SCHOOL OF AVIATION
Lund University

Reconstructing human contributions to accidents: The new view on error and performance

Sidney W. A. Dekker

Technical Report 2001-01

Lund University School of Aviation
Address: 260 70 Ljungbyhed, Sweden
Telephone: +46-435-445400
Fax: +46-435-445464
Email: research@tfhs.lu.se

Abstract

Problem

How can we reconstruct the human contribution to accidents? Investigators easily take the position of retrospective outsider, looking back on a sequence of events that seems to lead to an inevitable outcome, and pointing out where people went wrong. This does not explain much, however, and may not help prevent recurrence.

Method and results

In this paper I examine how investigators can reconstruct the human contribution to accidents in light of what has recently become known as the new view of human error. The commitment of the new view is to relocate controversial human assessments and actions back into the flow of events of which they were part and which helped bring them forth, to see why assessments and actions made sense to people at the time. The second half of the paper is dedicated to one way in which investigators could begin to reconstruct people's unfolding mindsets.

Impact on industry

In an era where a large portion of accidents gets attributed to human error, it is critical to understand why people did what they did, rather than judging them for not doing what we now know they should have done. This paper contributes by helping investigators avoid the traps of hindsight, and by presenting a method with which investigators can begin to see how people's actions and assessments could actually have made sense at the time.

Keywords: human error, investigations, accidents, hindsight, reconstruction

Introduction

In human factors today there are basically two different views on human error and the human contribution to accidents . One view, recently dubbed "the old view" (AMA, 1998; Reason, 2000), sees human error as a cause of failure. In the old view of human error:

- Human error is the cause of most accidents.
- The engineered systems in which people work are made to be basically safe; their success is intrinsic. The chief threat to safety comes from the inherent unreliability of people.
- Progress on safety can be made by protecting these systems from unreliable humans through selection, proceduralization, automation, training and discipline.

The other view, also called "the new view", sees human error not as a cause, but as a symptom of failure (Rasmussen & Batstone, 1989; Woods *et al.*, 1994; AMA, 1998; Reason, 2000; Hoffman & Woods, 2000). In the new view of human error:

- Human error is a symptom of trouble deeper inside the system.
- Safety is not inherent in systems. The systems themselves are contradictions between multiple goals that people must pursue simultaneously. People have to create safety.
- Human error is systematically connected to features of peoples tools, tasks and operating environment. Progress on safety comes from understanding and influencing these connections.

The new view of human error represents a substantial movement across the fields of human factors and organizational safety (Reason, 1997; Rochlin, 1999) and encourages the investigation of factors that easily disappear behind the label "human error"—long-standing organizational deficiencies; design problems; procedural shortcomings and so forth. The rationale is that human error is not an explanation for failure, but instead demands an explanation, and that effective countermeasures start not with individual human beings who themselves were at the receiving end of much latent trouble (Reason, 1997) but rather with the error-producing conditions present in their working environment. Most of those involved in accident research and analyses are proponents of the new view. For example:

"...simply writing off ... accidents merely to (human) error is an overly simplistic, if not naive, approach.... After all, it is well established that accidents cannot be attributed to a single cause, or in most instances, even a single individual." (Shappell & Wiegmann, 2001, p. 60).

However, our willingness to embrace the new view of human error in our analytic practice is not always matched by our ability to do so. When confronted by failure, it is easy to retreat into the old view. We seek out the "bad apples" and assume that with them gone, the system will be safer than before. An investigation's emphasis on proximal causes ensures that the mishap remains the result of a few uncharacteristically ill-performing individuals who are not representative of the system or the larger practitioner population in it. It leaves existing beliefs about the basic safety of the system intact.

> The pilots of a large military helicopter that crashed on a hillside in Scotland in 1994 were found guilty of gross negligence. The pilots did not survive—29 people died in total—so their side of the story could never be heard. The official inquiry had no

> problems with "destroying the reputation of two good men", as a fellow pilot put it. Potentially fundamental vulnerabilities (such as 160 reported cases of Uncommanded Flying Control Movement or UFCM in computerized helicopters alone since 1994) were not looked into seriously (Sunday Times, 25 June 2000).

Faced with a bad, surprising event, we seem more willing to change the players in the event (e.g. their reputations) than to amend our basic beliefs about the system that made the event possible. To be sure, reconstructing the human contribution to a sequence of events that led up to an accident is not easy. As investigators we were seldom—if ever—there when events unfolded around the people now under investigation. As a result, their actions and assessments may appear not only controversial, but truly befuddling when seen from our point of view. In order to understand why people could have done what they did, we need to go back and triangulate and interpolate, from a wide variety of sources, the kinds of mindsets that they had at the time. But working against us are the inherent biases introduced by hindsight (Fischoff, 1975) and the multiple pressures and constraints that operate on almost every investigation—political as well as practical (Galison, 2000).

