavior associated with an autoland, or the "softer" retard behavior associated with idling the throttles to slow the airplane down momentarily).

Aviation is not unique in running into the problems of seemingly economic propositional referencing: safety in fields ranging from spaceflight to anesthesiology to nuclear power generation has suffered from the underspecified nature of the token-referent relationship in propositional representation (Freund & Sharar, 1990; Moll van Charante et al., 1993; Woods, 1995). In the design of displays for the monitoring of complex, dynamic processes, this lack of support for event recognition has been shown to create the potential for serious performance breakdowns.

Such problems associated with propositional referencing have, over the past decades, gradually been reduced by the use of more advanced forms of referencing. In iconic referencing, for example, the relationship between token (the thing on the screen) and the referent (the behavior or thing in the system or the world) is mediated by resemblance: the thing on the screen is made to look like the system or the thing in the world. Icons are popular in desktop computing and also in newer-generation cockpit management systems (e.g. on

the Airbus). Though dependent on convention (or even previously stored knowledge or experience), icons usually lead to quicker recognition and user action than propositions do. A further form is analogical referencing in which the structure and the behavior of the tokens on the screen are related to the structure and behavior of what is represented, through some natural constraint. The trend vector on the speed tape could be said to present the image of an analogical reference, as the structure and behavior of the vector seems related to what the aircraft's speed is about to do (though technically it is not a trend but a representation based on a prediction (which in turn is based on certain assumptions)—see below).

It seems that automation surprises are connected specifically to propositional referencing, in that the surprise is a result of the automation behaving (or the automation making the aircraft behave) in unexpected ways (see Sarter, 1997; Nikolic, Orr & Sarter, 2001). Propositional references are ill-suited to represent behavior. They are even less well-suited to represent future behavior. This has been one of the defining themes from the earliest discussions about flight deck automation problems in the literature onwards (see Wiener & Curry, 1980). The question "what is it going to do next?" which has often been asked on flight decks (see Wiener, 1989; 1993; Billings, 1996) is a strong indication that the current generation of cockpit displays do not do a good job of supporting the task of predicting the immediate future in a dynamic environment. They do not help pilots understand the intentions and future behaviors of the automation very well. TK1951 is an example of this: the propositional reference of "RETARD" underspecified the future behavior of the automation to such an extent that the crew ended up being surprised by its outcome (which was not noticed until the stick shaker, and probably still then not even fully understood).

## "Moving thrust levers" that didn't move, and other cues

Another cue that is easy to point out in hindsight is that the thrust levers did not automatically advance once the aircraft had become established on the glideslope and airspeed had reduced toward the target speed. Noticing that the thrust levers don't move again involves the noticing of a lack of change, psychologically subject to the same difficulty as noticing the absence of change in airspeed reduction or mode annunciation (see above). Two decades ago, much was made of the difference between Boeing and Airbus aircraft—the latter (from the A320 family and upward) having thrust levers that do not move when the autothrottle is engaged. In fact, the thrust levers often serve more as a mode selector than as thrust levers in Airbus aircraft (e.g. Billings, 1996; FAA, 1996).

On the approach of TK1951 it may well have been that the F/O had his hands on the control wheel and the thrust levers, as this is standard practice. This does not mean, however, that he would have been precisely calibrated, on his seventeenth flight on the real aircraft ever, about the appropriate positions of either for this approach (his first to Amsterdam). In fact, the amount of experience may not even be an important factor here. As one pilot commented, "I have been on advanced/automated aircraft for about 12 years and basic flying skills have deteriorated somewhat. Using autothrottles causes you not to know basic power settings" (FAA, 1996, p. F-1). In other words, where the thrust levers should be in any phase of flight (what the "basic powersetting" is), or how much they should move during the latter stages of an expected airspeed deceleration on an approach for which you got vectored too high, could be an instance of a more general knowledge problem with advanced automated aircraft (CAA, 2004). The issue here too is that no movement of thrust levers could even be expected for the first 70 seconds of the deceleration.

Indeed, there is a more general problem of not knowing precisely what setting thrust levers have when they are governed by an autothrottle. In 2004, the CAA expressed this concern, saying that "dependence on automatics could lead to crews accepting what the aircraft was doing without proper monitoring." (CAA, 2004, p. vi). Other incidents corroborate the generality of this issue (see also FAA, 1996). A 2007 incident with a subtle autothrottle failure in a Boeing 737 on an ILS approach in the dark and in IMC (Instrument Meteorological Conditions) at Bournemouth confirms the difficulty of detecting subtle ways in which the autothrottle is not doing what crews may have expected it to do (AAIB, 2009), independent on whether crew members' hands may have been on the thrust levers or not, or whether the thrust levers move with autothrottle commands or not.

*A/T P/RST flashing amber*

A final cue (certainly in hindsight) may have been the small square A/T P/RST (Autothrottle disconnect) light, located above the respective Navigation Display on either side of the cockpit, between the A/P P/RST and FMC P/RST lights. If this particular Turkish B737 was equipped like this, this light may have flashed in amber (the condition, according to FCOM2 chapter 4 is that it flashes amber when an A/T airspeed error is detected. The notion of "airspeed error" is not defined more specifically in the FCOM2, but includes the condition that airspeed differs from the commanded value by +10 or -5 knots and is not approaching the commanded value. It could be (if installed) that these conditions had been fulfilled during the final seconds of the flight of TK1951.

Showing, however, that this indication may have been available in the cockpit at the time is not the same as knowing whether it was seen by any of the crewmembers, nor, if it was, what meaning they may have given it in the short time available. As said at the beginning of the report, it is seductive to conflate data availability with data observability, particularly in hindsight. This conflation should never, however, be confused with an explanation for why people did what they did. The demand to integrate the meaning of (and act on) a set of cues may seem simple and unproblematic with the benefit of hindsight and virtually unlimited analytic resources. In contrast, when that demand for meaning integration is situated in a rapidly evolving, cognitively noisy context that takes no longer than half a minute, those cues are impoverished and underspecified and only a few among a multitude of others. Also, such meaning integration had to be accomplished in a situation where the crew faced significant other workload and interleaving task demands (the topic of the next section).

*F/D command bars on the left PFD*

A similar question could be asked about the disappearance of the flight director roll and pitch command bars on the captain's Primary Flight Display (PFD), a by-product of the anomalous RA input from the right RA. Their disappearance can be construed as a cue that there is some anomaly that (in hindsight) would have required some kind of action. First, there is no basis in pilot training or Boeing 737 documentation available to pilots that helps make this link (between RA input anomalies and the disappearance of F/D bars). The FCOM even says that the F/D can be used with or without A/P and/or A/T, which suggests that the disappearance of the F/D bars has no influence on the functioning of either A/T or A/P. In the likely mental model of a pilot, the disappearance of the F/D bars says nothing about the functioning of malfunctioning of the A/T or the quality of inputs it receives. Second, there was no accompanying anomaly in autopilot behavior: it kept tracking the glideslope. Third, interleaving task demands and the reading of the landing checklist would have demanded the Captain's attention elsewhere (see the next section). As the Captain was pilot monitoring (PM), he would have been the one looking out for the runway or approach lights, whereas the pilot flying (PF) would have been more focused on his own Primary Flight Display which showed no anomalous RA readings or other issues.

## Workload and interleaving task demands

A few seconds after V/S was selected in an effort to capture the glideslope, TK1951 was asked to contact the tower, which the Captain did. TK1951 then received landing clearance on runway 18R from the tower controller. During this same time, the glideslope was captured and V/S mode changed into G/S mode and the A/P started to follow the glideslope. Immediately after that, the crew started completing the before-landing checklist. From the selection of V/S mode to the completion of the landing checklist, 80 seconds passed in which the crew accomplished a large number of duties and items within a very compressed time.



*A THY landing checklist being accomplished in a B737NG cockpit (left seat). The checklist, held between the left-seat pilot and the control yoke blocks the view of the PFD (Primary Flight Display) which includes the airspeed tape (Photograph: Author).*

The accomplishment of the landing checklist at THY involves the physical extraction of the checklist (an A4-size laminated card) from the center glareshield, with the PM (pilot monitoring) running through the items on the landing checklist on that card (possibly with her or his thumb indicating the

item to be accomplished, and sliding down when going through the list). Even a fast-paced accomplishment of the landing checklist, including the direction of attention to the various items, instruments and switches across the flight deck, takes more than a minute. During this time, the landing checklist is held in hand by the PM. Such work can detract visual attention from other monitoring tasks (see Shepherd, Findlay & Hockey, 1986), and increase the likelihood of missing incidents in the monitored process (Byrd, Adams & Ntuen, 2002).

An examination of the B737 cockpit from the left seat revealed that one of the most plausible positions for the checklist is between the pilot and the control yoke. There is almost no room for an A4-size checklist to the left of the control yoke (because of the proximity of the sidewall and window) and this would not enable the pilot to switch visual attention rapidly between checklist and the various items distributed through the flight deck. There is room on the right side of the left control wheel, which would place the checklist over the center console and throttle quadrant. This position is typically used when pilots are both to consult the list (or whatever document it may be), but not as often when one pilot is consulting the written document or checklist alone. Positioning the checklist between the pilot and the control yoke has the side effect of blocking the view to the PFD, including the airspeed tape (see the photograph).

Some airlines have tried to avoid this problem by only using a so-called control wheel (or "yoke") checklist for those checklists that need to be accomplished while the airplane is moving. In other words, a paper checklist, like the one used by THY, is still used, but only for checklists that are used when the airplane is stationary (e.g. preflight checklist, before start checklist, before taxi checklist, shutdown checklist, secure checklist). All other checklists (before take-off, after take-off, descent, approach and landing checklist) are put in a little list on the control yoke. This avoids the problem of having to pull out and consult a large paper checklist while the aircraft is in motion, which could entail missing crucial information either from the outside of the airplane (e.g. during taxi) or the instrument panel.

A paper checklist that needs to be consulted during flight or taxi is problematic only when it interferes with other important tasks, of course. The nature of flying in a busy environment (like AMS) is such that it doubtlessly will.

*Multi-tasking and checklists in dynamic situations*

While the Captain may have had his view of his PFD blocked by the checklist held in front of him, the F/O may still have seen his. And there is no other indication that something could have physically obstructed his view of the PFD.

And, of course, it is possible that the checklist was held differently after all by the Captain, so that he too may have had a view of his PFD. If this were the case, which we will never know, it does raise the question why it would have made sense for the two pilots not to notice the speed decay below Vref.

One answer to this lies in the coinciding of the accomplishment of the landing checklist on the approach of TK1951 with the unexpected airspeed decay below Vref. While the airspeed was going below Vref, the crew was going through the various items on the landing checklist, leading them to allocate cognitive and visual attention to the items on that checklist and how they are distributed across the cockpit (see the figure).
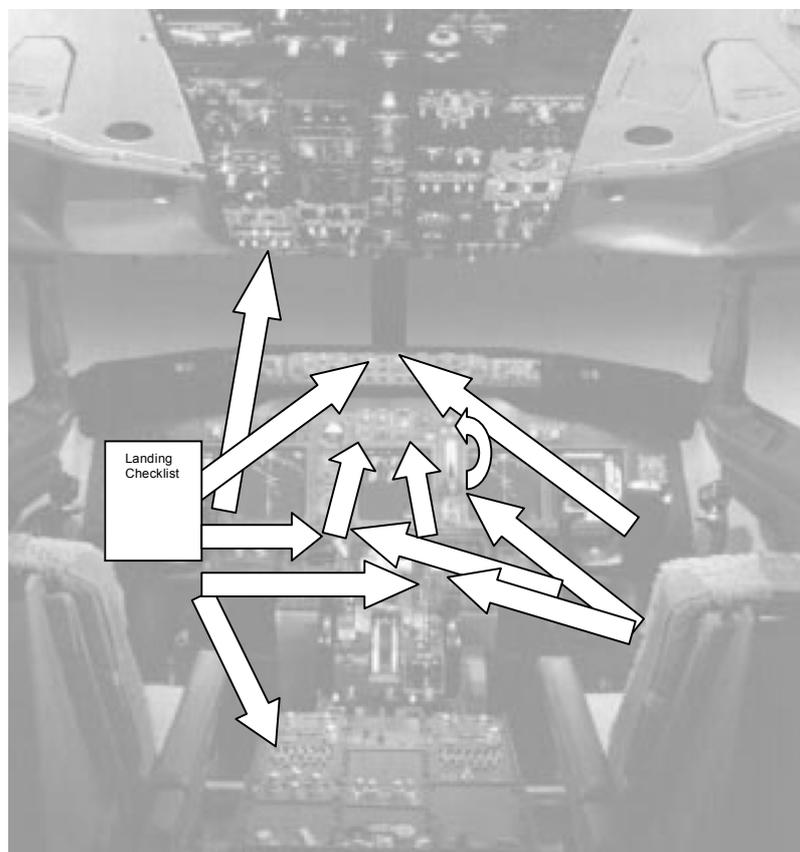


*Tracing the successive flows of visual attention during the accomplishment of the Landing checklist: Flaps 40, speed set; Speedbrake, speedbrake armed, green light; Landing gear, down, three green; Flaps, 40, green light; Cabin report confirmed; Missed approach altitude set; All lights on. Little attentional resources are available for other items such as the PFD. The accomplishment of the landing checklist during the approach of TK1951 coincided in time almost entirely with the speed decay below Vref (Cockpit photo courtesy of THY).*

Also, at the same time, with TK1951 about 2,5 nm from the runway threshold, consistent with the minimum reported visibility at that time (4500 meters), it is plausible that the runway or approach lights would have become visible at this point. The crew (in particular the captain as Pilot Monitoring) may therefore have directed part of their attention to the emerging runway environment outside the cockpit and its cues for the coming landing.

This problem of overlapping tasks and demands for mental and visual attention is of course one important reason behind the idea of a stabilized approach. If the reading of the landing checklist have not been completed below a certain altitude (often 1000 feet in conditions such as those of TK1951's approach), then the approach should be abandoned altogether. This issue is addressed under the heading "Why not make a go-around?" in the next section ("CRM and the intervention decision").



*The overlapping times of the speed decay below Vref and the completion of the landing checklist*

Checklist accomplishment, however, is not free from interference—and other tasks are not free from interference by checklist accomplishment—independent of whether the aircraft is above or below 1000 feet. Indeed, the problem of checklist and procedural design and how their assumptions fare in real-world flying is generic and has recently been the subject of both experimental and observational research by NASA and others (Degani & Wiener, 1990; Loukopoulos *et al.*, 2009).

*Assumptions of checklist accomplishment*

Airline Operations Manuals prescribe the standardized techniques that should be used in conducting an approach to landing. In particular, such written guidance prescribes the procedural flows to be used, and checklists to be accomplished, and when. It is possible to write such guidance only if a number of assumptions are made about the environment in which the guidance is to be followed. The most important of these assumptions have been captured by Loukopoulos and colleagues (2009), and include:

Assumption 1: The environment is linear: This assumption says that tasks always follow a prescribed order in a fixed sequence. The successive order in which tasks are to be accomplished is prescribed by Operations Manuals. The idea is that each item can be completely finished before the next item is taken on, or that a whole checklist can be finished before a new task can be taken on (e.g. flying the aircraft down from 1000ft). That way, tasks and checklist can be done in the prescribed order, and the possibility of changing, or having to change, the order of tasks is mentioned in such guidance only as an afterthought (e.g. "minor variations in order are acceptable"). There is no guidance on how to manage such variations or on how to ensure that an entire checklist is indeed accomplished, or that the delayed execution of a checklist does not intrude into concurrent or subsequent tasks. The assumed linearity implies seriality: only one activity is performed at the time, there are no parallel task demands. This of course contradicts the real nature of flying an approach (of which each approach, including that of TK1951, is an example: monitoring various things in the cockpit and coordination with ATC always accompanies the execution of checklists in parallel.

Assumption 2: The environment is predictable. This assumption says that tasks and events can all be exactly anticipated, both in nature and timing. Variations in pacing, timing and sequencing are hardly discussed in available written guidance. Also, the assumption implies that when pilots need to complete a certain task, then all information for completing that task is available unambiguously and in its entirety. There is no mentioning of the need to deploy prospective memory (the memory for accomplishing things yet to do, in the (near) future). A PM, for example, has to assume that the other pilot is ready to respond to a checklist when such a checklist is pulled out.

Assumption 3: The environment is controllable. Another important assumption in the construction of checklists and procedures is that pilots are in control of the environment in which they fly. That is, pilots are assumed to be in control over the timing, pacing and manner of execution of their tasks. This would also mean that sufficient time is available to actually complete the tasks, that pilots can devote their full attention to the task at hand, and that they can execute the tasks in the manner they had anticipated or planned. If other tasks intrude in for example checklist reading, or checklist reading intrudes in other tasks, there is no written guidance available on how priority should be given (or not) to the intervening tasks.

