SCHOOL OF AVIATION
Lund University

The re-invention of human error

Sidney W. A. Dekker

Technical Report 2002-01

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

Abstract

This paper contrasts the new view of human error with what is known as the old view, and examines the implications for progress on aerospace safety. In the old view, systems are basically safe, and it is human error that causes most accidents. The new view sees systems as not basically safe, and human errors are symptoms of contradictions, pressures and resource limitations deeper inside the system. By looking critically at two common methods for classifying human errors in aerospace, the paper traces ways in which attempts to introduce the new view become re-inventions of the old view of human error. This happens when we assume that (1) we can meaningfully count errors and assign them to categories, that classification is the same as analysis; (2) human error is caused by inherent processing limitations or motivational shortcomings; (3) judgments, for example of "people losing situation awareness" under fashionable guises can double as explanations of human error; and (4) we should seek sources of failure higher up, away from the local sharp end operators, by locating blame elsewhere in an organization. These error classification systems re-tread the old view around the wheels of supposed progress on safety, reinforcing the biases we already suffer when dealing with failure. In reality, progress on safety can be made by understanding how people create safety, and by understanding how the creation of safety can break down in resource-limited systems that pursue multiple competing goals. The paper argues for abandoning the construction of cause and instead deepening our insights into patterns of failure; the mechanisms by which failure succeeds.

Keywords: human error, mechanisms of failure, construction of cause, human factors, classification, creation of safety

Running head: The re-invention of human error

Introduction

There are basically two ways of looking at human error. We can see human error as a cause of failure, or we can see human error as a symptom of failure (Woods *et al.*, 1994). These two views have recently been contrasted as *the old view* of human error versus *the new view* (Cook, Render & Woods, 2000; AMA, 1998; Reason, 2000) and painted as fundamentally irreconcilable perspectives on the human contribution to system success and failure. In the old view of human error:

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

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.

It is not difficult to find proponents of the new view—in principle—in aerospace human factors. For example:

"...simply writing off aviation accidents merely to pilot 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. In fact, even the identification of a 'primary' cause is fraught with problems. Instead, aviation accidents are the result of a number of causes..." (Shappell & Wiegmann, 2001, p. 60).

In practice, however, attempts to pursue the causes of system failure according to the new view can become retreads of the old view of human error. In practice, getting away from the tendency to judge instead of explain turns out to be difficult; avoiding the fundamental attribution error remains very hard; we tend to blame the man-in-the-loop. This is not because we aim to blame—in fact, we probably intend the opposite. But roads that lead to the old view in aerospace human factors are paved with intentions to follow the new view. Error classification methods, for example, that attempt to efficiently lead an investigator to the deeper sources of human factors trouble (e.g. Helmreich *et al.*, 1999; Shappell & Wiegmann, 2001) can quickly become latter-day versions of what I call The Bad Apple Theory. This identifies bad apples (unreliable human components) somewhere in an organization, and gets rid of them or somehow constrains their activities. In contradistinction, the new view of human error knows no quick or efficient roads to insight into the human contribution to failure (see e.g. Dekker, 2001). The few true "new view" analyses of the human contribution to system failures in aerospace are all exceptionally large, and typically cover multiple years to probe a single event (e.g. Moshansky, 1992; Vaughan, 1996; Snook, 2000). Their size and length is testimony to the complexity of the many processes and interactions that underlie the routes to failure; testimony to the depths we need to go in order to really understand what happened.

In this paper, I try to find the reasons why aerospace human factors easily slips into the old view of human error, when it had no intention of doing so. By going through two methods used to analyze human error in aerospace, I look for clues and assumptions that signify their regression into the old view. I call what these approaches do "the re-invention of human error", because that is their major achievement. They substitute new, fashionable labels for the old "human error" label, mistaking this for deeper insight; they confuse judgment with explanation; they pretend to look further up the causal pathway to find distal contributions, yet in doing so they shift blame ahead, looking for a place to park it. From this analysis I move on to the re-invention of human error once again, this time by understanding that people create safety at all levels in organizations; that safety, as well as failure, is an emergent property of systems trying to succeed. The creation of safety can break down in resource-constrained systems that pursue multiple competing goals in an uncertain, dynamic world. By looking at ways in which people invest in their awareness of pathways to breakdown and devise strategies to forestall failure, I will argue for abandoning the construction of cause altogether, since it severely oversimplifies our ideas of success and failure. Instead, I suggest that aerospace should invest in its understanding of patterns of failure; common mechanisms by which failure succeeds.

Fitts & Jones '47

The groundwork for the new view to human error was laid at the beginning of aerospace human factors. Fitts and Jones (1947) found how features of World War II airplane cockpits systematically influenced the way in which pilots made errors. For example, pilots confused the flap and gear handles because these typically looked and felt the same and were co-located. Or they mixed up the locations of throttle, mixture and propeller controls because these kept changing across different cockpits. Human error was the starting point for Fitts' and Jones' studies—not the conclusion. The label "pilot error" was deemed unsatisfactory, and used as a pointer to hunt for deeper, more systemic conditions that led to consistent trouble. The idea these studies convey to us is that mistakes actually make sense once we understand features of the engineered world that surrounds people. Human errors are systematically connected to features of people's tools and tasks. The insight, at the time as it is now, was profound: the world is not unchangeable; systems are not static, not simply given. We can re-tool, re-build, re-design, and thus influence the way in which people perform. This, indeed, is the historical imperative of human factors—understanding why people do what they do so we can tweak, change the world in which they work and shape their assessments and actions accordingly.