In this paper I hope to make a contribution to our ability to reconstruct past human performance and how it played a role in accidents. I first capture some of the mechanisms of the hindsight bias, and observe them at work in how we routinely handle and describe human performance evidence. Trying to avoid these biases and mechanisms, I propose ways forward for how to reconstruct people's unfolding mindsets. Most examples will come from aviation, but they, and the principles they illustrate, should apply equally well to domains ranging from driving to shipping to industrial and occupational safety.

The mechanisms of hindsight

One of the safest bets we can make as investigators or outside observers is that we know more about the incident or accident than the people who were caught up in it—thanks to hindsight:

- Hindsight means being able to look back, from the outside, on a sequence of events that led to an outcome we already know about;
- Hindsight gives us almost unlimited access to the true nature of the situation that surrounded people at the time (where they actually were versus where they thought they were; what state their system was in versus what they thought it was in);
- Hindsight allows us to pinpoint what people missed and shouldn't have missed; what they didn't do but should have done.

Figure 1 shows the position of retrospective outsider, and contrasts it with the perspective from the people inside the sequence of events.

Insert figure 1 about here

From the perspective of the outside and hindsight (typically the investigator's perspective), we can oversee the entire sequence of events—the triggering conditions, its various twists and turns, the outcome, and the true nature of circumstances surrounding the route to trouble. In contrast, the perspective from the inside of the tunnel is the point of view of people in the unfolding situation. To them, the outcome was not known, nor the entirety of surrounding circumstances. They contributed to the direction of the sequence of events on the basis of what they saw on the *inside* of the unfolding situation. For investigators, however, it is very difficult to attain this perspective. The mechanisms by which hindsight operates on human performance data are mutually reinforcing. Together they continually pull us in the direction of the position of the retrospective outsider. The ways in which we retrieve human performance evidence from the rubble of an accident, represent it, and re-tell it, typically sponsors this migration of viewpoint.

Mechanism 1: Making tangled histories linear by cherry-picking and re-grouping evidence

One effect of hindsight is that ”people who know the outcome of a complex prior history of tangled, indeterminate events, remember that history as being much more determinant, leading ’inevitably’ to the outcome they already knew” (Weick, 1995, p28). Hindsight allows us to change past indeterminacy and complexity into order, structure, and oversimplified causality (Reason, 1990). In trying to make sense of past performance, it is always tempting to group individual fragments of human performance which *prima facie* point to some common condition or mindset. For example, "hurry" to land is a *leitmotif* extracted from the evidence in the following investigation, and that haste in turn is enlisted to explain the errors that were made:

> ”Investigators were able to identify a series of errors that initiated with the flightcrew’s acceptance of the controller’s offer to land on runway 19…The CVR indicates that the decision to accept the offer to land on runway 19 was made jointly by the captain and the first officer in a 4-second exchange that began at 2136:38. The captain asked: ’would you like to shoot the one nine straight in?’ The first officer responded, ’Yeah, we’ll have to scramble to get down. We can do it.’ This interchange followed an earlier discussion in which the captain indicated to the first officer his desire to hurry the arrival into Cali, following the delay on departure from Miami, in an apparent to minimize the effect of the delay on the flight attendants' rest requirements. For example, at 2126:01, he asked the first officer to ’keep the speed up in the descent’… (This is) evidence of the hurried nature of the tasks performed.” (Aeronautica Civil, 1996, p. 29)

But the fragments used to build the argument of haste come from over half an hour of extended performance. The investigator treats the record as if it were a public quarry to pick stones from, and the accident explanation the building he needs to erect. The problem is that each fragment is meaningless outside the context that produced it: each fragment has its own story, background, and reasons for being, and when it was produced it may have had nothing to do with the other fragments it is now grouped with. Also, behavior takes place in between the fragments. These intermediary episodes contain changes and evolutions in perceptions and assessments that separate the excised fragments not only in time, but also in meaning. Thus, the condition, and the constructed linearity in the story that binds these performance fragments, arises not from the circumstances that brought each of the fragments forth; it is not a feature of those circumstances. It is an artifact of the investigator. In the case described above, ”hurry” is a condition identified in hindsight, one that plausibly couples the start of the flight (almost 2 hours behind schedule) with its fatal ending (on a mountainside rather than an airport). ”Hurry” is a retrospectively invoked *leitmotif* that guides the search for evidence about itself. It leaves the investigator with a story that is admittedly more linear and plausible and less messy and complex than the actual events. Yet it is not a set of findings, but of tautologies.

Mechanism 2: Finding what people could have done to avoid the accident

Tracing the sequence of events back from the outcome—that we as investigators already know about—we invariably come across joints where people had opportunities to revise their assessment of the situation but failed to do so; where people were given the option to recover from their route to trouble, but did not take it. These are counterfactuals—quite common in accident analysis. For example, "The airplane could have overcome the windshear encounter if the pitch attitude of 15 degrees nose-up had been maintained, the thrust had been set to 1.93 EPR (Engine Pressure Ratio) and the landing gear had been retracted on schedule" (NTSB, 1995, p. 119). Counterfactuals prove what could have happened if certain minute and often utopian conditions had been met. Counterfactual reasoning may be a fruitful exercise when trying to uncover potential countermeasures against such failures in the future.