All three of these assumptions are routinely violated (in fact, they never really apply to any real aviation environment), and they were for the approach of TK1951 too. When descending toward 1000 feet, the crew was occupied with

managing the situation that the tight turn-in onto the localizer, and the fact that they had been kept at 2000 feet, had put them in. This involved the managing of different flight modes (V/S) and a monitoring of the descent relative to the glideslope and the appropriate airplane configuration and speed associated with that. Also, during the same time, TK1951 was transferred to the tower controller and the crew had to coordinate their landing clearance with him. There was no room to execute the landing checklist during this time: a typical example of how other demands can interrupt, interfere or delay the execution of tasks that are assumed to be linearly executable and sequential.

Of course, the question still is why the crew did not abandon the approach when they realized that the landing checklist could not be completed before 1000 feet. This is very much a question animated by hindsight, as we now know the outcome of the sequence of events. Every day across the world, approaches are flown unstabilized relative to flight parameters such as airplane configuration, airspeed, height or lateral displacement (which is knowable from flight data monitoring). A possibly even greater number (which is not knowable unless there is reason to listen to the CVR) is not stabilized in the sense that the checklist has not been read out loud (and thus not "completed") even though all the items on the checklist have actually procedurally been executed above the relevant altitude. It takes a special kind of encouragement from an airline, and probably a special kind of aircrew or very special kind of situation, to abandon an approach in which the aircraft is on speed and on path and in the right configuration, but where the only lacking item is the reading of a checklist.

This, indeed, was the item that made the approach of TK1951 procedurally unstabilized (and that the crew must have known about—the engines being at idle was not likely the one they knew or realized given its source in an automation surprise): the non-completion of the landing checklist after passing through 1000 feet in IMC. This has not systematically been studied since the spread of the stabilized approach concept around carriers across the world. But it could well be that of all the cues that should, procedurally, trigger a missed approach, the non-completion of a checklist is the weakest one. Being high on speed or low or high on glidepath, or off the localizer, or having the engines at a power setting that is persuasively not consistent with the phase of flight (again, not something that the TK1951 crew would likely have realized) may serve as more powerful or persuasive cues for breaking off an approach. Not having completed the reading of a checklist leads to a measure that many crews appear to find more plausible than going around: simply complete the checklist (even though a particular altitude window has been passed). Indeed, many crews do so (see Khatwa & Helmreich, 1999) and almost invariably these approaches are followed by safe landings.
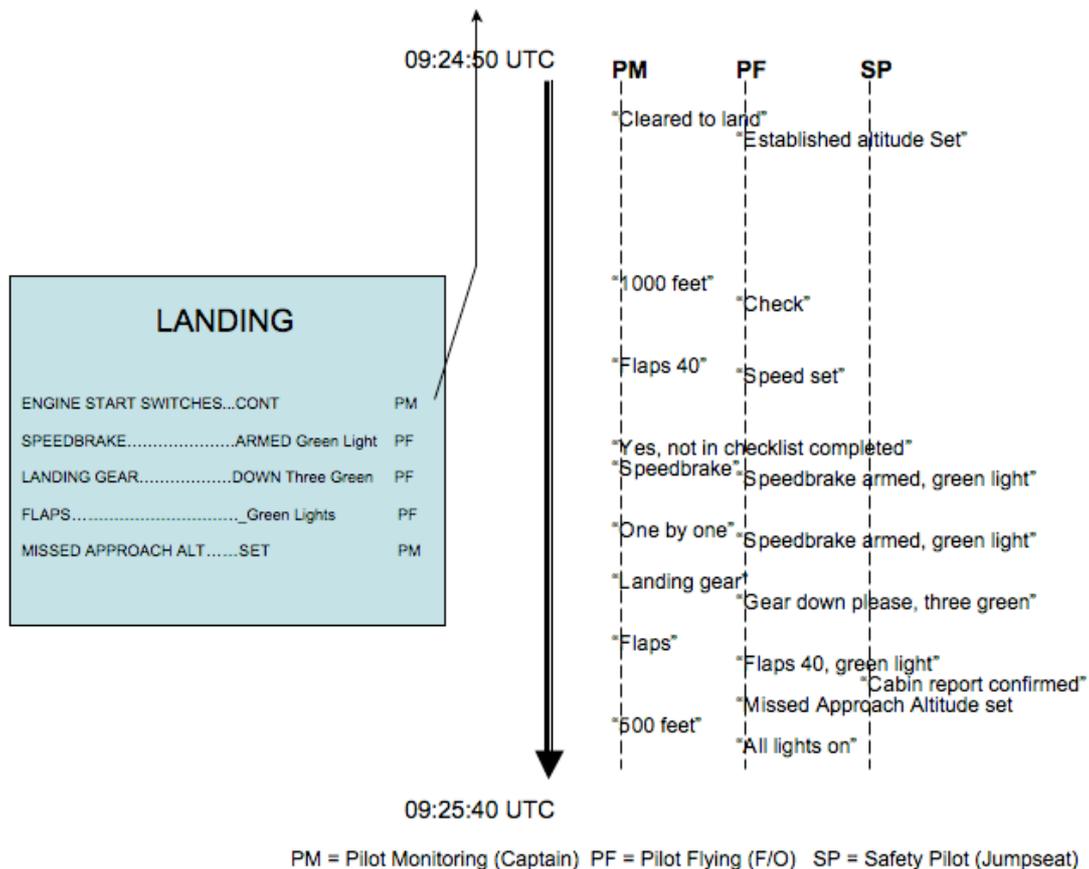
Also, when in VMC, or visual with the runway, an approach becomes typically classifiable as unstabilized only below 500 feet. The crew of TK1951 became visual with the runway between 800 and 700 feet and finished the landing checklist before 500 feet and by that time, the airplane was completely configured, on path and supposedly on speed. That the speed was still dropping, very likely unbeknownst to the crew, was the result of the automation surprise explained in the previous section. Again, more about the question of a go-around is dealt with in the next section of this report ("CRM and the intervention decision" under the heading "Why not make a go-around?").

Of course, none of the distractions, changes or interruptions are possible to anticipate in any meaningful way or in any detail in written procedures (such as the Operations Manual). And not all eventualities and complexities can be accommodated in written guidance. The most effective way to write up tasks to be done (e.g. in an Operations Manual) is in the sequence in which they normally (or desirably, ideally) are conducted. But that leaves two major problems (Degani & Wiener, 1990; Loukopoulos et al., 2009). First, procedures developed from the perspective of an ideal operating environment (based on the three assumptions above) tend to be brittle, or fragile. That is, their execution, order, or completeness, easily crumbles under the pressure of real task pacing, concurrent task demands or other complexities or surprises of real-world operations. Second, although pilots do learn to manage concurrent task demands, particularly through experience, they (as do other professional groups) tend to underestimate their own increased vulnerability to error in those circumstances (Loukopoulos et al., 2009).

Understanding real work on an approach to a busy airport requires understanding of the contrast between the ideal, assumed operating environment and how it fits procedures and checklists, and how these same procedures and checklists need to be interleaved, spliced, paused or paced in real-world operations. In many real-world situations, pilots cannot delay or defer one task long enough to complete another one fully. This means that multiple tasks need to be interleaved (e.g. monitoring instrumentation and automation when accomplishing a checklist). Interleaving unpracticed tasks, particularly those that involve novel or unexpected aspects (as was the case with TK1951), leads often to errors of omission (Loukopoulos *et al.*, 2009). It is interesting to note that much of the checklist research is focused on what happens if checklist-reading and execution is interrupted, and how various aspects of checklist and procedural design can, as much as possible, insulate checklist behavior from such interruptions and the omissions they engender (Degani & Wiener, 1990). Less attention is paid in this research to a case such as that of TK1951, where the execution of the checklist (procedurally demanded

as a precondition for landing), and itself complete and uninterrupted, intrudes into other tasks, such as instrument monitoring.

One more issue with respect to the landing checklist. The landing checklist as conducted by the crew of TK1951 was triggered by the 1000 foot call and the subsequent call by the F/O to select flaps 40 and set the speed (for that flap setting). The item "engine start switches—continuous" was not mentioned during this time. There are two likely reasons for this. The first is that the aircraft had been descending through cloud below +10° for a while, and was still in that cloud, which means that the engine start switches would have been selected to continuous for a while already (for the engine anti-ice system), and checking that again would have taken up precious time and resources (this is why other airlines elect to put the "engine start switches—continuous" item on the approach checklist, which happens higher up and during often lower workload phases of the flight).



09:24:50 UTC

| PM | PF | SP |
| --- | --- | --- |
| | | |

"Cleared to land"
"Established altitude Set"

"1000 feet"
"Check"

"Flaps 40"
"Speed set"

**LANDING**

ENGINE START SWITCHES...CONT     PM
SPEEDBRAKE.....................ARMED Green Light  PF
LANDING GEAR.................DOWN Three Green  PF
FLAPS.................................._Green Lights  PF
MISSED APPROACH ALT......SET     PM

"Yes, not in checklist completed"
"Speedbrake"
"Speedbrake armed, green light"

"One by one"
"Speedbrake armed, green light"

"Landing gear"
"Gear down please, three green"

"Flaps"
"Flaps 40, green light"
"Cabin report confirmed"
"Missed Approach Altitude set

"500 feet"
"All lights on"

09:25:40 UTC

PM = Pilot Monitoring (Captain)  PF = Pilot Flying (F/O)  SP = Safety Pilot (Jumpseat)

*The THY Landing Checklist and its accomplishment during the approach of THY1951. The item "Engine Start Switches...Continuous" would have nominally been completed long before this phase, as engine anti-ice would have been*

*selected on, given the meteorological conditions (this requires the engine start switches to be selected to continuous).*

The other possible reason is that the engine start switches is an item to be called and checked by the PM. What happened onboard TK1951 was that the Captain, after the F/O called for flaps 40 and the speed to be set, reminded the F/O that this was on the checklist that wasn't completed yet. The crew then took this as a cue to continue with the items on the checklist that the F/O should respond to. The second item on the checklist (and the first for the F/O to respond to), was speedbrakes. See the figure above.

An analysis of the landing checklist completion, and what the crew choreography around it may have meant is presented in the next section ("CRM and the Intervention decision").

## Speed tapes: How a cockpit knows its speed

From the sections above, there seems to be a preponderance of evidence that suggests that none of the three pilots may have seen or noticed the airspeed as the aircraft was descending below 1000 feet. But if they (or any one of them) had, then this raises interesting questions with respect to the human factors of the representation of airspeed through a linear tape, as it is on the PFD of the 737NG's flown by THY. It could be that such a representation, and the work associated with making it represent relevant (bug) speeds, may present some issues that make recognition of what is going on more difficult. Indeed, the awareness of the airspeed, built up through preparatory work in the cockpit leading up to the approach, may not be supported in the same way as it was in older-generation cockpits. All of this may contribute to a difficulty to recognize what was happening, why, and whether it was normal or consistent with expectations and instructions.

The calculation and representation of airspeeds, the cockpit work and crew cooperation associated with setting and keeping those speeds, as well as the manipulation of artifacts associated with airspeed, has undergone a dramatic change since the introduction of FMS's and PFD's into airliner cockpits. All ways of representing information about a dynamic process have advantages and disadvantages. They all affect the cognitive, collaborative and manual work that needs to go on in and among human operators in their own ways—both good and bad. New ways of representing and manipulating information may take certain traps and error opportunities away, but they will inevitably open new possible pathways to failure.

In 1995, ethnographer and pilot Ed Hutchins published a seminal paper called *How a cockpit remembers its speeds* (Hutchins, 1995) which traced in great detail the cognitive and collaborative work that goes on in non-FMS and non-PFD equipped airliners to look up, compute, cross-check, remember and then monitor speeds associated with different flap settings on an approach. This cockpit work begins after initiation of the descent and what it involves:

> "depends on the materials available, company policy, and crew preferences. For example, many older cockpits use the table in the operations manual and a hard plastic landing data card on which the arrival weather conditions, go-around thrust settings, landing gross weight, and landing speeds are indicated with a grease pencil. Still others use the table in the operations manual and write the speeds on a piece of paper (flight paperwork, printout of destination weather, and so forth). [Some use] a booklet of speed cards. The booklet contains a page for each weight interval (usually in 2,000 pound increments) with the appropriate speeds permanently printed on the card.
> The preparation of landing data consists of the following steps:
> 1. Determine the gross weight of the airplane and select the appropriate card in the speed card booklet.
> 2. Post the selected speed card [or other written representation of speeds] in a prominent position in the cockpit.
> 3. Set the speed bugs on both airspeed indicator (ASI) instruments to match the speeds shown on the speed card" (Hutchins, 1995, p. 271).

In preparing such a cockpit for the approach, in other words, there are activities such as looking up, cross-comparing (with a/c gross weight), writing down, positioning a physical artifact (speed booklet or other piece of paper with speeds on it). There will also very likely be some discussion, as well as the physical manipulation of speed bugs around the ASI and the visual and verbal cross-checking of those settings on the right and left ASI while or once they are completed. This rich variety of active work is very likely able to help the construction of a strong memory trace about what speeds will be appropriate for the coming approach. Not only is memory of speeds externalized to some extent (e.g. in the speed booklet and bugs); the physical manipulation and active verbal and visual engagement of these external artifacts forces crews to consciously consider each speed step down to Vref—both in advance of, and during airframe configuration changes (where speeds may be called out and cross-checked once more). This all contributes strongly to a memory trace in pilots' own minds. What marks such cognitive work, however, is not people simply recalling numbers. Instead, significant functions in the cockpit are achieved by people working with and interpreting material symbols: the

memory of which speeds to fly is not just in pilots' heads, it is distributed across the entire cockpit, built over time and represented in a number of physical artifacts with which the crews interact manually, visually and verbally.

This intricate cognitive architecture is altered with a change in technology. The introduction of the FMS and PFD has made almost all of these earlier crew manipulations of artifacts unnecessary. The FMS provides all the speeds based on aircraft gross weight, and sets the bugs automatically on the speed tape of the PFD. Vref for the desired final flap setting appears automatically in the FMS (based on aircraft gross weight), and gets cross-checked verbally and visually only once as part of the descent checklist, long before actually achieving and having to keep that speed. And, while "set speed" is called out under certain conditions during configuration changes, there is no active verbal engagement with the actual speed value for that particular configuration.

The new design evidently takes away a host of error traps inherent in the previous arrangement (e.g. where possibilities included selecting the wrong gross weight, selecting the wrong page in the booklet or writing errors on the piece of paper, setting bugs incorrectly along the ASI). But, again, there is no neutral ground in design. The new design changes the work associated with figuring out and remembering approach speeds and reduces the cognitive depth of crews' engagement with the task. There is much less physical manipulation and transferal of speed figures across different representations, less cross-checking, and less verbal coordination that involves mentioning the various figures. It is likely that this has consequences for crews' appreciation of configuration/speed relationships, as well as for the build-up of memory traces about speeds in their own minds.

*Monitoring round dials versus tapes*

One of the strong advantages of the round ASI turned out to be an unintended by-product of the design: it, together with the bug settings, allows crews to monitor in terms of *speed spaces* (Hutchins, 1995). Rather than matching the pointer of the dial with a particular bug (let alone a particular airspeed number), the visual matching task becomes one in which the pointer needs to be kept inside a particular space (like a pie slice on the round dial) for any particular configuration—a "speed space." Over time, crews will build up a memory of which speed spaces (or regions or pie slices on the dial) roughly match particular airframe configurations (even with adjustments for high or low gross weights on landing) and will construct expectations of the immediate visual appearance of the ASI for particular phases of the approach. Pilots learn to see *regions* of the airspeed indicator scale as having meanings that are not

expressed in numbers. Even a furtive glance will reveal immediately whether the pointer is in the right space or not.

There are other features of a round ASI dial that support the visual matching task (and thereby speed monitoring). One is that the end-points are visible simultaneously, providing the floor and ceiling and thereby the entire context for possible travel of the airspeed pointer. The pointer sitting in, or traveling into or out of a particular region, relative to the visible endpoints of the scale, carries immediate meaning about configurations and configurability that needs no further staring or inquiry (Gibson, 1979). Also, the velocity with which the pointer travels into or out of a region (and into the next one) is an immediate visual representation of a speed trend—a representation that requires no technological enhancement to be meaningful to a human observer.