But saying what people could have done in order to prevent a particular outcome does not explain why they did what they did. This is the problem with counterfactuals. When they are enlisted as explanatory proxy, they help circumvent the hard problem of investigations: finding out why people did what they did. Stressing what was not done (but if it had been done, the accident would not have happened) explains nothing about what actually happened, or why.

In addition, counterfactuals are a powerful tributary to the hindsight bias. They help us impose structure and linearity on tangled prior histories. Counterfactuals can convert a mass of indeterminate actions and events, themselves overlapping and interacting, into a linear series of straightforward bifurcations. For example, people could have perfectly executed the go-around maneuver but did not; they could have denied the runway change but did not. As the sequence of events rolls back into time, away from its outcome, the story builds. We notice that people chose the wrong prong at each fork, time and again—ferrying them along inevitably to the outcome that formed the starting point of our investigation (for without it, there would have been no investigation).

But human work in complex, dynamic worlds is seldom about simple dichotomous choices (as in: to err or not to err). Bifurcations are extremely rare—especially those that yield clear previews of the respective outcomes at each end. In reality, choice moments (such as there are) typically reveal multiple possible pathways that stretch out, like cracks in a window, into the ever denser fog of futures not yet known. Their outcomes are indeterminate; hidden in what is still to come. In reality, actions need to be taken under uncertainty and under the pressure of limited time and other resources. What from the retrospective outside may look like a discrete, leisurely two-choice opportunity to not fail, is from the inside really just one fragment caught up in a stream of surrounding actions and assessments. In fact, from the inside it may not look like a choice at all. These are often choices only in hindsight. To the people caught up in the sequence of events there was perhaps not any compelling reason to re-assess their situation or decide against anything (or else they probably would have) at the point the investigator has now found significant or controversial. They were likely doing what they were doing because they thought they were right; given their understanding of the situation; their pressures. The challenge for an investigator becomes to understand how this may not have been a discrete event to the people whose actions are under investigation. The investigator needs to see how other people's "decisions" to continue were likely nothing more than continuous behavior—reinforced by their current understanding of the situation, confirmed by the cues they were focusing on, and reaffirmed by their expectations of how things would develop.

Mechanism 3: Judging people for what they did not do but should have done

Where counterfactuals are used in investigations, even as explanatory proxy, they themselves often require explanations as well. After all, if an exit from the route to trouble stands out so clearly to us, how was it possible for other people to miss it? If there was an opportunity to recover, to not crash, then failing to grab it demands an explanation. The place where investigators look for clarification is often the set of rules, professional standards and available data that surrounded people's operation at the time, and how people did not see or meet that which they should have seen or met. Recognizing that there is a mismatch between what *was* done or seen and what *should* have been done or seen—as per those standards—we easily judge people for not doing what they should have done.

Where fragments of behavior are contrasted with written guidance that can be found to have been applicable in hindsight, actual performance is often found wanting; it does not live up to procedures or regulations. For example, ”One of the pilots…executed (a computer entry) without having verified that it was the correct selection and without having first obtained approval of the other pilot, contrary to procedures.” (Aeronautica Civil, 1996; p. 31). Investigations invest considerably in organizational archeology so that they can construct the regulatory or procedural framework within which the operations took place, or should have taken place. Inconsistencies between existing procedures or regulations and actual behavior are easy to expose when organizational records are excavated after-the-fact and rules uncovered that would have fit this or that particular situation. This is not, however, very informative. There is virtually always a mismatch between actual behavior and written guidance that can be located in hindsight (Suchman, 1987; Woods *et al.,* 1994). Pointing that there is a mismatch sheds little light on the *why* of the behavior in question. And for that matter, mismatches between procedures and practice are not unique to mishaps (Degani & Wiener, 1991).

Another route to constructing a world against which investigators hold individual performance fragments, is finding all the cues in a situation that were not picked up by the practitioners, but that, in hindsight, proved critical. Take the turn towards the mountains on the left that was made just before an accident near Cali, Colombia in 1995 (Aeronautica Civil, 1996). What should the crew have seen in order to notice the turn? They had plenty of indications, according to the manufacturer of their aircraft:

> ”Indications that the airplane was in a left turn would have included the following: the EHSI (Electronic Horizontal Situation Indicator) Map Display (if selected) with a curved path leading away from the intended direction of flight; the EHSI VOR display, with the CDI (Course Deviation Indicator) displaced to the right, indicating the airplane was left of the direct Cali VOR course, the EaDI indicating approximately 16 degrees of bank, and all heading indicators moving to the right. Additionally the crew may have tuned Rozo in the ADF and may have had bearing pointer information to Rozo NDB on the RMDI” (Boeing, 1996, p. 13).