The speed tape, in contrast, carries no such inherent meaning. This does not mean that crews do not gradually get used to a speed tape as opposed to a round dial, because experience shows that they do. But on a speed tape, the entire notion of what is up and what is down is in fact negotiable, as is the question of what should go up or down during a speed increase or decrease. Endpoints are not visible during an approach, which means there is little context available in which to place movements of the tape. Speed spaces or regions do not exist in any discriminable way, because, to the extent that there are any, they look like a linear portion of a linear tape, not a pie slice of a unique size, uniquely angled along a circle in a unique direction.

Research has shown that noticing change on a speed tape is harder than on a round dial (e.g. Roscoe, 1968). A first reason is that the difference between large changes and small changes are easy to note on a round dial. In contrast, the difference between large and small changes almost gets lost in the linear format of a speed tape. This, in part, has made it necessary to artificially engineer a visual trend back into the display, so as to make movements salient. The trend vector on a speed tape, however, is a representation of a number, and technically not a representation of a trend in the way that a pointer moving across a round dial is. The trend vector, after all, represents a numerical prediction: what the airspeed will be at a fixed number of seconds into the future (based on the assumption of conditions remaining the same). A second reason is connected to what is called *display compatibility*. To achieve compatibility between the conceptual analog quantity of airspeed in the world and the representation of airspeed in the cockpit, the display of speed should be analog too (and not digital, not just a number). That means that an airspeed scale should be fixed. Moving a pointer across that fixed scale brings out the conceptual analog quantity of airspeed, both in the movement and the position of the pointer. This became what Roscoe and his colleagues later called the

*principle of the moving part* (Roscoe, Corl & Jensen, 1981). The moving part in a compatible representation of airspeed is the pointer, not the scale. Yet in a speed tape, this principle is violated: it is the scale that moves and the pointer that remains fixed. This also means that airspeed on a speed tape comes closer to the digital representation that Roscoe recommended against in 1968: the pointer is fixed against a digital number that changes as the scale slides by in the background.

One advantage of a speed tape versus a round dial is its more parsimonious use of cockpit real estate, but for that gain, there are costs in the way it supports the cognitive and visual work of speed monitoring. The visual matching of the speed tape pointer with the Vref bug is one remaining task that carries immediately visible meaning (i.e. the pointer is either above or below the bug, but it is, of course, the bug that moves on a tape, not the pointer, again violating the principle of the moving part). Color codings only carry meaning if the color codes are known (these are not inherently meaningful), and these can be represented on a round dial as well, and have been, of course. In the case of TK1951, the airspeed number is surrounded by a flashing yellow box if it decays too much below Vref. This would have happened on board of TK1951. The flashing yellow box around the airspeed number would have been a way to capture a pilot's attention if it were directed at the airspeed tape in the first place. The fact that this did not happen during the sequence of events of this approach may suggest that visual attention was indeed directed elsewhere (particularly the various items on the landing checklist distributed across the cockpit).

This confirms that airspeed representation alone of course does not explain the sequence of events of TK1951 (nor is it the point of Hutchins' research: it is the entire workflow, from top of descent on down, surrounding the production of a series of airspeed numbers and their representation across different media in the cockpit that helps build and maintain the memory and awareness, not just the way in which the final, actual airspeed number is represented). An interesting contrast case with airspeed and autothrottle (though with high speed rather than low speed) on approach occurred in 1984 on an aircraft with conventional round airspeed dials. The incident happened to a DC-10 at J.F. Kennedy airport on 28 February 1984. After crossing the runway threshold at the proper height but 50 knots above reference speed, the airplane touched down 4700 feet beyond the threshold of the 8400 feet long runway and could not be stopped on the runway. The crew steered it to the right and the aircraft came to rest in water 600 feet from the end of the runway. The weather had been poor and the runway was wet.

This DC-10's autothrottle system had been unreliable for a month prior to the incident, and had not reduced speed when commanded during the first leg (Stockholm-Oslo) of this flight either. The Captain had selected an approach speed of 168 knots to compensate for the threat of windshear. The throttles did not retard when the airplane passed 50 feet and did not respond to the autothrottle speed control system commands either. Other available data that, in hindsight, pointed to the real nature of the situation, included fasts-low indicators on the attitude directors of the airplane, and a vertical speed of 1840 feet/minute down on the approach on account of the high speed. The investigation discussed the problem of reliance on normally reliable autoflight systems at length, even at the expense of monitoring basic flight parameters (NTSB, 1984).

TK1951, however, shows again how the problem of automation surprises is difficult to overcome by asking crews to mistrust their systems, stare harder and intervene earlier. The monitoring of airspeed, independent of the format of representation (tape or round dial) is usually done quite sharply when flying manually, and there is research that suggests that such monitoring is also quite keen with automated systems that are not as reliable as today's (CAA, 2004).

# CRM AND THE INTERVENTION DECISION

An accident such as TK1951 is never just the result of a single individual not watching or integrating particular things; never just the result of an entire industry not ensuring adequate understanding of the automated systems it makes its operators work with; never just the result of design and certification decisions made entirely outside the control of most airlines and pilots. At the end of the day, all the pedagogical, design-, procedural and organizational preparations get channeled through the hands of a small team of pilots who have to make the safety-critical work happen at the sharp end of the system. It is therefore important to examine this team—its composition, interactions—to begin to understand how things could go so badly for them on this particular day.

This (final) section of the report examines what has become known as CRM (Crew Resource Management) onboard of TK1951, and the intervention decision by the Captain (i.e. to take over control of the airplane). It does so by first looking at the composition of the flight crew onboard TK1951 on the 25th of February, then by laying out the basis that THY pilots have in (CRM) training, and then by considering the question whether the approach of TK1951 can be said to have involved a "breakdown of CRM" by setting elements of the voice record against the background of one promising forensic CVR technique and some cultural considerations. It then asks and tries to answer whether TK1951 represented a "rushed approach." After that, this section of the report studies the (obviously counterfactual and retrospectively judgmental) question why the crew did not make a go-around, by invoking research on naturalistic decision making and plan continuation. It concludes with the intervention and recovery attempt.

## The flight crew of TK1951

The Captain was 54 years of age at the time of the accident. Retired as a LTC (Lieutenant Colonel) from the Turkish Air Force, he had been hired by THY in

1996 and been type trained on the Boeing 737 in the same year, and flew the type as First Officer until 2005 when he was upgraded to Captain. He had been an instructor at THY since 2004, and had already brought considerable experience with him to THY as instructor in the Turkish Air Force. He had accumulated 10,885 flight hours on the Boeing 737 (of a total of 17,000 flight hours). Everybody interviewed at THY in the course of this investigation offered that this Captain was one of the most appreciated, likable, accessible and genial instructors and captains around, even going back to his years as Air Force instructor.

Even if such comments are normalized for the influences of post-mortem reflections about a deceased person, the residue of a well-liked Captain and instructor seems to stand. He was known as an instructor who would take his students' initiative and progress seriously, who would create for them the space to explore their expanding capabilities, allow them to make and recover from the sorts of technical mistakes consistent with their (low, growing) experience, and neither prematurely intervene in, meddle with, cut off, nor denigrate a student's assessments and actions. From interviews during this investigation, it became obvious that everybody liked flying with him.

The first officer was 42 years of age at the time of the accident. He had been hired by THY about half a year earlier. He too, had come from the Air Force, where he had achieved the rank of Captain and flew fighter jets (with an accumulated total of about 4,000 hours). The F/O was under line training (formally called LiFUS, or Line Flying Under Supervision), and there were no negative comments about his performance on his training records. LiFUS is the phase of training, on revenue flights, that occurs after the so-called base check (which involves a number of take-offs and landings on the type, in this case the B737, without passengers on board). The base check in turn comes after (in THY's case) several months of ground school and simulator training (see under the next heading "Training (CRM) at THY"). TK1951 was the F/O's 17th sector. He had accumulated about 50 hours of flight time on the B737 at that time. For the first 20 sectors of LiFUS training at THY, there is a Safety Pilot onboard. After these 20 sectors, there is a progress check. Thereafter, there are another 20 sectors of LiFUS flying, but then without a Safety Pilot onboard.

The Safety Pilot was 28 years of age at the time of the accident. He had been hired by THY in mid-2006, and came from a civilian background. His total flight time was over 2,100 hours, of which 720 were on the B737. As safety pilot, he occupied the cockpit's observer seat, behind and between the two pilots.

# Training (CRM) at THY

*Training and proceduralization at THY*

When it comes to proceduralization and procedural compliance, the DSB investigation shows THY to generally at least match industry standards. One example is an elaborate cruise checklist (provided on paper in the cockpit) that is used across fleets at the top of climb in which fuel predictions, drift-down procedures, en-route alternates, and any other specifics of the flight are calculated or discussed and then checked against the printed checklist. Though some airlines encourage pilots to do this (though many don't), few have imbued this briefing with the sort of organized, structured format that THY has. Another example is the use of a list against which to check whether all relevant items have been discussed in the departure and arrival/approach briefings. Even the crew rooms boast a wall-mounted checklist to help ensure that all relevant items have been taken along in the pre-departure briefing. There is no specific item on this list for briefing the role of the Safety Pilot during early LiFUS flights.

When it comes to training, one example of at least matching industry standards is the facility that THY owns and operates (and leases out to a number of other carriers from across the world). It includes not only full-flight simulators but a mixed cabin mockup (Airbus and Boeing) which abuts a large swimming pool on the one side for ditching evacuation practice and a dry side on the other for land-side evacuation practice. The simulator also has a cockpit mockup and includes everything from smoke, noise, motion and other effects to reproduce the events during a forced landing. Most interesting, and another example that puts THY above the industry standard, is the recent development (over the last two years) of elaborate LOFT (Line Oriented Flight Training) scenarios for cabin and cockpit together. Inspiration for this came in part from other recent B737 accidents like the one with Helios flight 522. Events such as a rapid decompression and emergency descent, fire, pilot incapacitation, rejected take/off and evacuation are all practiced with cockpit and cabin crews together as LOFT scenarios, including a scenario where all information flows from the cockpit cease and the cabin chief has to take over initiative for what to do next.

Another example is the provision of equipment, in the THY aircraft simulator facility, to audio- and videotape entire four-hour training sessions (where the instrument panel inside the simulator is captured by one camera, the flight crew by another). This record can then be used in debriefing sessions by the instructor, significantly enhancing the pedagogical return on investment of the simulator session, as the instructor can take the students back to particular portions of the flight and the assessments and actions and communications that surrounded them. These video records of training flights, by the way, are subject

to tight disclosure rules so that their only purpose is to enhance the instructional environment and enrich the pedagogical encounter. They are not a record of student performance to be kept for any other purposes.

*Initial ground school and CRM training*

THY's pilot corps is divided about 50/50 between pilots from civilian and military backgrounds. It is custom, at THY, to try to allocate Air Force pilots to the B737 fleet first (as opposed to the Airbus fleet), so as to allow them to build up handling skills with a non-fly-by-wire transport jet aircraft. The F/O had been allocated to the B737 (as had the Captain, 13 years earlier).

THY often shows itself, in many of its procedures, the documentation and compliance around them, as well as their facilities for training, to meet, and in several cases, exceed what could be regarded as the industry standard. One obvious example is the length of ground training for new hires. Whereas some airlines/TRTO's are comfortable training a B737 pilot inside of six weeks (including ground training, systems training and examination, simulator training, base check, company conversion, CRM, emergency training), THY spends considerable more time preparing its new hires. One reason for the extended period may be issues with English proficiency, but given that this takes 14 hours of explicit training (in English Radio Telephony), it certainly does not explain the length of the training period. Rather, judging from the curriculum and its content, THY takes the preparation of new hires seriously.

Like all new hires at THY, the F/O of TK1951 (already a highly experienced Air Force pilot) underwent classroom training for months before entering the simulator to start flight training. This stands in contrast with some other TRTO's (like the one used for comparison purposes in an earlier section in this report). The F/O of TK1951 had been given, in the classroom, 4 hours training in initial low visibility operations, 3 hours of Air Law, 14 hours meteorology and hazardous weather, 7 hours of THY aerodrome and route briefings, 14 hours of emergency procedures and equipment training, 4 hours of aircraft security procedures, 3 hours of special airspace procedures, 7 hours of Jeppesen documents training, 7 hours of dangerous goods training, 28 hours training in the THY Operations Manual and 14 hours initial CRM training. This was followed by 84 instructor-supervised hours of systems training on the Boeing 737 CBT (Computer-based training), and 28 hours of classroom instruction in B737 performance, flight planning and load and balance. The total time of ground school thus amounted to 2 months (as compared to a regulatory minimum of about three weeks). After this extensive ground school, the F/O did 12 hours of Multi-Crew Cooperation (MCC) training in a flight simulator, which

was followed by another 2 weeks of classroom training in standard operating and abnormal procedures.

THY has been conducting CRM training for over a decade. In an example of aiming above the industry standard, THY's CRM instructors are not only all active THY line pilots, and either active instructors or ex-instructors, they also have advanced graduate degrees (often from American Universities) in subjects relevant to CRM (this contrasts sharply with a number of airlines in Western Europe, who might not even have their own CRM instructors, and where those who instruct CRM often have no tertiary education whatsoever, in any subject, no airline or flying experience and no instructional experience (see also Dahlström *et al.*, 2008). The initial CRM training conducted by THY covers the topics of human error and reliability, error chain, error prevention and detection; company safety culture, standard operating procedures, organizational factors; stress, stress management, fatigue and vigilance; information acquisition and processing, and situation awareness; decision making; communication and coordination inside and outside the cockpit; leadership and team behavior synergy; automation, philosophy of the use of automation; specific type-related issues or differences; and case-based studies.

Rather than just talking about these topics (for which THY CRM instructors show an impressive array of audio-visual materials and well-equipped, modern classrooms), or showing borrowed or imported material, THY has decided to develop a number of CRM scenarios itself, and film them. These scenarios play out in different places (the flight deck, but also in the briefing room, and elsewhere in the aircraft) and involve THY crewmembers as actors. Each scenario is aimed at bringing out a specific set of CRM behaviors that students themselves are then asked to pick out, reflect on and discuss, and apply to their own training and situation. This is consistent with the experience of most CRM instruction that crew resource management training is not just (or even mainly) about imparting knowledge, but about raising awareness and beginning with laying the basis for skill development (Wiener *et al.*, 1993; Salas *et al.*, 2006). This demands different sorts of pedagogical engagement, particularly oriented at learning through reflection and discussion and practice. THY's approach to CRM training supports this well.

In fact, to the extent that there could be any personal or cultural sensitivities (for which it is not easy to find evidence in THY, actually), the solution to have their own scenarios as an inspiration for discussion turned out to be quite a good move. Again, to the extent that there could be any issues of some THY crewmembers feeling that they might lose face when confronted by others about their (CRM) performance, the videotaped scenarios offered a class (initial CRM but particularly recurrent CRM training where everybody, also senior

crewmembers, attends) the possibility to discuss and reflect on the behavior in the scenario rather than on anybody's putative personal shortcomings. Even with this considered, THY appears to represent and embody a profoundly modern and secular Turkish culture, in which speaking up to senior members of either staff (or family) is not seen as problematic but rather as the self-evident duty of everybody. Interviewees at THY stressed how young people in Turkey are generally better educated than they ever were, and have learned to express themselves better and more clearly, independent of whom they are speaking to.

## TK1951: A breakdown in CRM?

One of the biggest triggers for the institutionalization of CRM in aviation, say Dahström, Laursen & Bergström (2008), was the 1977 accident that resulted in the largest-ever number of fatalities to date, when two Boeing 747s collided on the ground on Tenerife (see also Wiener, Kanki & Helmreich 1993). Along with other major crashes during the 1970s, this marked the beginning of a new era in flight safety. Fight safety no longer seemed to be a matter primarily of pilots' skills at handling their planes,  nor an issue of technical reliability. Instead, pilots' skills relating to interaction with other people were found to be at least as important.

Two years after the collision on Tenerife, the National Aeronautics and Space Administration (NASA) held a seminar on resource management on the flight deck. Around that time, analyses of accidents and incidents began to focus on how the majority were not caused by an inability to control the plane, but linked instead to issues of information handling, decision-making, communication and leadership. NASA helped introduce the concept of CRM at the seminar, which originally stood for Cockpit Resource Management. Training programs for effective gathering of information, leadership, making decisions and collaborating with others were recommended. A large number of the airlines that participated in the seminar began to gradually implement such training (Wiener *et al.*, 1993; Dahlström *et al*, 2008).

When CRM training was first implemented in the industry, it generated resistance from some crewmembers, who may have felt that it was too oriented to psychology (or even psychotherapy) and that the commander's authority was undermined by the content (Wiener *et al.*, 1993). The examination of the THY CRM training programs, and interviews with their instructors, revealed that there are no such issues at play in THY.