This is a standard response after mishaps: point to the data that would have revealed the true nature of the situation. Knowledge of the ”critical” data comes only with the omniscience of hindsight, but if data can be shown to have been physically available, it is assumed that it should have been picked up by the practitioners in the situation. The problem is that pointing out that it should have does not explain why it was not, or why it was interpreted differently back then (Weick, 1995). There is a dissociation between data availability and data observability (Woods *et al.*, 1994)—between what can be shown to have been physically available and what would have been observable given the multiple interleaving tasks, goals, attentional focus, interests, and—as Vaughan (1996) shows—culture of the practitioner.

There are also less obvious or not documented standards. These are often invoked when a controversial fragment (e.g. a decision to accept a runway change (Aeronautica Civil, 1996), or the decision to go around or not (NTSB; 1995)) knows no clear pre-ordained guidance but relies on local, situated judgment. For these cases there are always ”standards of good practice” which are based on convention and putatively practiced across an entire industry. One such standard in aviation is ”good airmanship”, which, if nothing else can, will explain the variance in behavior that had not yet been accounted for.

While micromatching, the investigator frames people's past assessments and actions inside a world that s/he has invoked retrospectively. Looking at the frame as overlay on the sequence of events, s/he sees that pieces of behavior stick out in various places and at various angles: a rule not followed here; available data not observed there; professional standards not met overthere. But rather than explaining controversial fragments in relation to the circumstances that brought them forth, and in relation to the stream of preceding as well as succeeding behaviors which surrounded them, the frame merely boxes performance fragments inside a world the investigator now knows to be true. The problem is this after-the-fact-world may have very little relevance to the actual world that produced the behavior under investigation. The behavior is contrasted against the investigator’s reality, not the reality surrounding the behavior in question at the time. Judging people for what they did not do relative to some rule or standard does not explain why they did what they did. Saying that people failed to take this or that pathway—only in hindsight the right one—judges other people from a position of broader insight and outcome knowledge that they themselves did not have. It does not explain a thing yet; it does not shed any light on why people did what they did given *their* surrounding circumstances. The investigator has gotten caught in what William James called "the psychologist's fallacy" a century ago: he has substituted his own reality for the one of his object of study.

It appears that in order to explain failure, we seek failure. In order to explain missed opportunities and bad choices, we seek flawed analyses, inaccurate perceptions, violated rules—even if these were not thought to be influential or obvious or even flawed at the time (Starbuck & Milliken, 1988). This search for people's failures is another well-documented effect of the hindsight bias: knowledge of outcome fundamentally influences how we see a process. If we know the outcome was bad, we can no longer objectively look at the behavior leading up to it—it must also have been bad (Fischoff, 1975; Woods *et al.*, 1994; Reason, 1997).

Local rationality

What is striking about many accidents in complex systems is that people were doing exactly the sorts of things they would usually be doing—the things that usually lead to success and safety. Mishaps are more typically the result of everyday influences on everyday decision making than that they are isolated cases of erratic individuals behaving unrepresentatively (e.g. Woods et al., 1994; Reason, 1997; AMA, 1998; Sanne, 1999). People are doing what makes sense given the situational indications, operational pressures and organizational norms existing at the time. Accidents are seldom preceded by bizarre behavior. People's errors and mistakes (such as there are in any objective sense) are systematically coupled to their circumstances and tools and tasks. Indeed, a most important empirical regularity of human factors research since the mid-forties is the local rationality principle. What people do makes sense to them at the time—it has to, otherwise they would not do it. People do not come to work to do a bad job; they are not out on crashing cars or airplanes or grounding ships. The local rationality principle, originating in Simon (1969), says that people do things that are reasonable, or rational, based on their limited knowledge, goals, and understanding of the situation and their limited resources at the time (Woods *et al.*, 1994). Avoiding the mechanisms of the hindsight bias means acknowledging that failures are baked into the nature of people's work and organization; that they are symptoms of deeper trouble or by-products of systemic brittleness in the way business is done. It means having to find out why what people did back there and then actually made sense given the organization and operation that surrounded them.

To explain outcome failure, we need to convert the search for human failures into a search for human sensemaking (Snook, 2000). The question is not "where did people go wrong?", but "why did this assessment or action make sense to them at the time?". Such real insight is derived not from judging people from the position of retrospective outsider, but from seeing the world through the eyes of the protagonists at the time. When looking at the sequence of events from this perspective, a very different story often struggles into view.

The reconstruction of unfolding mindset

How do we capture the perspective from inside the tunnel, so that we can generate meaningful results from our probe? The investigator is confronted by a problem similar to that of the field researcher—how to migrate from a context-specific set of data to more concept-based results that are interpretable and falsifiable; that are more than just another anecdote (Woods, 1993; Xiao & Vincente, 2000). Falsifiability means the investigator has to leave a trace that others can follow. In human factors it is not uncommon to make the shift from context-specific to concept-dependent in one big leap (e.g. "this underestimate of the closing rate signifies a loss of situation awareness"); which produces conclusions that no one else can verify. The challenge is to build up an account that moves from the context-specific to the concept-dependent gradually, leaving a clear trace for others to follow, verify, and debate (e.g. Hollnagel *et al.*, 1981). To be sure, any explanation of past performance that we arrive at remains a fictional story; an approximation; a tentative match—open to revision as new evidence may come in. In the words of Woods (1993, p. 238): "A critical factor is identifying and resolving all anomalies in a potential interpretation. We have more confidence in, or are more willing to pretend that, the story may in fact have some relation to reality if all currently known data about the sequence of events and background are coherently accounted for by the reconstruction". Below I present five steps by which the investigator could begin to reconstruct a concept-dependent account from context-specific incident data.