United Airlines was the first airline to implement CRM training in conjunction with NASA's seminar (Orlady & Orlady, 1999). Since this time, CRM has been

continually changed and developed (Helmreich, Merritt & Willhelm, 1999). From having initially been a voluntary aid for airlines in internal training of pilots, in most parts of the world CRM has become a mandatory part of initial training, conversion training and a recurring training element for several personnel categories related to flight operations. Even in other fields—shipping, the chemical and nuclear power industries, and healthcare—training of this type is now being conducted. In the aviation industry, the number and kind of personnel categories that undergo CRM training has gradually expanded (maintenance, ramp, administrative). This is illustrated by the CRM abbreviation presently referring to entire crews (C now stands for Crew instead of Cockpit) or even linked to the functioning of the entire company (with C standing for Company) (Helmreich *et al.*, 1999; Dahlström et al., 2008).

*How to show a "breakdown of CRM"*

It may be an attractive idea to suggest that a "breakdown of CRM" ("Crew Resource Management") may have sped the pilots of TK1951 toward the outcome on the 25th of February near Amsterdam. There are two significant problems with making (let alone substantiating) such a suggestion—one of them specific to TK1951, the other a generic problem. The issue specific to TK1951 is the instructional context of the flight. This not only put the Captain in a dual role (as commander and final responsibility for the safety of the flight on the one hand; as pedagogical source and instructor on the other); it also introduced the presence of a third pilot, the safety pilot, into the cockpit. This of course affects the crew dynamics, roles and communication patterns in various ways.

The generic problem with citing "breakdown of CRM" as contributory to the sequence of events and its outcome is that there is no generally accepted definition of what "good CRM" might be (because what is "good" is so context-dependent), and, therefore, no good definition of what a breakdown of CRM may represent (see Salas *et al.*, 2006). Despite the conviction in the aviation industry that CRM has had a major influence in increasing flight safety over the past decades, this is something that is difficult to actually prove (Dahlström, Laursen & Bergström, 2008). Assessing which behaviors represent "good CRM" (or the opposite) is a topic rich with social-scientific debate and marked more by its diversity than its agreement or closure: A variety of proposals for how to assess CRM behavior have been generated over the years (e.g., van Avermaete, 1998; Flin & Martin, 2001; Baker & Dismukes, 2002; O'Connor *et al.*, 2002; Goldsmith & Johnson, 2002; Thomas, 2004; Klinect, 2005; Nevile & Walker, 2005, Salas *et al.*, 2006, Dahlström *et al.*, 2008).

In assessing CRM behaviors in the sequence of events leading up to an accident, the risk of the hindsight bias is ever-present too. Once we know that the outcome of a sequence of events is bad, we tend to look for the supposedly "bad" assessments and "bad" decisions that led up to that outcome and that supposedly helped produce it. This cause-consequence equivalence assumption denies us the opportunity to look at crew interactions for what they were (without knowledge of outcome, for the crew didn't have that knowledge either).

CRM assessment methods have been developed with the LOFT (Line Oriented Flight Training) or line-check context in mind (van Avermaete, 1998; Flin & Martin, 2001; Baker & Dismukes, 2002; O'Connor *et al.*, 2002; Goldsmith & Johnson, 2002; Thomas, 2004; Klinect, 2005; Nevile & Walker, 2005, Salas *et al.*, 2006, Dahlström *et al.*, 2008). Such settings offer the analyst a range of behaviors and data that is not at all available with the post-hoc, forensic analysis of a CVR recording. LOFT or line-check settings offer more time than is generally available on a CVR, they allow the observer or analyst to be part of briefings and debriefings, they offer the observer or analyst all kinds of data on body language and other non-verbal communicative behavior, and allow the analyst the opportunity to ask questions for clarification or background.

None of this is possible with the forensic analysis of a (relatively short) CVR recording. The problem of CVR recordings is that they are both short and underspecified. They only capture voice recordings (no gestures, no body language), and only for a particular amount of time. Though in the case of TK1951 the CVR recording was 2 hours, it still takes a number of leaps of faith to assert solid conclusions about the nature of people's relationships on a record of only two hours of patchy (and mostly task-oriented) conversation. A research direction that has shown promise with this setting, however, is one developed by Maurice Nevile in Australia (Nevile, 2004; Nevile & Walker, 2005), using social-scientific micro-analytic techniques such as conversation analysis to study voice traces of naturally occurring interaction. Inspired by this, excerpts of the TK1951 CVR recording are considered below for any evidence of a "breakdown of CRM."

*Analysis of TK1951 CVR portions*

Any claims made about a sequence of events, particularly if that sequence involves social interactions between people at another time and place, should first reside in evidence available in the data itself, not in generic assertions about suspected cultural or organizational inclinations that are freely ascribed to those people. In other words, what really needs to be looked at is what these

people did and said, and just how and when they did and said so, as their interaction developed. Claims about people's understandings and actions must be based on and demonstrated in analysis of the transcription data, asking "why that now?" (Nevile & Walker, 2005).

The key is to avoid preconceptions about the participants or the setting in which they worked, and not to ascribe any mental, motivational or emotional state to people without solid backing in both the record and the social-scientific basis against which to hold and examine it. Also, single utterances or actions cannot be meaningfully isolated for analysis or held up as examples that are supposedly representative of a crew's entire relationship. Actions and utterances always occur in context, in some sequence relative to each other. What is not said, or when things are said relative to other things, is certainly as important as what is said (Predmore, 1991; Silverman, 1998; Nevile, 2002). The sequential ordering of talk (who is talker and listener when), as well as other attributes of the utterances (pitch, length) can convey important social meanings that are picked up by participants without much conscious attention, but with consequences for what they may say or do next or how they may go about that (Silverman, 1998; Nevile & Walker, 2005).

In a conversation analysis of the CVR of a particular accident flight in Australia, Nevile and Walker (2005), showed how CRM between two crewmembers broke down before the aircraft crashed on a non-precision approach. The analysis offered in their special report is canonical in the sense that it is one of the first (but see also Predmore, 1991), that imports social scientific analysis into the micro-details of a voice record in an attempt to substantiate any claims about the quality of the CRM behavior that went on there. Their analysis showed that at least three aspects of the interaction between the crew could form a basis for claiming that there was a breakdown of effective crew resource management during that approach.

First, in the Australian accident analyzed by Nevile & Walker (2005), there were many instances of overlapping talk. That is, both pilots would speak at the same time. This obviously is problematic from an interactive standpoint as it forces participants to merge the roles of speaker and listener. Overlapping talk occurs when somebody other than the original speaker starts to talk. In another study (Nevile, 2004) found that this is actually unusual for flight crews: they normally wait with speaking before the other speaker is entirely finished (this may be in part an artifact of the noisy environment and the often electronically mediated communication through microphones and headsets, such as in some B737 cockpits). Overlapping talk does occur occasionally on flight decks, but then only in non-task oriented talk.

Second, in this Australian accident, there were many instances when the Captain of the aircraft said something and the F/O said nothing in reply, even though some kind of response would have been appropriate and relevant and an expected next action. Overwhelmingly, interactive conversation normally involves the production of talk in response to other talk. One person says something, the other says something that may be deemed appropriate as a follow-up or response. This is formally called the first pair part and the second pair part of conversation. People are sensitive to the sequential nature of conversation. Silences, even short ones, are meaningful, and might signal a problem with the first pair part (e.g. it wasn't heard or deemed irrelevant or inappropriate).

Third, the analysis of the CVR of the Australian accident showed how the Captain often corrected, or repaired, the F/O's remarks, even when there was no sign of any problems, from the F/O's point of view, in his actions or utterances. Repair refers to those points in conversation where participants try to recover from communicative problems of some sort. There is a marked tendency in everyday conversation for self-repair, which is seen also among flight crews (Nevile, 2004). Even when the other person initiates the repair (which may even be done with certain body language), the speaker him or herself will mostly repair his or her own talk. In the more exceptional cases where repair is both initiated *and* executed by another person in the conversation (itself unusual, it is called other-initiated-other-repair), then this is often done in such a way (hedging, qualifying, delaying, softening) so as to smoothen the impact of the repair for the recipient.

Interestingly, the latter, and richest portion of the TK1951 CVR reveals little of these three problems. In other words, it is difficult to substantiate an assertion about a breakdown in CRM on the basis of the primary data available (the CVR). There is essentially no overlapping talk. The crew members allow each other to finish their utterances without wrapping it up for the other person or interrupting or butting in. Where actions are (likely) taken by one crewmember that fall within the remit of the other (e.g. the selection of V/S, or of landing flaps), they are the by-product of the instructor-student relationship between the Captain and the F/O of TK1951 and the compressed time that resulted from the short turn-in onto final for 18R. There are very few silences after a first pair part; utterances are appropriately responded to. And the only cases where there is other-initiated-other-repair by the Captain, are also those that are the by-product of the instructor-student relationship; they cannot be persuasively related to any pre-existing hierarchical or authority-gradient issues.

The coupling between first pair part and second pair part, and how it represents not only standard operating procedures, normal crew interaction but also the

instructional relationship, can be seen nicely from the two heading changes received in the last stages of the approach (times indicated here are UTC times (hh:mm:ss)):

| 09:19:44 | ATC | "Turkish 1-9-5-1, turn left heading 2-6-5" |
| 09:19:49 | FC | "Left 2-6-5, 1-9-5-1" |
| 09:19:51 | FC | "Left 2-6-5" |
| 09:19:53 | F/O | "2-6-5" |

In this sequence, the Captain (FC) reads back "Left 2-6-5, 1-9-5-1" to the approach controller and immediately repeats "Left 2-6-5" to the F/O in the next second. Such repetition can be an artifact of the instructor-student relationship, making sure that the F/O is up-to-date (and the airplane has to change lateral mode as well, from LNAV to HDG SEL). This utterance to the F/O forms the first pair part, the second pair part, appropriately produced by the F/O is "2-6-5," accompanied by a selection of that heading on the MCP. Not much later, there is a similar sequence, but it differs subtly and interestingly:

| 09:22:40 | ATC | "Turkish 1-9-5-1, turn left heading 2-1-0, cleared approach 1-8 right" |
| 09:22:44 | FC | "Left 2-1-0, cleared I-L-S, Turkish 1-9-5-1." |
| 09:22:49 | F/O | "2-1-0 set, Hocam." |

In this sequence, the F/O uses the read-back of the Captain (FC) to ATC as first pair part to his second pair part "2-1-0 set, Hocam," possibly demonstrating how he is up to pace, learning, and anticipating events. The Captain, as instructor, is shown that he no longer needs to trigger the student to perform and read back a particular setting. Such immediate, sequential ordering without the perceived need for prompting other participants with additional first pair parts, could be a sign of an increasingly smoothly functioning crew, and a student who shows the expected learning gradient. There are neither silences that would indicate a breakdown of CRM, nor gratuitous verbal fillings with additional prompts (or first pair parts) that would indicate an instructor who is experiencing a student who is behind the curve or not learning.

In another example, even though it is a short piece of interaction, the completion of the landing checklist onboard TK1951 could be a worthwhile illustration of the relationships between the crewmembers and their respective engagement with what is going on during the final stages of the approach.

```
09:24:50 UTC        PM          PF          SP
                    "Cleared to land"
                              "Established altitude Set"

┌─────────────────────────────────────┐
│            LANDING                  │   "1000 feet"    "Check"
│                                     │
│ ENGINE START SWITCHES...CONT    PM  │   "Flaps 40"   "Speed set"
│                                     │
│ SPEEDBRAKE.............ARMED Green Light  PF │   "Yes, not in checklist completed"
│                                     │   "Speedbrake" "Speedbrake armed, green light"
│ LANDING GEAR.............DOWN Three Green  PF │
│                                     │   "One by one" "Speedbrake armed, green light"
│ FLAPS.....................Green Lights   PF │
│                                     │   "Landing gear"
│ MISSED APPROACH ALT......SET    PM  │            "Gear down please, three green"
└─────────────────────────────────────┘
                                         "Flaps"
                                                "Flaps 40, green light"
                                                        "Cabin report confirmed"
                                                "Missed Approach Altitude set
                                         "500 feet"
                                                "All lights on"
09:25:40 UTC

PM = Pilot Monitoring (Captain)  PF = Pilot Flying (F/O)  SP = Safety Pilot (Jumpseat)
```
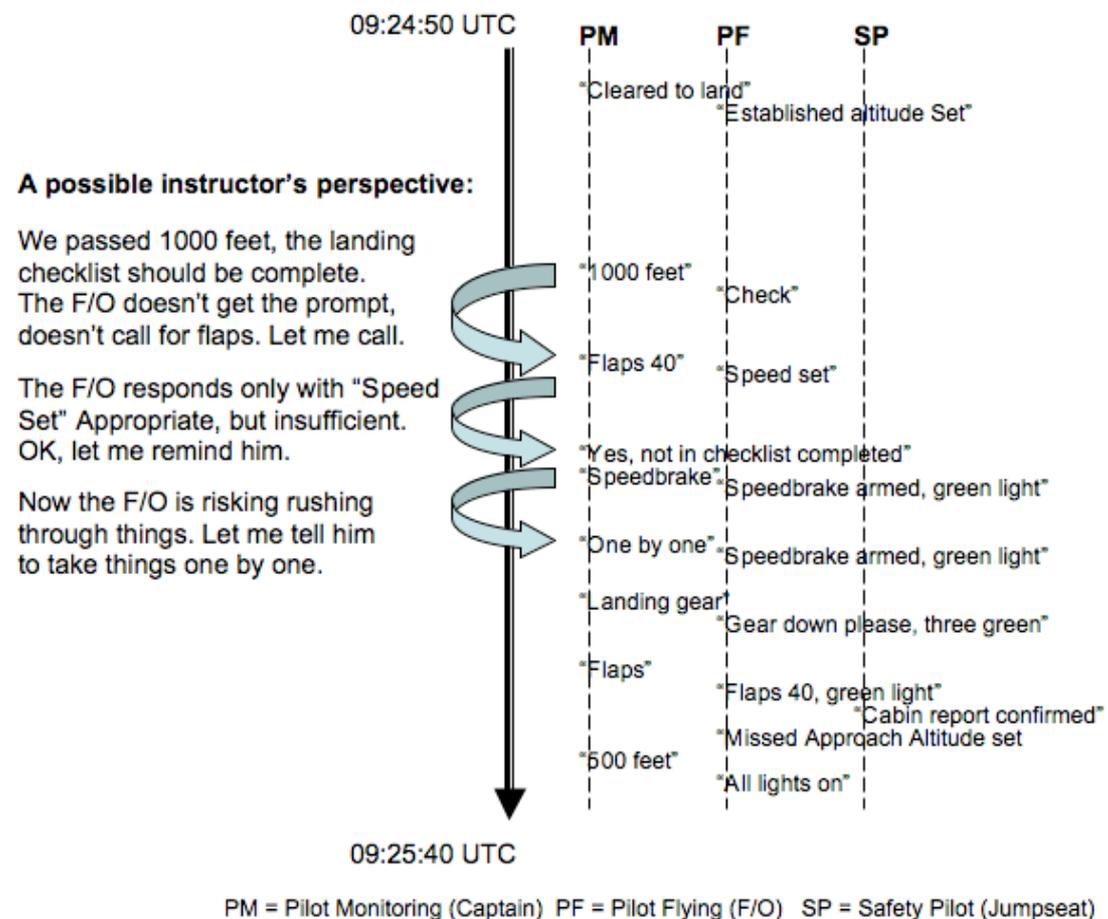
*The completion of the landing checklist during the approach of TK1951*

At 09:24:52, the Captain reads back that TK1951 is cleared to land. Two seconds later, cued by the landing clearance and the descent below their last assigned altitude of 2000 feet, the F/O announces "Established altitude set." This likely refers to the missed approach altitude, which he has presumably set in the altitude window on the MCP when the landing clearance was received. It is interesting to note that this is also the last item on the landing checklist (see the figure), which may possibly, in this rapidly unfolding context, have briefly prompted the F/O to believe that the list was completed (see Degani, Heymann & Shafto (1999) for similar examples). The receipt of landing clearance and subsequent setting of missed approach altitude can indeed be taken, on a typical approach, for the last items before landing. Again, the F/O had only had a few operational flights on the B737, which could mean that his experience with the pacing and completion of checklists in actual practice was quite thin at the time.

Nine seconds later (at 09:25:04), the aircraft descends through 1000 feet, which the Captain verbally announces, and the F/O responds to with "Check." This response, after a slight latency of five seconds, triggered the Captain to then call "Flaps 40" (and possibly select them himself as a result of his own call-out),

98

because the F/O didn't call for it (which is the role of the PF). The F/O in the next seconds says "Speed set" (which would be his role on the MCP at that point). The brief wait by the Captain—from the F/O's response of "Check" to the Captain's initiative to call for and select landing flaps—was followed by another brief wait (of about 4 seconds). During this time, the F/O responded only with the immediately successive item of setting the concomitant flap speed on the MCP and saying that he did so.

It is interesting that the F/O, as PF, calls the last item "Missed approach altitude set" (which he presumably set himself too, or had already done so earlier on when they received landing clearance). Immediately after the Captain's call of "Flaps" and the F/O's response "Flap 40, green light", the Safety Pilot comes in with "Cabin Report confirmed."



*A possible instructor's perspective on the completion of the landing checklist during the approach of TK1951*

The second, brief pause on part of the Captain may have been intended to prompt the F/O to call for the landing checklist, which would be consistent with the Captain's reputation as an instructor who will create the space for his students to make and recover by themselves from the sorts of technical mistakes that are the natural by-product of low experience. The other-initiated repair by the Captain ("Yes, not in landing checklist completed") is consistent with his role as instructor, not inconsistent with crew resource management principles for the conduct of this flight. As the F/O then starts to go through the items on the landing checklist that are his to call out, the Captain counsels him to take his time and not rush through things, by saying "One by one." The F/O repeats the first item that is his to call out (Speedbrakes) and then continues on down the list. A possible instructor's perspective is represented in the figure above.



09:24:50 UTC

| PM | PF | SP | **A possible student perspective:** |
| --- | --- | --- | --- |
| "Cleared to land" | "Established altitude Set" | | Last item on the landing checklist. |
| "1 000 feet" | "Check" | | |
| "Flaps 40" | "Speed set" | | I need to set the speed for the flaps. |
| "Yes, not in checklist completed" "Speedbrake" | "Speedbrake armed, green light" | | He's right, landing checklist, quick. |
| "One by one" | "Speedbrake armed, green light" | | OK, again. |
| "Landing gear" | "Gear down please, three green" | | Of course, gear is already down. |
| "Flaps" | "Flaps 40, green light" "Cabin report confirmed" | | |
| "500 feet" | "Missed Approach Altitude set" | | Let me help here, I set it already. |
| | "All lights on" | | Let me help here again. |

09:25:40 UTC

PM = Pilot Monitoring (Captain)  PF = Pilot Flying (F/O)  SP = Safety Pilot (Jumpseat)

*A possible student perspective on the completion of the landing checklist during the approach of TK1951*

Possibly sensing that the Captain's attention is captured by this, however briefly, and understanding that there is not much time to complete things, the F/O offers the missed approach altitude item himself (also because he possibly set it previously and called it out then, though not entirely with the right appellation).

All of this happens inside of eight seconds. This would be the sort of evidence of adaptation and resilience that Loukopoulos and colleagues (2009) observed in flight crews during approaches in the real environment: making trade-offs about items whose setting is and has been common knowledge for a while, and covering for each other (the Safety Pilot doing the cabin report, the F/O as PF picking up the last item of the checklist that was for the PM), all entirely smoothly and without any need for meta-coordination. Some would say that this is exactly how they would expect a professional, well-working team to operate with a checklist (Degani & Wiener, 1990; Loukopoulos et al, 2009). A possible student perspective is represented in the figure above.

Of course, these are only possible (at the most they are plausible) perspectives on what may have gone on in those few seconds of the approach of TK1951. As said in the beginning of this report, it is impossible in a human factors analysis to complete close the gap between data and interpretation, between the "what" and the "why." What matters most is offering a plausible (or multiple plausible) explanation(s) of why it might have made sense for people to do what they did, and then pulling possible learning leverage out of those conjectures.

*The role of the Safety Pilot*

TK1951 had a three-man flight crew on the 25th of February, consistent with LiFUS rules and practices. In fact (and again, something that puts THY above industry standard), the THY Operations Manual contains a number of pages of internal rules for who can fly with whom. Both first officers and captains are divided into three experience groups each (the requisite experience to make it into the next group is specified very exactly), and then there are a number of further subdivisions of these into other categories such as instructors. The Operations Manual features a table in which it is specified exactly which kind of pilot can fly with what other kind of pilot, when a safety pilot is needed, and all of this is taken into the scheduling of flight crews. In other words, THY has invested considerable and careful planning into setting up crew pairings, ensuring that the right mix of experience is always present in any cockpit. Who can be a safety pilot is also subject to conditions and restrictions (related to experience and experience on type).

The role of, and briefing to, the safety pilot during line training flights is described in THY Operations Manual Part-D Training, p. 2-8-7:

> "9. SAFETY PILOTS
> a) A safety pilot is a pilot who is qualified on a specific type of aeroplane and carried on board the aeroplane (during aeroplane training) for the

purpose of taking over control should the person acting as a PIC become incapacitated or when the trainee is unable to perform the required training.

b) The role of the safety pilot is to observe the flight training from the allocated observer seat.

c) …The safety pilot while seated in the observers seat is responsible for advising the PIC when any irregularities are noted.

d) …Prior to the commencing (of) a training flight, the PIC shall instruct the safety pilot on when to intervene and take control of the aeroplane.

e) During the briefing, PIC must advise the safety pilot when he wishes the safety pilot to conduct any flight crew ancillary tasks. Any specific duties shall be clearly specified prior to departure or any specific maneuver so as to avoid confusion between the PIC and the safety pilot. This is important when events are moving quickly and the aeroplane is near the surface, for example, during take-off or final approach to landing. Safety pilot must at all times remain alert.

f) The most important phraseology between the PIC and the safety pilot should always be respected "I HAVE CONTROL" from the safety pilot and the responding words "YOU HAVE CONTROL" from the PIC or visa-versa. It is simple, appropriate and ensures safety of crew and aeroplane."

It is impossible to assemble from the data available what was briefed before the flight. It is not usual, however, to specifically brief the role of the Safety Pilot, particularly not for those flights that are toward the end of the LiFUS period in which a Safety Pilot is still onboard. Before the first line flight of a new F/O, a Captain may offer the Safety Pilot some reminders and coordinate roles and possible interventions, but that would not be as likely for the seventeenth flight of an F/O. Given the Safety Pilot's role to "advise the PIC when any irregularities are noted," his remarks about the radio altimeter while on the approach could be expected and are consistent with the Operations Manual:

> 09:24:35    SP    "Hocam, radyo altimetre arizamiz var, hocam"
>                  ("We have radio altimeter failure")
> 01:24:38    FC    "Tamaaam" ("Okaaay")

The Captain's (FC) response to the Safety Pilot's (SP) advice of the irregularity can be seen in the context not only of the actual words used (radio altimeter failure) but also of what he and the F/O had done to the cockpit setup in order to insulate themselves from any problems associated with the left radio altimeter. Technically, it was not a failure (again, there were no failure flags, warnings, annuncations associated with the left RA, only a reading that was inconsistent with the right RA and seemed not to make sense relative to the

stage of the flight). And the use of FCC B (i.e. the right F/D selected as Master and the right autopilot, or CMD B, ON), would, according to pilots' available knowledge on the basis of training and documentation, effectively insulate their flight from any problems that the left RA might be having.

While the Safety Pilot does not respond to the Captain's "Okaaay", it cannot be established that this is the kind of silence (the lack of a second pair part) identified by Nevile and Walker (2005) earlier on as contributory to a breakdown in CRM. Directly after the Captain says it, he goes onto the radio to announce to the tower controller that they are established on the ILS for 18R. The tower controller then clears TK1951 to land, after which the F/O continues with utterances related to the landing checklist (see above). There is no opportunity for the Safety Pilot to say much at all. This means that it is equivocal whether the silence on his part is the result of a failed second pair part in the sequence of conversation with the Captain, or the artifact of the events on the approach and other necessary utterances by other participants. Also, the Safety Pilot does not hesitate to remind the Captain twice of his airspeed not much later on.

Whether any pre-existing difference between pilots from a military or civilian background could have played a role in the events onboard TK1951 is difficult to say on the basis of the available data. All instructors who were interviewed for the investigation, however, said that this difference seems greater to outsiders than it in reality is within the airline. With everybody similarly trained, in the same cockpit, with the same uniform, it is difficult to assert that a difference in background (which, in the Captain's case, was more than 13 years ago) would conspire against the ability of one crewmember to make himself heard by two others (one of whom he even outranks with his experience in the airline).

An interesting mystery in this, of course, is the sound that occurs 9 seconds before the Safety Pilot says "We have radio altimeter failure". The sound could be consistent with the release of seat belts (at least a shoulder harness). The autopsy report shows that the Safety Pilot, at the time of the impact, was not wearing any seatbelts. It is almost impossible to shove the throttle handles fully forward from a seated and belted position in the center cockpit observer seat; the forward reach for that is too far. This remark came well before the speed dropped below Vref, but V/S mode had just been selected and the airplane had started its descent toward the glideslope from above. His not wearing seatbelts, however, can also simply be the consequence of not having prepared for landing yet (one of the flight attendants in the back of the aircraft was not strapped in either yet), for a reason that cannot be determined on the basis of the data available. The sound on the CVR could then be spurious.

*"Hocam"*

On various occasions, the co-pilots used the appellation "Hocam," when addressing the Captain of TK1951. Interviews and other examination of the use of this word revealed that this is a reflection of common practice in Turkey that cuts evenly across civilian and military backgrounds (i.e. today those with a military background will not necessarily use the appellation more often than those with a civilian background, independent of whom they address). Though the term technically comes close to "teacher," it is not used exclusively in instructor-student relationships. In fact, it is seen as a safe placeholder label to apply whenever unsure about what to call somebody. A ground agent may call any member of the aircrew "Hocam," for example, or a child may call his or her father, or other more senior family member, "Hocam." The word may imply deference to some extent (but can really be a common politeness or comfortable familiarity more than anything), and it does not imply reverence.

For the purposes of the analysis here, it could amount to over-interpretation to read an unusually steep authority gradient into the usage of "Hocam" by the two co-pilots in their address of this Captain. It is neither unique to their relationships, nor to the Captain's role as instructor on this flight, nor unique to this Captain as person, nor demanded or expressly expected by him in any way. Also, despite the Captain's role as instructor on this flight, a steep authority gradient would be inconsistent with what was known about this Captain: a cordial, welcoming personality who created the space to allow his junior colleagues to make mistakes, while encouraging them to learn from them. Finally, it is common for colleagues at THY to call each other "Hocam" even if they have known each other (and worked together on many occasions) for twenty years.
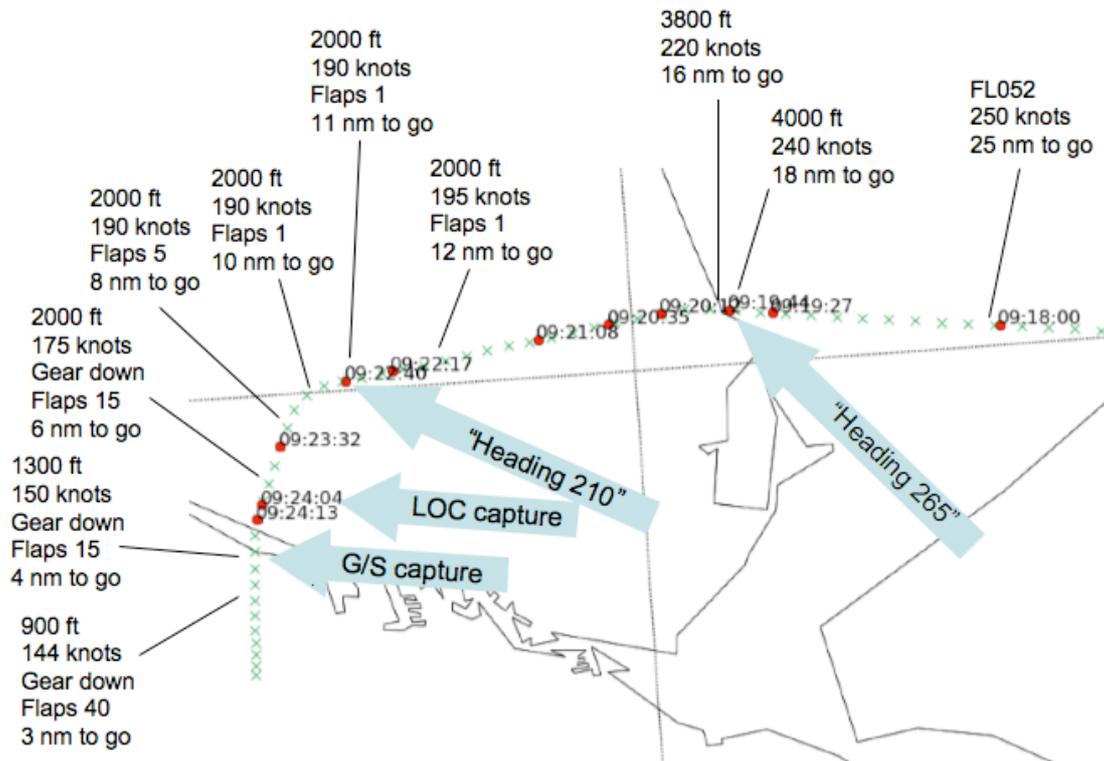
It is not easy to place the use of "Hocam" aboard airplanes in a meaningful relationship with the cultural-anthropological research base. The international research literature on Turkish culture focuses in large part on issues of gender, outward expressions of faith, intra-national ethnicity, migration, integration and adjustment in Western (European) societies, and in small part on domestic migration, tradition and patriarchy. There is little research focus on patriarchy and hierarchy in professional relationships, or the military/civilian interface as it plays out in professional relationships. Nonetheless, most research about Turkish culture points to the importance of relationships, and a proper understanding of and around them, as a key to the smooth social functioning of Turkish society. There are scores of appellations for various kinds and levels of relationships, "Hocam" being just one of them. It is not as if a Turkish pilot would call a colleague "Hocam" and then call everybody else in his or her

environment (it being kin, family, acquaintance, professional) by first name. On the contrary, similar appellations, appropriate to the relationship, are used all the time. "Hocam" does not stand out, nor should it be overrated for its meaning onboard TK1951.

Of course, no human interaction is flawless. Hence the often-used opportunities for repair (as discussed above). It is always possible to point to glitches in people's interaction, particularly in a micro-analysis of a sequential conversation. The point, however, is to be able to persuasively link such glitches to any causal argument. That becomes very difficult for this sequence of events. The analysis above shows none of the previously known aspects of a breakdown in CRM to be overly present (overlapping talk, silences, other-initiated repair) other than the few that are a by-product of the instructional relationship between the Captain and the F/O. The Captain, once again, was known for his geniality and well-liked as an instructor. It is difficult, on the basis of the data available, to claim that TK1951 involved a breakdown of CRM, or that this would have had any major role to play in how the events turned out. To think that we can prescribe a mere cultural or collaborative recipe against the events as they occurred and combined on the approach of TK1951 on this particular morning would probably generate a red herring, given the nature of the automation surprise (which is culturally independent) and the kind of B737 systems training that went before it (which is industry-wide).

## Was TK1951 a "rushed" approach?

Just like it may have been appealing to consider whether a "breakdown of CRM" occurred on the approach of TK1951, it would be plausible to consider the aircraft ending up "hot and high" and the approach ending up being rushed. In hindsight, with the landing checklist being completed below 1000 feet, this could seem very plausible, and even contributory to the outcome. There is, of course, no generic set of criteria that would make an approach "rushed." In fact, the experience of a rushed approach is likely that—an experience. Being turned onto a tight final, for example, even above the glideslope, does not have to automatically make an approach rushed—if the flight crew is aware of ATC's intentions and has a chance to prepare. The experience of a rushed approach, then, would increase with the unexpected compression of the flight path (both lateral and vertical). That indeed could apply to TK1951. But even there, the basis for any conclusions about whether the approach was rushed must be sought in the data that are available. This does not include the crew's experience of the approach, but only the voice and flight data records that they, and their flight, left behind.