1. Laying out the sequence of events in context-specific language

The record and other data about an incident typically reveals a sequence of activities—human observations, actions, assessments, decisions; as well as changes in the state of the process or system. This sequence of events forms the starting point for an examination of the inside of the tunnel. The goal is to examine how people's mindset unfolded in parallel with the situation evolving around them, and how people, in turn, helped influence the course of events. There is a fundamental reciprocity in human information processing (Neisser, 1976; Clark, 1997) from which the investigator can benefit by triangulation and interpolation. Cues and indications from the world influence people's situation assessments, which in turn inform their actions, which, in turn, change the world and what it reveals about itself, and so forth. This means that if certain actions or assessments are difficult to interpret, then the circumstances (and particularly what was observable about them) in which they appeared can hold the key to their sensibility. Indeed, the reconstruction of mindset often begins not with the mind, but with the situation in which the mind found itself. Similarly, if there is a lack of data from system or process sources, certain behaviors that are canonical in particular process states can help you reconstruct the state of cues and indications observable at the time. This makes that there are various entries to scour the record for events and activities:

- <u>Shifts in behavior.</u> There can be points where people may have realized that the situation was different from what they believed it to be previously. You can see this either in their remarks or their actions. These shifts are markers where later you want to look for the indications unfolding around them that people may have used to come to a different realization.
- <u>Actions to influence the process</u> may come from people's own intentions. Depending on the kind of data that the domain records or provides, evidence for these actions may not be found in the actions themselves, but in process changes that follow from them. As a clue for a later step, such actions also form a nice little window on people's understanding of the situation at that time.
- <u>Changes in the process</u>. Any significant change in the process that people manage must serve as event. Not all changes in a process managed by people actually come from people. In fact, increasing automation in a variety of workplaces has led to the potential for autonomous process changes almost everywhere—for example:
    - Automatic shut-down sequences or other interventions;
    - Alarms that go off because a parameter crossed a threshold;
    - Uncommanded mode changes;
    - Autonomous recovery from undesirable states or configurations.

  Yet even if they are autonomous, these process changes do not happen in a vacuum. They always point to human behavior around them; behavior that preceded it and behavior that followed it. People may have helped to get the process into a configuration where autonomous changes were triggered. And when changes happen, people notice them or not; people respond to them or not. Such actions, or the lack of them, again give you a strong clue about people's knowledge and current understanding.

The way to capture these events and activities during this stage is in context-specific language—meaning a minimum of psychological diction; instead a version of what happened in terms that domain people use to talk about their own work. The goal is to miss as few details as possible. Skipping to higher-level descriptions of human performance is seductive, even at this stage, but should be avoided. Seemingly low-level concepts, such as "decision making" or "diagnosis", already are large—meaning they contain a lot of behavior—and are easily mistaken for detailed insight into psychological issues (Woods, 1993; Hollnagel, 1998).

Time (and/or space) can be powerful organizing principles to help lay out the activities and events. Behavior, and the process in which it took place, unfolded over time and, probably, in some space. By organizing data spatially and temporally (e.g. through drawing maps or timelines or both), actions and assessments can become more clearly coupled to the process state and location in which they took place; they can recover their spot in the flow of events of which they were part and which helped bring them forth. Such organization likely yields further clues about why actions and assessments made sense to people back there and then.

2. Divide the sequence of events into episodes, if necessary

Accidents do not just happen; they evolve over a period of time. Sometimes this time may be long (e.g. 34 hours, see NTSB, 1996), and where it is, it may be fruitful to divide the sequence of events into separate episodes that each deserve their own further human performance analysis. Cues about where to chunk up the sequence of events can mostly come from the domain description arrived at above, especially at discontinuities in human assessments or actions or process states.

There is of course inherent difficulty in deciding what counts as the overall beginning of a sequence of events (especially the beginning—the end often speaks for itself). Since, philosophically, there is no such thing as a root cause, there is technically no such thing as the beginning of a mishap. Yet the investigation needs to start somewhere. Making clear where it starts and explaining this choice is a good step toward a structured, well-engineered human performance investigation. Here is one option: Take as the beginning of your first episode the first assessment, decision or action by people or the system close to the mishap—the one that, according to you, set the sequence of events in motion. This assessment or action can be seen as a trigger for the events that unfold from there. Of course the trigger itself has a reason, a background, that extends back beyond the mishap sequence—both in time and in place. The whole point of taking a proximal action or event as starting point is not to ignore this background, but to identify concrete points to begin the investigation into them.