The diagram shows the approach track with the following annotations (from upper right to lower left):

FL052 / 250 knots / 25 nm to go — 09:18:00

4000 ft / 240 knots / 18 nm to go — 09:19:14 / 09:19:27

3800 ft / 220 knots / 16 nm to go — 09:20:00 / 09:20:35

09:21:08

2000 ft / 195 knots / Flaps 1 / 12 nm to go — 09:22:17

2000 ft / 190 knots / Flaps 1 / 11 nm to go — 09:22:40

2000 ft / 190 knots / Flaps 1 / 10 nm to go

2000 ft / 190 knots / Flaps 5 / 8 nm to go — 09:23:32

2000 ft / 175 knots / Gear down / Flaps 15 / 6 nm to go

1300 ft / 150 knots / Gear down / Flaps 15 / 4 nm to go — 09:24:04 / 09:24:13

900 ft / 144 knots / Gear down / Flaps 40 / 3 nm to go

"Heading 265"

"Heading 210"

LOC capture

G/S capture

*The approach to 18R by TK1951 in summarized form. It illustrates how the approach may not have been experienced as "rushed." On the contrary, as an example, the crew realizes that the 210 heading they are given will make them intercept the localizer close-in, so they configure the aircraft ahead of it (gear down, flaps 15, slowing down to 175 knots).*

The figure here captures the rudimentary details of the flight, its distance and configuration at some of the critical junctures. As can be seen, there is a reduction in expected distance to go when TK1951 is turned onto a heading of 265. The crew is cleared to descend the aircraft to 2000 feet, and starts slowing it down to 195 knots, extending the flaps to 1. Compared to a typical nearing of an ILS approach in a B737-800, it appears that up to the moment of getting a heading of 210, there is no evidence of any "rush." Nearing the intercept heading with flaps 1, at 190 knots and 2000 feet (intercept altitude) is quite normal. The call by the Captain to Ground Ops at Amsterdam, coming while the aircraft is still on its base leg to the intercept to the ILS, could be considered late, but according to a post-accident interview with Ground Ops, neither the timing nor the content of that call was unusual.

It is the intercept heading, together with being kept at 2000 feet, that begins to compress things (there is no coordination with the crew about this, other than that the controller gives them the heading of 210 and clears them for the approach). The crew's responses to this clearance, however, show that they stay

106

ahead of the consequences. They lower the gear (while still being kept at 2000 feet and flying on the intercept heading), set flaps to 15 and slow down further to 175 knots. Rather than this being evidence of a rushed approach, it seems that the crew is anticipating and staying ahead of events (gear down and flaps 15 is normally done on final approach, when established on the localizer). Apparently realizing that the intercept heading will put them on a short final, the crew configures the aircraft ahead of localizer capture.

The aircraft captures the localizer with only one configuration step to go (flaps 40 and the concomitant final approach speed). This, at 5,3 nm from the threshold, is in itself not a problem, and not evidence of a rushed approach either. The only aspect that is problematic now is having been kept at 2000 feet. TK1951 is not "hot" because of the crew's actions, but they are still "high" because of air traffic control. This prompts the TK1951 crew to select V/S mode to capture the glideslope from above. As they do this, they not only need to monitor vertical speed and check glideslope capture, but are also required to coordinate with approach control and the tower controller for a handover and a landing clearance. This requires engine power to go to idle, it increases the airspeed slightly again, and uses up about 750 feet in altitude and about 1,5 nm of their final approach path.

By the time TK1951 captures the glideslope, they are really neither "hot" nor "high." Airspeed is slowing down to the target, they are on the glide path and only have the final flap setting and its concomitant speed reduction to go (and are still above 1000 feet). They have already been cleared to land. The thing that the capturing of the glideslope from above did, however, was push the completion of the landing checklist further down the approach. The onset of the landing checklist is further delayed for a few seconds by (likely) the instructional context of the flight (see above) where the F/O at first does not call for the landing checklist until prompted by the Captain. This drives the accomplishment of the landing checklist into the altitude window between about 900 and 500 feet. From the data available, this still does not give the Captain the experience of a "rushed" approach. He counsels the F/O, after all, to take the landing checklist items "one by one" without rushing through them with the risk of skipping things.

Was this a "rushed" approach? Up to receiving the intercept heading of 210, things looked entirely normal. After receiving that heading, crew actions (gear, flaps, speed) represented testimony to them staying ahead of how things were developing; there was no evidence of the crew getting rushed or falling behind. The capturing of the glideslope from above, which was successfully done by about 1300 feet, put the aircraft on glide and slowing down to target speed, with only one configuration step to go and with landing clearance already

received. While it necessitated the idling of engine power, that too is no strong cue. None of this is persuasive evidence of a "rushed" approach; the aircraft was neither "hot" nor "high" at that point. Again, the main consequence was the delay in accomplishing the landing checklist, which was not done above 1000 feet. Did this make the approach feel rushed? It does not appear so from the response of the Captain ("one by one"). But it may leave some with the question why the crew did not go around. This is dealt with under the next heading.

## Why not make a go-around?

In hindsight, one can ask why the crew didn't make a go-around when they realized that they would not be able to complete the landing checklist before 1000 feet. With an approach-and-landing accident such as TK1951, this question is as easy to ask as it is difficult to answer.

Continuing an approach against written guidance (particularly when events are considered in hindsight) is not a problem unique to TK1951. The Flight Safety Foundation sponsored a study in the late 1990's to analyze the factors that play into approach and landing accidents (Khatwa & Helmreich, 1999), and concluded in part, predictably, that executing a missed approach is one of the best investments that pilots can make in safety to prevent approach and landing accidents. Such advice should not be confused with an explanation for why many crews do not do so. In fact, getting crews to execute go-arounds, particularly in cases of unstabilized approaches, remains one of the most vexing problems facing most chief pilots and managers of flight operations across the world. Characterizations such as "*press-on-itis*" (Khatwa & Helmreich, 1999) do little to explain why crews press on; such words really only label a very difficult problem differently without offering any deeper understanding.

Concomitant recommendations directed at flight crews are, as a result, difficult to implement. For example, one recommendation is that "flight crews should be fully aware of situations demanding a timely go-around" (Khatwa & Helmreich, 1999, p. 53). Even if crews can be shown to possess such knowledge in theory (e.g. they are able to recite the criteria for a stabilized approach from the airline's Operations Manual), becoming aware that a timely go-around is demanded hinges precisely on a particular awareness of the situation itself. The data from continued approaches suggest that crews do not primarily interpret situations in terms of stabilized approach criteria, but in terms of their ability to continue the approach. Soft and hard gates (e.g. 1000 feet, 500 feet), when set in context of the end of a flight at a busy, major airport on a scheduled flight become norms against which to plan and negotiate the actual approach vis-à-

vis the air traffic and weather situation, not iron-fisted stop rules for that approach.

Thus, there seems little mileage in just reminding crews of the situations that demand a go-around, as almost all crews are able to talk about such situations and offer advice about how to do the right thing—until they are players in a rapidly unfolding situation themselves. When on approach themselves, it is not primarily generic criteria that crews see against which they can make some cognitive calculus of the legality and wisdom of continuing their approach. When on an approach themselves, crews see a situation that still looks doable, a situation that looks like they can make it, a situation that may hold a useful lesson for their LIFUS student in the right seat, a situation that suggests things will be all right before passing over the threshold (see Orasanu & Fischer, 1997). This is entirely consistent with decades of human factors research (e.g. Simon, 1957; Rasmussen, 1997). Operational decisions are not based on a "rational" analysis of all parameters that are relevant to the decision. Instead, the decision, or rather a continual series of assessments of the situation, is focused on elements in the situation that allow the decision maker to distinguish between reasonable options. The psychology of decision making is such that a situation is not assessed in terms of all applicable criteria (certainly not quantitative ones), but in terms of the options the situation appears to present (Simon, 1957; Reason, 1990; Klein, 1998).

It could be that there is, in this regard, a possible difference between pilots with a civilian versus a military background. One source at THY suggested that pilots with a military background (the Captain and the F/O of THY1951 had military backgrounds, the safety pilot did not) could be indoctrinated, experienced and trained to be more mission-oriented, that is, retaining a focus on accomplishing the flight even with minor adversities or problems (such as a late glideslope capture). It is impossible to say whether this mission-orientation, such as it may be, played a role in how the pilots of THY1951 saw and interpreted the doability of the unfolding situation, and indeed, human factors research from a number of decades strongly points to the generality of this issue, not its dependence on either personality or background.

Regardless of pilots' background, one promising countermeasure seems to be not to remind crews of the criteria for a stabilized approach, but to offer generic rewards for all the cases in which crews execute a missed approach. Chief pilots or managers of flight operations who offer a no-questions-asked reward (e.g. a bottle of wine) to crews who make a go-around, generally report modest success in reducing the number of unstabilized approaches. It is crucial, in setting up such policies, to communicate to crews that breaking off an approach is not only entirely legitimate, but actually desired: getting pilots to buy into the

idea that "each landing is a failed go-around." The handing out of such rewards should be advertised broadly to everybody else. Such encouragement is of course difficult to uphold in the face of production- and economic pressures and incredibly easy to undermine by sending subliminal messages (or more overt ones) to crews that on-time performance or cost/fuel savings are important. The paying of bonuses for on-time performance, which has been, and still is, custom in some airlines, is an obvious way to increase the likelihood of unstabilized approaches that are not broken off. To be sure, THY had and has no such bonuses in place, and in fact, several pilots attested in interviews to the fact that go-arounds have become an entirely accepted and expected category of events in the airline (in a shift from how they might have been seen a decade ago).

An interesting line of research has come from NASA Ames Research Center (Orasanu *et al.*, 2003). A phenomenon which Judith Orasanu and her colleagues called "plan continuation" captures a considerable amount of the data available from cases where an approach was continued despite cues that, in hindsight or in written guidance, pointed to the wisdom of a go-around. In fact, three out of four cases in which crews made tactical decisions that turned out erroneous in hindsight fit the plan-continuation pattern. The NASA research takes as its starting point the psychology of decision-making consistent with the last decades of research into it (e.g. Simon, 1957; Tversky & Kahneman, 1974; Zsambok & Klein, 1997). Decision-making in complex, dynamic settings such as an approach is not an activity that involves a weighty comparison of options against pre-specified criteria. Rather, such decision-making is "front-loaded:" this means that most human cognitive resources are spent on assessing the situation and then re-assessing it for its continued do-ability. In other words, decision-making on an approach is hardly about making decisions, but rather about continually sizing up the situation. The "decision" is often simply the outcome, the automatic by-product of the situation assessment. This is what turns a go-around decision into a continually (re-)negotiable issue: even if the decision to go around is not made on the basis of an assessment of the situation *now*, it can be pushed ahead and be made a few or more seconds later when new assessments of the situation have come in.

Even more important than the cognitive processes involved in decision making, are the contextual factors that surround a crew at the time (Orasanu *et al.*, 2003). The order in which cues about the developing situation come in, and their relative persuasiveness, are two key determinants for plan continuation. Conditions often deteriorate gradually and ambiguously, not precipitously and unequivocally. In such a gradual deterioration, there are almost always strong initial cues that suggest that the situation is under control and can be continued without increased risk. This sets a crew on the path to plan continuation.

Weaker and later cues that suggest that another course of action could be safer then have a hard time dislodging the plan as it is being continued, and as evidenced from the situation as it has so far been handled.

TK1951 seems to match these aspects of plan continuation. Strong initial cues suggested doable weather conditions for the training flight with ceilings way in excess of Cat I criteria and not a lot of wind. Other traffic ahead of and around TK1951 was streaming into the airport without problems. By 09:20:35, or less than minute after being turned onto 265°, the aircraft had slowed down to the target speed of 220 knots. Still on its nominal base leg, it leveled off at 2000 feet. At 09:21:08, there was a sound of a person whistling in the cockpit, most likely the Captain. This could be a sign that the situation was judged to be under control and even in its continuation entirely manageable (and indeed, there are or were no persuasive cues that suggest that it would not be). While still level at 2000 feet, the aircraft was slowed down to 195 knots for the next flap setting, again strongly confirming that the approach was doable.

Weak cues that began to suggest that the approach was about to become more hurried included the intercept turn that made for a late *and* shallow intercept to the localizer for 18R. The selection of V/S, another such possible cue, resulted in capturing the G/S rather than in a deterioration of the situation, once again not convincing the crew that the situation was unmanageable (in fact, showing the opposite). The further slowing down of the aircraft and the completion of its landing configuration showed the same. Passing the 1000ft gate without the aircraft proper configuration did not serve as a strong cue that would have suggested an abandonment of the approach. It hasn't with many crews other crews either (Khatwa & Helmreich, 1999). What mattered *in situ* was the series of judgments about the continued do-ability of the unfolding situation (Orasanu *et al.*, 2003) , judgments that the Captain/instructor would likely have made on a continual basis, as would any other Captain/instructor.

Note how plan continuation is different from a characterization of "confirmation bias." Confirmation bias suggests that crews seek out the evidence that supports (or confirms) their hypothesis of the situation, at the expense of other evidence. In TK1951, as in most cases, there is little that suggests the crew or the Captain was actively avoiding evidence that spoke against the plan as it was being continued. Indeed, the "bias" in confirmation bias seems to be produced almost exclusively in the mind of the retrospective observer, the one calling it a "confirmation bias," rather than in the mind of the person observed. Once again, it is hindsight that endows certain indications with a particular salience over others—a hindsight interpretation against which observed performance can be judged to have been "biased." This, of course, is hardly a meaningful conclusion: only hindsight would have shown which cues were more important

than others, and people inside the situation didn't have hindsight, so cannot meaningfully be judged to have been "biased" relative to it.

Finally, who should have made the decision to go around? Given the social, functional and hierarchical context of this crew composition, this becomes a peculiar question. Of course, the Captain as responsible for the safety of the flight would be an obvious crewmember to do so, but initiating a go-around may have conflicted with his other goals, e.g. pedagogical ones.

The Safety Pilot would be another reasonable candidate for initiating a go-around (at least verbally), as it would seem compatible with the role (which is to observe the flight training and advising the PIC of any irregularities that are noted). The THY Operations Manual (Part-D Training, p. 2-8-7), however, also makes it clear that the Safety Pilot is instructed by the PIC about when and how to intervene (even before the flight has commenced), which clearly would take some initiative out of the Safety Pilot mandate. Nonetheless, the Operations Manual also states, and pilots at THY confirm this, that the Captain has the obligation to obey anybody's call for a go-around (including the Safety Pilot's), and then ask questions later.

The F/O, and trainee, would be the third candidate for initiating a go-around, but there is evidence from the record of TK1951 that in the last minutes he may have fallen somewhat behind the airplane as the sequence of events was quickening after the tight turn-in for the localizer. For example, he was prompted by the Captain that the landing checklist was not yet completed. This, if anything, would make it unlikely that the F/O would have been accurately calibrated with respect to whether the criteria for a stabilized approach had been met or not, denying him a basis of confidence on which to initiate a missed approach. That, combined with his role as student and lowest-experience pilot in the cockpit, and with him never having been at AMS before, may have made the initiation of a go-around on his part unlikely.


## Intervention and attempted recovery

Being both a Captain, and finally responsible for the safety of the flight, and a flight instructor, responsible for the pedagogical progress of the student pilot represents a goal conflict. The job is one of managing the flight (even when that means trying to anticipate and understand how someone else, the student, is managing the flight), while also managing somebody's training. Pedagogy needs to be balanced against safety. This goal conflict is articulated most acutely in the decision whether and when to intervene in how the student is flying the aircraft (or, indeed, in how any student is operating any system). Intervention decisions

are notoriously difficult (e.g. Sheridan, 1987; Moray, Lee & Hiskes, 1994; Dekker & Woods, 1999). Intervention decisions are almost impossible to time right:

- Either the intervention comes too early. It is based on only partial or incomplete evidence that trouble may possibly or likely occur if no intervention happens. The intervention then eradicates the evidence of its own necessity. This represents a sacrifice to pedagogical goals (which can, in the long run, even have safety consequences too).

- Or the intervention comes too late, in which case trouble may have escalated, and the need for an intervention has become obvious. Process margins may have become eroded by this time, making a successful intervention less likely.

An intervention decision isn't typically a "decision" in the traditional sense (i.e. gather evidence, weigh options, compare outcomes and go for the one with the greatest benefit). Rather, intervention decisions in flight instruction seem matched much more closely to the models known from naturalistic decision making research (just like in the go-around "decision" discussed above), where instructor cognitive resources are spent on assessing the situation, gauging the student's evolving understanding of the situation, and then re-assessing the situation for its continued do-ability. The intervention decision, accordingly, is hardly a decision, but rather the outcome of continually sizing up the situation as well as somebody else's (the student's) evolving understanding of that situation. The whistling heard on the CVR could be an interesting indication in that direction (where the instructor could think something along the lines of "okay, let's see what the student makes of it here").

Given the continuous nature of the assessment of the do-ability of the situation, flight instruction typically involves a gradation of intervention tactics. The instructor may begin by asking a question ("Are you going to slow down yet?"), then offer suggestions ("Yes, not in checklist completed yet"), help with certain things or subtasks (e.g. (likely) the selection of V/S) and finally take over altogether ("I have"). This gradation can help an instructor balance safety and pedagogical goals, trying to achieve both as much as possible against the background of the availability of margins. It is interesting to note that as the approach progresses, time compresses and margins shrink, the Captain of TK1951 gradually shifts his intervention tactics, from asking a question to becoming increasingly (and finally totally) active in manipulating the aircraft controls. This is an expected escalation of intervention options that mimics the progression, pacing and leeway available during an approach to a major airport like AMS.

Providing generic guidance on when and how to intervene has proven almost impossible (e.g. Dekker & Woods, 1999; CAA, 2004). Instructor training may involve the discussion of a number of scenarios that allow new instructors to think around various options and how the progress of a flight or phase of flight may exclude certain options while opening up or demanding others. Such discussions, however, hardly happen in any extended way during instructor training or subsequent instructor meetings at the airline level. Indeed, interviews with line instructors at THY revealed that their monthly instructor meetings have not dealt with the issue of the "intervention decision" generically, and that this is believed to be something that really should fall earlier in basic instructor training.

The counsel to the F/O to pace his landing checklist calls (between 900 and 500 feet) is an indication that there is no awareness of the brewing trouble and quickly eroding margin, something that is consistent with the automation surprise research (e.g. Sarter, 1997). The Captain would probably have expected the F/O to disengage the A/T and A/P by the time the checklist was completed and the field was in sight and for the F/O to then land the aircraft manually on 18R. Flaps were at 40, speed would be Vref plus 5. When the stickshaker did go off at 09:25:47.5 (at about 450 feet) and airspeed turned out to have become something quite different than expected, it came as the kind of (automation) surprise that the research predicts for that type of situation (Sarter, Woods & Billings, 1997). It matches the 1996 FAA report's observation that,

> "…contrary to the belief of many flightcrews, some autoflight systems will take the airplane outside of the normal flight envelope (e.g., speed below stall warning speed or above the maximum operating limit speed), or attempt maneuvers that would not be expected of a human pilot. These characteristics can have potentially hazardous consequences, especially if the flightcrew is unaware of them." (p. 35)

The intervention decision (not in what the F/O was doing, but rather in what the automation had been doing) and the sequence of events for the final seconds of TK1951 then looked as follows:

| | | |
|---|---|---|
| 09:25:47.5 | stick shaker onset | |
| 09:25:48 | Power applied to throttles. Throttles halfway up, then back to idle. | |
| 09:25:49 | | |
| 09:25:50 | Captain: "*I have.*" | Safety pilot: "*Speed.*" |
| 09:25:51 | A/T OFF | |
| 09:25:52 | Full power applied | Safety pilot: "*100 knots.*" |
| 09:25:53 | A/P OFF | Safety pilot: "*Speed.*" |

09:25:54
09:25:55
09:25:56
09:25:57
09:25:58
09:25:59
09:26:00
09:26:01
09:26:02      End of recording

It is not clear who pushed the throttles forward at 09:25:48. The A/T was not disconnected when this is done, so the A/T system pulled the throttles back to idle. It is possible that the delay produced by this was contributory to the failure of the recovery attempt. As said earlier, it is interesting that the A/T was not disconnected, as it may confirm the nature and depth of the automation surprise. The A/T had gone and stayed in RETARD flare mode (which does not allow the crew to manually advance the throttles). As explained earlier in the report, however, a normally trained and experienced B737 crew has no basis on which to expect the A/T to go and stay in RETARD flare mode while in flight. They do know that the A/T can have idled the throttles but are in the full expectation that they can then manually advance them (which is possible in ARM mode).

The fact that the throttles pull back to idle by themselves could indeed represent a secondary automation surprise that can help explain the few seconds delay before the Captain announces that he has control of the aircraft, realizes that the A/T need to be disengaged, after which the autopilot is also disengaged. The subsequent recovery attempt is not successful. While (approach to) stall recovery in landing configuration is something that is practiced in flight training, even on the Boeing 737 (in the simulator), there is no specific guidance or training on what to do if such a recovery is to be carried out very close to the ground, with little or no margin below the airplane. In fact, when close to the ground, recovery from a stall or an approach to a stall involves the juggling of multiple conflicting goals and airplane responses.

The Boeing 737 Flight Crew Training Manual (chapter 7) describes how, for stall recovery, nose down pitch control must be applied (which uses up altitude) and maintained until the wings are unstalled. It also describes how, under certain conditions, it may be necessary to reduce thrust in order to prevent the angle of attack from continuing to increase (which is possible with high thrust on under-wing mounted engines such as on the B737). Reducing thrust obviously conflicts with the goal of building up airspeed quickly in a critical situation. Indeed, the manuals say that when ground contact is a factor, crews should

smoothly advance the thrust levers to maximum thrust and adjust pitch to avoid ground contact. Intermittent stick shaker should be used as the upper limit for pitch attitude for recovery when ground contact is a factor.

That the TK1951 crew was ultimately unsuccessful in recovering from the automation surprise is not inconsistent with research (see Gawron, Berman & Dismuouth, 2003). Indeed, the 2007 Boeing 737 incident at Bournemouth (AAIB, 2009) is also an example of the difficulty for pilots to perform stall recovery maneuvers in that aircraft in a serious situation. There is no way to establish whether the following would have applied to the TK1951 crew, but the FAA in 1996 noted its concern about pilots becoming

> "less confident in their own airmanship skills relative to the capabilities they perceive to be present in the automation, particularly in a stressful situation. In some cases, where this perception of the automation's capabilities is particularly inaccurate, it can have potentially hazardous consequences." (p. 35)

Similarly, Billings (1996) expressed concern about the policy of some carriers that requires their pilots to effectively fly at the highest level of automation possible, all the time (a policy that some carriers have to this day). In 1993 (and in fact throughout the 1990's), this was the topic of a brisk debate in the aviation literature, where some opined that "excessive reliance  on equipment to help pilots fly 'smarter and safer' has become institutionalized to the point of becoming dangerous". One pilot commented how he was "…admonished by the chief pilot for daring to hand-fly a raw-data standard instrument departure" (see Billings, 1996). The FAA Human Factors team concluded:

> "Based on the incident data, accident data, and pilot and operator input evaluated by the HF Team, we have concerns about pilot basic airmanship skills and general airmanship knowledge in several areas. One area is the degradation of manual flying skills of pilots who use automation frequently." (FAA, 1996, p. 103).

There is no evidence in the record of the accident that the use of automation during pilots' careers may have contributed to skill erosion in the case of TK1951. However, it is interesting to note that this does represent an industry-wide concern with possible consequences for how pilots might or might not be able to recover from automation-induced upsets or stalls.

# FINDINGS AND CONCLUSIONS

## Findings

The Flight Crew Human Factors Investigation into the TK1951 accident near AMS on the 25th of February, and detailed in this report, has produced the following findings:

1) During the descent towards Amsterdam, the TK1951 crew discussed and showed awareness of an anomaly of the (left) radio altimeter.
   TK1951 was kept at 2000 feet while being vectored to capture the localizer at less than 5,5nm from the 18R runway threshold at AMS. This put it above the glideslope. The crew appeared aware of the tight vectoring given to them by ATC (the landing gear was already down and the flaps were at 15 even before localizer intercept). No prior warning or coordination from ATC occurred, which would be normal and desirable with such a tight vectoring for approach.

2) Upon the crew's selection of vertical speed mode to capture the glideslope from above (as a result of the tight vectoring received from ATC), the 737's autothrottle (A/T) retarded to idle, consistent with crew expectations. The aircraft had to descend and simultaneously slow down to the next (flap 40) target speed. Upon selecting V/S mode, the A/T window of the Flight Mode Annunciator (FMA) on the Primary Flight Display (PFD) in the cockpit showed "RETARD."

3) The B737 has two RETARD modes that combine autothrottle and autopilot functions: (a) Retard flare and (b) retard descent. Retard descent commands the thrust levers to the aft stop to allow the autopilot to follow a planned descent path. Retard descent mode is normally followed by the ARM mode, in which the A/T protects the flight envelope and maintains a selected speed. ARM mode also allows crews to manually set the thrust levers forward again.

117

In contrast, the Retard flare mode is normally activated just prior to touchdown when an automatic landing is performed. The A/T does the retard part, the autopilot the flare part, so as to jointly make a smooth landing. In retard flare mode, the A/T no longer offers flight envelope protection, does not maintain any selected speed, and it will keep the thrust levers at the idle stop (or pull them back there if the crew would push them forward).

The A/T window on the FMA offers no way for a flight crew to distinguish one RETARD mode from the other.

4) While the A/T window would normally have shown "MCP SPD" upon selecting V/S mode (knowledge that a pilot flying his 17th leg on a B737 is unlikely to have ready at hand), the RETARD mode made aircraft behavior insidiously consistent with crew expectations. They needed to go down and slow down so as to capture the glideslope (from above) and get the aircraft's speed down to the speed for the next flap setting. The A/T announced that it did what the crew commanded it to do: it retarded, and aircraft behavior matched crew expectations: the aircraft went down, slowed down and then captured and started tracking the glideslope.

5) As it only showed "RETARD" (and not "FLARE"), the FMA annunciation gave the appearance as if the A/T went into RETARD descent mode. However, the A/T went automatically into the unexpected RETARD flare mode—not because the crew had selected V/S, but because a number of conditions had now been fulfilled, and the A/T was acting according to its own logic: the aircraft was going below 2000 feet RA (in fact, it was at -7 feet RA, according to its only available (and corrupted) input to the A/T system), the flaps were more than 12,5 degrees out and the F/D mode was no longer in ALT HOLD.

While the A/T had, in effect, decided it was time to land, FCC B was still commanding the F/D and Autopilot B to stay on glideslope. One part of the automation was doing one thing (landing), while the other part was doing something else (flying). The part that was landing (the A/T) had control over the airspeed, the part that was flying (Autopilot B) did not; it only tracked the descent path on the glideslope.

6) Based on their training and documentation, the TK1951 crew would have believed that they had protected their aircraft and its flight from any pre-existing problems with the left RA. The right autopilot (known as Autopilot B or CMD B) had been selected on, and the right Flight Control Computer (known as FCC B) was giving it inputs.

Boeing pilot training materials and documentation do not reveal that the autothrottle always gets its height information from the left Radio Altimeter,

that, on pre-2005 737NG models, it doesn't cross-check its RA input data with other RA data; and that the right RA does not provide input to the autothrottle—even when FCC B has been selected as Master and Autopilot B is flying (which was the case for TK1951).

7) The crew was completing their landing checklist during the thirty seconds when the airspeed decayed below the selected landing speed as a result of this automation mismatch. Interleaving task demands, speed tape design issues and the erosion of cognitive work surrounding the calculation of final approach speeds in automated airliners, becoming visual with the runway, and landing checklist design could all have interacted with the crew's attention during the 30 seconds of speed decay below approach speed.

8) There is no persuasive basis in the record to conclude that the approach was "rushed." The crew anticipated the late glideslope capture by lowering the gear and selecting flaps 15 even before capturing the localizer, and the only items to be completed after glideslope capture were final flap setting and the landing checklist. Landing clearance had already been obtained.

9) TK1951 fits human factors research on plan continuation. Decisions to go around or intervene in a student pilot's actions involve the assessment and re-assessment of the unfolding situation for its continued do-ability. The dynamic emergence of cues about the doability of the TK1951 approach, suggested to the crew that continuing was possible and not problematic.

10) A breakdown in CRM (Crew Resource Management) cannot be substantiated for TK1951. Other than artifacts of the instructional context of the flight, there is little to no evidence in the primary data source (the CVR) for overlapping talk, for second-pair part silences, or for other-initiated repair—three aspects of conversational interaction that have recently been implicated in CRM breakdowns. The Captain was well-liked, and a popular instructor at THY.

11) The length of B737 type training at THY, as well as procedural compliance at THY, appear to at least match industry standard. The Captain had close to 11,000 hours on the Boeing 737 alone. This combination of training standards and experience is apparently not enough to protect crews from the subtle effects of automation failures during automated, human-monitored flight. The documentation and training available for flight crews of the Boeing 737NG leaves important gaps in the mental model that a crew may build up about which systems and sensor inputs are responsible for what during an automatically flown approach.

12) TK1951 fits the substantial research base on automation surprises. For 70 seconds, automation and aircraft behavior were consistent with crew expectations (the A/T had insidiously reverted to an unexpected mode that seemed—and was annunciated—as if it followed crew instructions). After that period, the really difficult task for the crew of TK1951 was to discover that the automation was actually not (or no longer) following their instructions. This was discoverable not through a change in aircraft behavior (as it usually is during automation surprises), but through *a lack of change* of aircraft behavior (and a lack of mode change). The aircraft did not stop slowing down, and the automation did not change mode. The crew would have had to discover one or two non-events, in other words. Research shows that discovering non-events is very difficult, particularly when initial system behavior is consistent with expectations, when design does not show the behavior but only the status of the system, and when there is no basis in a crew's mental model to expect the non-event(s).

13) Believing, on the basis of their training, documentation and experience, that they had insulated their cockpit set-up from any problem with the left RA, the TK1951 flight crew was surprised by the automation when it turned out that it had not been keeping the airspeed. When they did notice as a result of the stick shaker, and tried to intervene, it was too late in this situation, for this crew, to recover.

14) Post-accident manufacturer recommendations that, in effect, tell flight crews to mistrust their machine and to stare harder at it not only mismatch decades of human factors and automation research, but also leave a single-failure pathway in place.

A large amount of scientific research and, perhaps even more importantly, studies sponsored and conducted by regulatory aviation safety agencies and independent aviation safety boards from across the world (FAA, 1996; BASI, 1998; CAA, 2004) have pointed for years to the insufficiency of automation training standards, the difficulty of relying on human monitors with normally very reliable automated systems, and the possibly devastating effects of subtle automation failures. TK1951 may have been a surprise for the aircrew involved; it can hardly come as a surprise to the industry.

## Conclusion

Back in 1996, the FAA Human Factors team recommended the creation of interim certification policy guidance, which would fill the gap until design and training guidance material would be complete. It said:

"Specifically, the HF Team believes that the following areas should be addressed by interim guidance:
Pilot/autopilot interactions that create hazardous out-of-trim conditions;
Autopilots that can produce hazardous energy states and may attempt maneuvers that would not normally be expected by a pilot; and
Improved airplane flight manual wording regarding the capabilities and limitations of the autopilot." (FAA, 1996, p. 98)

No results of this recommendation could be found in the training material given to the newly type-rated F/O of flight TK1951. Indeed, manufacturers clearly take another approach to this. Shortly after the accident, Boeing issued a bulletin to all 737 operators and announced that it "will warn crews about fundamentals like flying the aircraft, monitoring airspeed, [and] monitoring altitude" (Learmount, 2009).

The only defense against a designed-in single-failure path, in other words, are the pilots who are warned to mistrust their machine and to stare at it harder. Such a reminder, oriented only at the human operator in the system, is hardly credible after three decades of in-depth research into automated airliner flying and the subtle and pervasive ways in which automation on the flight deck (and particularly its subtle failure) affects human performance (e.g. Wiener & Curry, 1980, Sarter *et al.*, 1997). For flight crews of Boeing 737's, like the crew of TK1951, there is no sufficient training, no written guidance or documentation, and no likelihood of line experience that would insulate them from the kind of automation surprise that happened near Amsterdam on the 25th of February.

# REFERENCES

Air Accidents Investigation Branch (2009). *Report on the serious incident to Boeing 737-3Q8, registration G-THOF on approach to runway 26, Bournemouth airport, Hampshire on 23 September 2007* (Aircraft Accident Report no. 2/2009). London: Department for Transport, AAIB.

Air Transport Association (1989, April). *National plan to enhance aviation safety through human factors improvements*. Washington, DC: Author.

Bainbridge, L. (1983). Ironies of automation. *Automatica, 19*(6), 775-779.

Baker D. & Dismukes R. (2002). A Framework for Understanding Crew Performance Assessment Issues. *International Journal of Aviation Psychology, 12*(3), 205-222.

Bureau of Air Safety Investigation (1998). *Advanced technology aircraft safety survey report*. Canberra, ACT: BASI, Department of Transport and Regional Development.

Berman, B. A., & Reed, C. B. (2009). *Current state of pilot training for emergency/abnormal situations: A preliminary report*. Moffet Field, CA: SJSU/NASA Ames Research Center.

Byrd, K. S., Adams, K. A., Ntuen, C. A. (2002). *Visual attention loss function during automation monitoring task*. Department of Industrial and Systems Engineering, North Carolina A&T State University, Greensboro, North Carolina.