3. Find out how the world looked or changed during each episode

This step is about reconstructing the unfolding world that people inhabited: find out what their process was doing; what data was available. This is the first step toward coupling behavior and situation—toward putting the observed behavior back into the situation that produced and accompanied it. Laying out how some of the critical parameters changed over time is nothing new to investigations. Many accident report appendices contain read-outs from data recorders, which show the graphs of known and relevant process parameters. But building these pictures is often where investigations stop today. Tentative references about connections between known parameters and people's assessments and actions are sometimes made, but never in a systematic, or graphic way. The point here to marry all the events that have been identified with the unfolding process—to begin to see the two in parallel, as an inextricable, causal *dance-a-deux*. The point of step three is to build a picture that shows these connections.

The record will most likely contain (some kind of) data about how process parameters were changing over time (speed until impact, for example, but also traces of changing pressures, ratios, settings, quantities, automation or computer modes, rates, and so forth) and how these were presented to the people in question. Considerable domain knowledge (either from the investigator him/herself or from outside) may be necessary to determine which of the parameters could have counted as a stimulus for the behavior under investigation. The difficulty (reflected in the next step) will be to move from merely showing that certain data was physically available, to arguing which of these data was actually observable and made a difference in people's assessments and actions—and why this made sense to them back then.

4. Identify people's goals, focus of attention and knowledge active at the time

So what, out of all the data available, did people actually see and how did they interpret it? Given that human behavior is goal-directed and governed by knowledge activated *in situ* (Woods *et al.*, 1994), clues are available from looking at people's goals at the time, and at the knowledge activated to help pursue them.

Finding what goals people were working on does not need to be difficult. It often connects directly to how the process was unfolding around them:

- What was canonical, or normal at this time in the operation? Tasks (and the goals they represent) relate in systematic ways to stages in a process.
- What was happening in the process managed by the people? Systems were set or inputs were made—changes which connect to the tasks people were carrying out.
- What were other people in the operating environment doing? People who work together on common goals often divide the necessary tasks among them in predictable or complementary ways. There may be standard role divisions, for example between pilot flying and pilot not-flying, that specify the work for each.

It is seldom the case, however, that just one goal governs what people do. Most complex work is characterized by multiple goals, all of which are active or must be pursued at the same time (on-time performance and safety, for example). Depending on the circumstances, some of these goals may be at odds with one another, producing goal conflicts. Any analysis of human performance has to take the potential for goal conflicts into account. Goal trade-offs can be generated by the nature of the work itself. For example, anesthesiologists need to maximize pre-operative workup time with a patient to guard patient safety and quell liability concerns, while their schedules interlock with other professions that exercise pressure with respect to e.g. timing. Goal conflicts can also precipitate from the organizational level. In this case, not all goals (or their respective priorities) are written down in guidance or procedures or job descriptions. In fact, most are probably not. This makes it difficult to trace or prove their contribution to particular assessments or actions. However, previous occurrences in similar circumstances or in the same organization may yield powerful clues. They can substantially influence people's criterion setting with respect to a goal conflict. For example, a decision to take off or not to take off in bad weather may be precluded by earlier incidents, or, conversely, encouraged by organizational reactions to lack of on-time performance.

When it comes to knowledge, not all knowledge people once showed to possess is necessarily available when called for. In fact, the problem of knowledge organization (is it structured so that it can be applied effectively in operational circumstances?) and inert knowledge (even if it is there, does it get activated in context?) should attune investigators to mismatches between how knowledge was acquired and how it is to be applied in practice. For example, if material is learned in neat chunks and static ways (books, most computer-based training) but needs to be applied in dynamic situations that call for novel and complex combinations, then inert knowledge is a risk (Woods *et al.*, 1994).

What people know and what they try to accomplish jointly determines where they will look; where they will direct their attention—and consequently, which data will be observable to them. Recognize how this is, once again, the local rationality principle. People are not unlimited cognitive processors (there are no unlimited cognitive processors in the entire universe). People do not know and see everything all the time. So their rationality is limited, or bounded. What people do, where they focus, and how they interpret cues makes sense from their point of view; their knowledge, their objectives and their limited resources (e.g. time, processing capacity, workload). Re-establishing people's local rationality will help you understand the gap between data availability and what people actually saw or used. In dynamic situations, people direct their attention as a joint result of:

- What their current understanding of the situation is, which in turn is determined partly by their knowledge and goals. Current understanding helps people form expectations about what should happen next (either as a result of their own actions or as a result of changes in the world itself).
- What happens in the world. Particularly salient or intrusive cues will draw attention even if they fall outside people's current interpretation of what is going on.

Keeping up with a dynamic world, in which situations evolve and change, is a demanding part of much operational work, and implies two different kinds of "errors". People may fall behind during rapidly changing conditions, and update their interpretation of what is happening constantly, trying to follow every little change in the world. Or people may become locked in one interpretation, even while evidence around them suggests that the situation has changed (see De Keyser & Woods, 1990).

5. Step up to a conceptual description

The goal here is to build an account of human performance that runs parallel to the one created in step 1. This time, however, the language that describes the same sequence of events is not one of domain terms, it is one of human factors or psychological concepts. One reason for the importance of this step perhaps goes beyond the mandate of an individual investigation. Getting away from the context-specific details—in a language that may not communicate well with other context-specific sequences of events—opens a crucial way to learn from failure: discovering similarities between seemingly disparate events. When people instead stress the differences between sequences of events, learning anything of value beyond the one event becomes difficult (Rochlin, 1999). Similarities between accounts of different occurrences can point you to common conditions that helped produce the problem under investigation. Figure 2 shows the steps involved in the reconstruction of unfolding mindset.

Insert figure 2 about here

Conclusion

The systematic investigation of human contributions to accidents is not yet a very well-established practice with common methods or assumptions. Investigators are often forced to rely almost exclusively on domain knowledge and common sense, but this exposes them to the mechanisms of hindsight. Human performance evidence can get disembodied from the flow of events that accompanied it and brought it forth; and conclusions about the human contribution easily become counterfactual and judgmental—stressing what people should have done to avoid the accident, but failed to do. None of this explains what really happened or why. There may be a need for stronger appreciation among investigators of the methodical challenges and pitfalls associated with retrospective analyses of human performance. Even clearer is the need for further development of ways in which investigators can systematically reconstruct the human contribution to accidents and avoid the biases of hindsight.

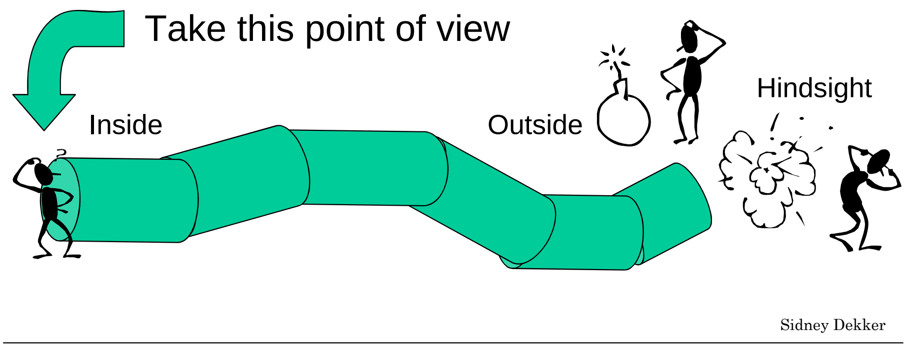

Fig. 1: See the unfolding world from the point of view of people inside the situation—not from the outside or from hindsight.

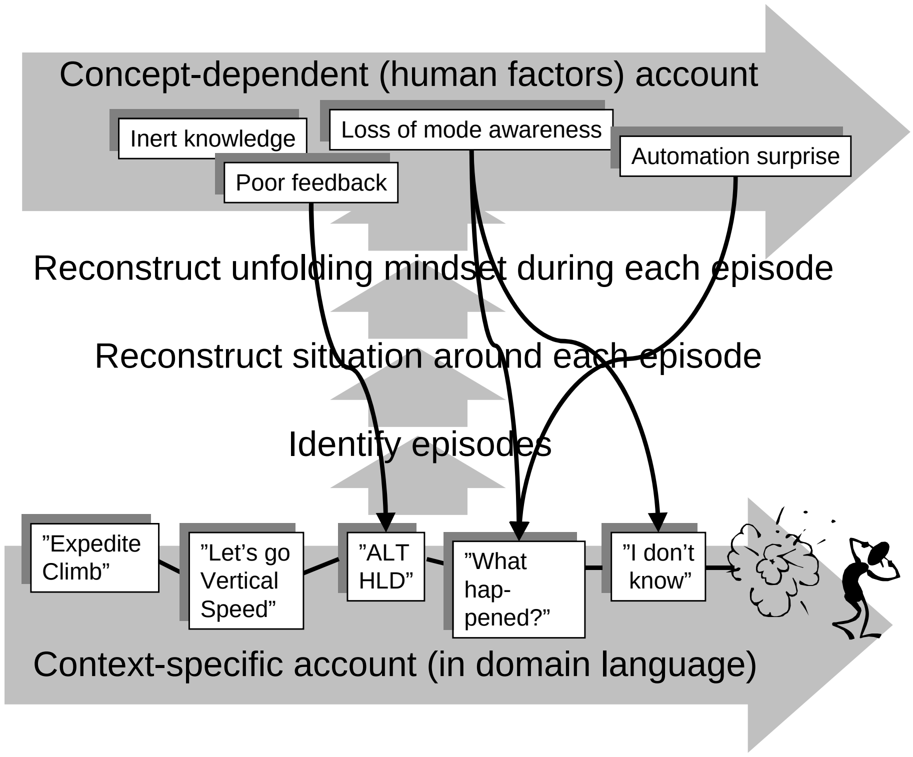

Fig. 2: Close the gap from data to interpretation by following and documenting the various steps between a context-specific account of what happened and a concept-dependent one, linking back the concepts found to specific evidence in the context-specific record.