Christoffersen, K., Woods, D. D. (2003). *Making sense of change: Extracting events from dynamic process data* (Institute for Ergonomics/Cognitive Systems Engineering Laboratory Report ERGO-CSEL 01-TR-02). Columbus, OH: The Ohio State University.

Chow, R. Christoffersen, K., & Woods, D. D. (2000). A model of communication in support of distributed anomaly response and replanning. In *Proceedings of the IEA 2000/FHES 2000 Congress*, Human Factors and Ergonomics Society, July 2000.

Civil Aviation Authority (UK) (2004). *Flight crew reliance on automation* (CAA report no. 2004/10). Gatwick: CAA Safety Regulation Group (authored by S. Wood, Cranfield University).

Corwin, W. H. (1995). Understanding mode annunciation: What is the pilot's mental model? In R. S. Jensen & L. A. Rakovan (Eds.), *Proceedings of the Eighth International Symposium on Aviation Psychology* (pp. 249–253). Columbus: The Ohio State University.

Degani, A., & Wiener, E. L. (1990). *Human factors of flight deck checklists: The normal checklist* (Contractor Report NCC2-377). Moffett Field, CA: NASA Ames Research Center.

Degani, A., Heymann, M., & Shafto, M. (1999). Formal aspects of procedures: The problem of sequential correctness. In *Proceedings of the 43rd Annual Meeting of the Human Factors and Ergonomics Society*. Houston, TX: Human Factors Society.

Dekker, S. W. A. (2000). Crew situation awareness in high-tech settings: Tactics for research into an ill-defined phenomenon. *Transportation Human Factors, 2*, 49–62.

Dekker, S. W. A., & Hollnagel, E. (Eds.) (1999). *Coping with computers in the cockpit*. Aldershot, UK: Ashgate Publishing Co.

Dekker, S. W. A., & Woods, D. D. (1999). To intervene or not to intervene: The dilemma of management by exception. *Cognition, Technology and Work, 1*, 86-96.

Dornheim, M. A. (1995). Dramatic incidents highlight mode problems in cockpits. *Aviation Week and Space Technology, 142*(5), 57–59.

Dornheim, M. A. (1996). Airlines Improve Automation Training. *Aviation Week & Space Technology 145*(10), 128-129.

Dornheim, M. A. (1996, September 2). Airlines improve automation training. *Aviation Week & Space Technology, 145*(10), 128-129.

Dornheim, M.A. (1995). Dramatic Incidents Highlight Mode Problems in Cockpits. *Aviation Week and Space Technology*, (1/30/95), 6-8.

Eldredge, D., Dodd, R.S., and Mangold, S.J. (1991). *A Review and Discussion of Flight Management System Incidents Reported to the Aviation Safety Reporting System*. (Battelle Report, prepared for the Department of Transportation). Columbus, OH: Volpe National Transportation Systems Center.

Flin, R. & Martin, L. (2001). Behavioral Markers for Crew Resource Management: A Review of Current Practice. *International Journal of Aviation Psychology 11*(1), 95-118.

Fox, J., Merwin, D., Marsh, D., McConkie, G., & Kramer, A. (1996). Information extraction during instrument flight: An evaluation of the validity of the eye–mind hypothesis. In *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting* (pp. 77–81). Santa Monica, CA: Human Factors Society.

Funk, K. H., Lyall, E. A., & Niemczyk, M. C. (1997). Flightdeck automation problems: Perceptions and reality. In M. Mouloua & J.M. Koonce (Eds.),

*Human–automation interaction: Research and practice* (pp. 29–34). Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Gawron, V., Berman, B., & Dismukes R. K. (2003, July-August). New airline pilots may not receive sufficient training to cope with airplane upsets. *Flight Safety Foundation's Flight Safety Digest*, pp. 19-32.

Gibson, J. J. (1979). An ecological approach to perception. Boston, MA: Houghton Mifflin.

Goldsmith, T. & Johnson, P. (2002). Assessing and Improving Evaluation of Aircrew Performance. *International Journal of Aviation Psychology, 12*(3), pages 223-240.

Goteman, Ö. E., & Dekker, S. W. A. (2007). Flight crew callouts and aircraft automation modes. *International Journal of Applied Aviation Studies, 6*(2), 235-248.

Group (2002). Developing a Method for Evaluating Crew Resource Management Skills: A European Perspective. *International Journal of Aviation Psychology 12*(3), 263-285.

Heft, H. (2001). *Ecological psychology in context: James Gibson, Roger Barker and the legacy of William James's radical empiricism*. Mahwah, NJ: Lawrence Erlbaum Associates.

Helmreich, R., Merritt, A., & Willhelm, J. (1999). *The Evolution of Crew Resource Management Training in Commercial Aviation*. (University of Texas)

Huettig, G., Anders, G., & Tautz, A. (1999). Mode awareness in a modern glass cockpit attention allocation to mode information. In R. Jensen (Ed.), *Proceedings of the 1999 Ohio State University Aviation Psychology Conference*. Dayton: Ohio State University.

International Civil Aviation Organization (1995). *Operational implications of automation in advanced technology flight decks* (Human Factors Digest no. 5). Montréal: ICAO.

Jonides, J. and Yantis, S. (1988). Uniqueness of Abrupt Visual Onset in Capturing Attention. *Perception and Psychophysics, 43*(4), 346-354.

Klein, G. (1998). *Sources of power*. Cambridge, MA: MIT Press.

Klinect, J. (2005). *Line Operation Safety Audits (LOSA)*: *Definition and operating characteristics*. Austin, TX: University of Texas.

Learmount, D. (2009). Pilots in the dock. Flight International, p. 7.

Loukopoulos, L. D., Dismukes, R. K., & Barshi, I. (2009). *The multitasking myth: Handling complexity in real-world operations*. Aldershot, UK: Ashgate Publishing Co.

Lyall, B. (1998, July). *Autoflight mode awareness issues: An overview*. Paper presented at the FAA Mode Awareness Workshop, Seattle, WA.

Mathews, G., Davies, M. & Holley, P.J. (1993). Cognitive predictors of vigilance. *Human Factors*, 35, 3-24.

Maurino, D.E., Reason, J., Johnston, N., & Lee, R.B. (1995). *Beyond aviation*

*human factors: Safety in high technology systems*. Aldershot, UK: Avebury Aviation.

Mayer, E. R. (2000). *Human Monitoring of Automation Failure and Recovery Performance*, MSIE Thesis, Department of Industrial and Systems Engineering, North Carolina A&T State University, Greensboro, North Carolina

Mindell, D. A. (2008). *Digital Apollo*. Cambridge, MA: MIT Press.

Moll van Charante, E., Cook, R.I., Woods, D.D., Yue, L., and Howie, M.B. (1992). Human-Computer Interaction in Context: Physician Interaction with Automated Intravenous Controllers in the Heart Room. In H. G. Stassen, editor, *Analysis, Design and Evaluation of Man-Machine Systems*, Pergamon Press, 1993, p. 263-274.

Moray, N. & Rotenberg, I., (1989). Fault management in process control: Eye movements and action. *Ergonomics*, 32, 1139-1147.

Moray, N. (1986). Monitoring Behavior and Supervisory Control. In K. R. Boff, L. Kaufman, and J. P. Thomas (Eds.), *Handbook of Perception and Human Performance* (Vol. 2, Chapter 40). New York: Wiley.

Moray, N., Lee, J., & Hiskes, D. (1994). Why do people intervene in the control of automated systems? In *Proceedings of the first automation technology and human performance conference*, Washington, DC, April 1994.

Mumaw, R. J., Sarter, N. B., & Wickens, C. D. (2001). Analysis of pilots' monitoring and performance on an automated flight deck. In Proceedings of the Eleventh International Symposium in Aviation Psychology (p. 6). Columbus: Ohio State University.

National Transportation Safety Board (1984). *Scandinavian Airlines DC-10-30, J. F. Kennedy Airport, New York, 2/28/84* (NTSB Rep. No. AAR-84-15). Washington, DC: Author.

Nevile, M. & Walker, M. B. (2005). *A context for error: Using conversation analysis to represent and analyse recorded voice data* (Aviation Research Report B2005/0108). Canberra, ACT: Australian Transport Safety Bureau.

Nevile, M. (2002). Coordinating talk and non-talk activity in the airline cockpit. *Australian Review of Applied Linguistics*, 25(1), 131-146.

Nevile, M. (2004). *Beyond the black box: Talk-in-interaction in the airline cockpit*. Aldershot, UK: Ashgate Publishing Co.

Nikolic, M. I., Orr, J., & Sarter, N. B. (2001). The effects of display context on the effectiveness of visual onsets for attention capture. *Proceedings of the 20th IEEE Conference on Digital Avionics Systems, DASC*, Volume 1, pp. 5A3/1-5A3/7.

Nikolic, M. L.,&Sarter, N. B. (1999). Peripheral visual feedback in support of mode awareness and attention allocation on highly automated flight decks source. In R. S. Jensen, B. Cox, J. D. Callister, & R. Lavis (Eds.), *Proceedings of the Tenth International Symposium on Aviation Psychology* (pp. 111–116). Columbus: Ohio State University.

Norman, D. A. (1988). *The Psychology of Everyday Things*. New York, Basic Books.

Norman, D. A. (1990). The 'Problem' with Automation: Inappropriate Feedback and Interaction, not 'Over-Automation'. *Philosophical Transactions of the Royal Society of London*, B 327, 585-593.

O'Connor, P., Hörmann, H., Flin, R., Lodge, M., Goeters, K., The JARTEL

Orasanu, J., & Fischer, U. (1997). Finding decisions in natural environments: The view from the cockpit. In C. Zsambok & G. Klein (Eds.), *Naturalistic decision making* (pp. 343-357). Mahwah, NJ: Erlbaum.

Orasanu, J., Martin, L., & Davison, J. (2003). Cognitive and contextual factors in aviation accidents: Decision errors. In E. Salas & G. Klein (Eds.), *Applications of naturalistic decision making*. Mahwah, NJ: Erlbaum.

Orlady, H., & Orlady, L. (1999). *Human Factors in Multi- Crew Flight Operations*. Aldershot, UK: Ashgate Publishing Co.

Patterson, E. S., & Woods, D. D. (2001). Shift changes, updates and the on-call model in space shuttle mission control. *Computer Supported Collaborative Work, 10*, 317-346.

Perry, S. J., Wears, R. L., Cook, R. J. (2005). The role of automation in complex system failures. *Journal of Patient Safety, 1*(1), 56-61.

Predmore, S.C. (1991). Microcoding of communications in accident investigation: Crew coordination in United 811 and United 232. In Jensen, R.S. (Ed.), *Proceedings of the Sixth International Symposium of Aviation Psychology*, 350-355. Columbus, OH: Ohio State University.

Rasmussen, J. (1997). Merging paradigms: Decision making, management, and cognitive control. In R. Flin, E. Salas, M. Strub, & L. Martin (Eds.), *Decision making under stress: Emerging themes and applications* (pp. 67-81). Aldershot, UK: Ashgate.

Reason, J. T. (1990). *Human error*. Cambridge, UK: Cambridge University Press.

Reason, J. T. (1997). *Managing the risks of organizational accidents*. Aldershot, UK: Ashgate Publishing Co.

Rohloff, K. R. (2005). Sensor Failure Tolerant Supervisory Control. *Proceedings of the 44th IEEE conference on Decision and Control 2005 and the 2005 European Control Conference, CDC-ECC*, pp. 3493-3498.

Roscoe, S. N. (1968). Airborne displays for flight and navigation. *Human Factors, 10*, 321-332.

Roscoe, S. N., Corl, L., & Jensen, R. S. (1981). Flight display dynamics revisited. *Human Factors, 23*, 341-353.

Salas, E., Wilson, K. A., Burke, C. S., & Wightman, D. C. (2006). Does Crew Resource Management Training work? An update, an extension and some critical needs. *Human Factors, 48*(2), 392-412.

Sarter N. B., Woods D. D., & Billings, C. E. (1997). Automation surprises. In G. Salvendy (ed) *Handbook of Human Factors/Ergonomics*. New York: Wiley.

Sarter, N. B. (1991) The flight management system-pilots' interaction with cockpit automation. *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Decision Aiding for Complex Systems, Volume 2*. pp. 1307-1310 vol.2

Sarter, N. B. (1995). "Knowing when to look where": Attention allocation on advanced automated flight decks. In R. S. Jensen & L. A. Rakovan (Eds.), *Proceedings of the Eighth International Symposium on Aviation Psychology* (pp. 239–242). Columbus: Ohio State University.

Sarter, N. B. (1997). 'Strong and silent' machine agents: in need of improved communication and coordination skills. *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, Volume 2*, pp. 1163-1167.

Sarter, N. B. and Woods, D. D. (1992). Pilot Interaction with Cockpit Automation: Operational Experiences with the Flight Management System. *International Journal of Aviation Psychology, 2*(4), 303-321.

Sarter, N.B. and Woods, D. D. (1994). Pilot Interaction with Cockpit Automation II: An Experimental Study of Pilots' Model and Awareness of the Flight Management and Guidance System. *International Journal of Aviation Psychology, 4*(1), 1-28.

Sarter, N.B. and Woods, D.D. (1995b). How in the World Did We Ever Get Into That Mode? Mode Error and Awareness in Supervisory Control. *Human Factors, 37*(1), 5-19.

Shepherd, M., Findlay, J.M., & Hockey, G.R.J. (1986). The relationship between eye movements and spatial attention. *Quarterly Journal of Experimental Psychology*, 38, 475-491.

Sheridan, T. B. (1987). Supervisory control. In G. Salvendy (ed). *Handbook of human factors*. New York: Wiley.

Sheridan, T. B. (1995). Human centered automation: oxymoron or common sense? *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics, 1995. 'Intelligent Systems for the 21st Century'*, Volume 1, pp. 823-828.

Silverman, D. (1998). *Harvey Sacks: Social science and conversation analysis*. Oxford, UK: Oxford University Press.

Simon, H. (1957). *Models of man: Social and rational*. New York: Wiley.

Thomas, M. (2004). Predictors of Threat and Error Management: Identification of Core Non-technical Skills and Implications for Training Systems Design. *International Journal of Aviation Psychology 14*(2), 207–231.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty. *Heuristic and Biases Science, 185*, 1123-1124.

Vakil, S. S., Hansman, R. J., Midkiff, A. H., & Vaneck, T. (1995). Mode awareness in advanced autoflight systems. In T. B. Sheridan (Ed.), *Analysis, design and evaluation of man–machine systems* (Vol. 1, pp. 371–376). Oxford, England: Pergamon.

van Avermaete, J. (1998). *Non-technical skill evaluation in JAR-FCL*. (Report TP-98518 from the Dutch National Aerospace Laboratory). Amsterdam, NL: NLR.

Wiener, E. L. (1989). *Human factors of advanced technology (glass cockpit) transport aircraft* (NASA contractor report No. 177528). Moffett Field, CA: NASA Ames Research Center.

Wiener, E. L., & Curry, R. E. (1980). *Flight deck automation: Promises and problems* (NASA Tech. Mem. 81206). Moffett Field, CA: NASA Ames Research Center.

Wiener, E. L., Chute, R. D., & Moses, J. H. (1999). *Transition to glass: Pilot training for high-technology transport aircraft* (NASA contractor report No. 208784). Moffett Field, CA: NASA Ames Research Center.

Wiener, E., Kanki, B., & Helmreich, R. (1993). *Cockpit Resource Management*. San Diego, CA: Academic Press.

Wiener, E.L. (1993). Crew Coordination and Training in the Advanced-Technology Cockpit. In E.L. Wiener, B.G. Kanki, and R.L. Helmreich (Eds.), *Cockpit Resource Management* (pp. 199-223), Academic Press: San Diego.

Wiener, E.L. and Curry, R.E. (1980). Flight-deck automation: promises and problems. *Ergonomics, 23*(10), 995-1011.

Winograd, T. and Flores, F. (1986). *Understanding computers and cognition*. Reading, MA: Addison-Wesley Publishing Company.

Woods, D. D. (1994). *Representation design in the computer medium*. Columbus, OH: Cognitive Systems Engineering Laboratory, The Ohio State University.

Woods, D. D., & Sarter, N. B. (2000). Learning from automation surprises and going sour accidents. In N. B. Sarter & R. Amalberti (Eds.), *Cognitive engineering in the aviation domain*. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Woods, D. D., Dekker, S. W. A., Cook, R. J., Johannesen, L., Sarter, N. B. (2009). *Behind human error*. Aldershot, UK: Ashgate Publishing Co. (in press).

Zsambok, C., & Klein, G. (Eds.). (1997). *Naturalistic decision making*. Mahwah, NJ: Erlbaum.