References

Aeronautica Civil (1996). *Aircraft accident report: Controlled flight into terrain, American Airlines flight 965, Boeing 757-223, N651AA near Cali, Colombia, December 20, 1995*. Bogota, Colombia: Aeronautica Civil.

American Medical Association (1998). *A tale of two stories: Contrasting views of patient safety*. Report from a workshop on assembling the scientific basis for progress on patient safety. Chicago, IL: National Patient Safety Foundation at the AMA.

Boeing Commercial Airplane Group (1996). *Boeing submission to the American Arilines Flight 965 Accident Investigation Board*. Seattle, WA: Boeing.

Clark, A (1997). *Being there: Putting brain, body and world together again*. Cambridge, MA: MIT Press.

Degani, A., & Wiener, E. L. (1991). Philosophy, policies and procedures: The three P's of flightdeck operations. *Paper presented at the Sixth International Symposium on Aviation Psychology, Columbus, OH*, April.

De Keyser, V., & Woods, D. D. (1990). Fixation errors: Failures to revise situation assessment in dynamic and risky systems. In A. G. Colombo & A. Saiz de Bustamante (Eds.). *System reliability assessment*, pp. 231-251. Dordrecht, NL: Kluwer Academic.

Fischoff, B. (1975). Hindsight is not foresight: The effect of outcome knowledge on judgement under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 1(3),* 288-299.

Galison, P. (2000). An accident of history. In P. Galison, & A. Roland (eds.): *Atmospheric flight in the twentieth century*. Dordrecht, NL: Kluwer Academic Publishers.

Hoffman, R. R., & Woods, D. D. (2000). Studying cognitive systems in context. *Human factors, 42*(1), 1-7.

Hollnagel, E., Pederson, O. M., & Rasmussen, J. (1981). *Notes on human performance analysis* (Tech. Rep. Risoe-M-2285). Denmark: Risoe National Laboratory.

Hollnagel, E. (1998). Measurements and models, models and measurements: You can't have one without the other. In *Proceedings of NATO AGARD conference*, Edinburgh, UK.

National Transportation Safety Board (1995). *Aircraft accident report: Flight into terrain during missed approach, USAir flight 1016, DC-9-31, N954VJ, Charlotte Douglas International Airport, Charlotte, North Carolina, July 2, 1994* (NTSB/AAR-95/03). Washington, DC: NTSB.

Neisser, U. (1976). *Cognition and reality*. San Francisco, CA: Freeman.

Rasmussen, J., & Batstone, R. (1989). *Why do complex organizational systems fail?* Environment Working Paper No. 20. Washington, DC: World Bank.

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

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

Reason, J. T. (2000). Grace under fire: Compensating for adverse events in cardiothoracic surgery. *Paper presented at the 5th conference on naturalistic decision making*, Tammsvik, Sweden. May, 2000.

Rochlin, G. I. (1999). Safe operation as a social construct. *Ergonomics*, 42, 1549-1560.

Sanne, J. M. (1999). *Creating safety in air traffic control*. Lund, Sweden: Arkiv.

Shappell, S. A., & Wiegmann, D. A. (2001). Applying Reason: the human factors analysis and classification system (HFACS). *Human Factors and Aerospace Safety, 1*(1), 59-86.

Simon, H. (1969). *The sciences of the artificial*. Cambridge, MA: MIT Press.

Snook, S. (2000). *Friendly fire*. Princeton, NJ: Princeton University Press.

Starbuck, W. H. & Milliken, F. J. (1988). Challenger: Fine-tuning the odds until something breaks. *Journal of managemernt studies, 25*, 319-340.

Suchman, L. A. (1987). *Plans and situated actions: The problem of human-machine communication*. Cambridge, UK: Cambridge University Press.
Vaughan, D. (1996). *The Challenger launch decision*. Chicago, IL: University of Chicago Press.
Weick, K. (1995). *Sensemaking in organizations*. London: Sage.
Woods, D. D. (1993). Process-tracing methods for the study of cognition outside of the experimental laboratory. In G.A. Klein, J. Orasanu, R. Calderwood & C. E. Zsambok (Eds.), *Decision making in action: Models and methods*, pp. 228-251. Norwood, NJ: Ablex.
Woods, D. D., Johannesen, L. J., Cook, R. I., & Sarter, N. B. (1994). *Behind human error: Cognitive systems, computers and hindsight.* Dayton, OH: CSERIAC.
Xiao, Y., & Vincente, K. J: (2000). A framework for epistemological analysis in empirical (laboratory and field) studies. *Human factors, 42*(1), 87-101.

Sidney Dekker (Ph.D., 1996, The Ohio State University) is normally Associate Professor in the Division of Industrial Ergonomics, Linköping Institute of Technology, Sweden, and currently Senior Visiting Fellow At Nanyang Technological University in Singapore. A pilot and flight instructor, he was previously employed in the UK at the British Aerospace Dependable Computing Systems Centre. He is full member of the Human Factors and Ergonomics Society and the European Assocation for Cognitive Ergonomics. His latest book is *The Field Guide to Human Error Investigations* (Cranfield University Press/Ashgate, 2002).