Years later, aerospace human factors extended the Fitts and Jones work—extended it, not changed it. Increasingly, we realized how trade-offs by people at the sharp end are influenced by what happens at the blunt end of their operating worlds; their organizations (Maurino *et al.*, 1995). Organizations make resources available for people to use in local workplaces (tools, training, teammates) but put constraints on what goes on there at the same time (time pressures, economic considerations), which in turn influences the way in which people decide and act in context (Woods *et al.*, 1994; Reason, 1997). Again, what people do makes sense on the basis of the circumstances surrounding them, but now circumstances that reach far beyond their immediate engineered interfaces. This realization has put the Fitts and Jones premise to work in organizational contexts, for example changing workplace conditions or reducing working hours or de-emphasizing production to encourage safer trade-offs on the line (e.g. the "no fault go-around policy" held by many airlines today, where no (nasty) questions will be asked if a pilot breaks off his attempt to

land). Human error is still systematically connected to features of people's tools and tasks, and, as acknowledged more recently, their operational and organizational environment.

The re-invention of human error I: Human error by any other name

In order to lead people (e.g. investigators) to the sources of human error as inspired by the organizational extension of Fitts and Jones '47, a number of methods have been developed in aerospace. I will look at two in more detail here, the Threat and Error Management Model (e.g. Helmreich, 2000) and the Human Factors Analysis and Classification System (HFACS, Shappell & Wiegmann, 2001). The biggest trap in both error methods is the illusion that classification is the same as analysis. While both intend to provide investigators more insight into the background of human error, they actually risk trotting down a garden path toward judgments of people instead of explanations of their performance; toward shifting blame higher and further into or even out of organizational echelons, but always onto others. How does the road, paved with new view intentions, end up in the old view of human error? The error tools or models mentioned above have much in common, and it is these commonalities that provide a clue about the mechanisms by which we, unwillingly, may regress into the old view of human error.

The myth of the error count

The starting point for both methods is the error count. "Between 70 and 80% of aviation accidents can be attributed, at least in part, to human error" (Shappell and Wiegmann, 2001, p. 60). "Research...into aviation accidents has found that 70% involve human error" (Helmreich, 2000, p. 781). These percentages, putative testimony to the size and nature of the problem the industry faces, may continue to muster political momentum and resources for human factors, but it has neither a basis in scientific practice, nor a future in productive countermeasures. When it comes to error, counting is particularly difficult. Human error "in the wild"—that is, as it occurs in natural complex settings—resists tabulation because of the complex interactions, the long and twisted pathways to breakdown and the context-dependency and diversity of human intention and action.

Consider the following case of a "human error"—failing to arm the ground spoilers before landing. Passenger aircraft have "spoilers"—panels that come up from the wing upon landing, to help brake the aircraft during its landing roll-out. To make these spoilers come out, pilots have to manually "arm" them by pulling a lever in the cockpit. Many aircraft have landed without the spoilers being armed, some cases even resulting in runway overruns. Each of these events gets classified as "human error"—after all, the human pilots forgot something in a system that is functioning perfectly otherwise. But deeper probing reveals a system that is not at all functioning perfectly. Spoilers typically have to be armed after the landing gear has come out and is safely locked into place. The reason is that landing gears have compression switches which communicate to the aircraft when it is on the ground (how else would an aircraft know that, right?). When the gear compresses, the logic tells the aircraft that it has landed. And then the spoilers come out (if they are armed, that is). Gear compression, however, can also occur *while* the gear is coming out, because of air pressure from the enormous slip stream around a flying aircraft, especially if landing gear folds open *into* the wind (which many do). This would create a case where the aircraft thinks it is on the ground, but it isn't, really. If the spoilers would already be armed at that time, they would come out too—which would be really bad if you are still airborne. To prevent this from happening, all these aircraft carry procedures that say that the spoilers

may only be armed when the gear is fully down and locked. It is safe to do so, because the gear is then orthogonal to the slipstream, with no more risk of compression. But the older an aircraft gets, the longer a gear takes to come out and lock into place. The hydraulic system no longer works as well, for example. In some aircraft, it can take up to half a minute. By that time, the gear extension has begun to seriously intrude into other cockpit tasks that need to happen by then—selecting wing flaps for landing, capturing and tracking the electronic glide slope towards the runway, and so forth. These are items that come *after* the "arm spoilers item on a typical before-landing checklist. If the gear is still doing its thing, while the world has already pushed you further down the checklist, not arming the spoilers is a slip that is only too easy to make. Combine this with a system that, in many aircraft, never warns pilots that their spoilers are not armed; a spoiler handle that sits over to one, dark side of the center cockpit console, obscured for one pilot by power levers, and whose difference between armed and not-armed may be all of one inch, and the question becomes: is this mechanical failure or human error?

"Human error", if there were such a thing, is not a question of an individual single-point failure; a failure to notice or process that can simply be counted and added up—not in the spoiler story and probably not in any story of breakdowns in flight safety. Practice that goes wrong spreads out over time and in space, touching all the areas that usually make people successful. It extends deeply into the engineered, organized, social and operational world in which people carry out their work. Were we to trace "the cause" of failure, the causal network would fan out immediately, like cracks in a window, with only we determining when to stop looking because the evidence will not do it for us. Labeling certain assessments or actions in the swirl of human and social and technical activity as causal, or as "errors" and counting them in some database, is entirely arbitrary and ultimately meaningless.

There are additional problems with the 70% myth. For example, what do we refer to when we say "error"? In safety debates there are three ways of using the label "error":

- Error as the *cause* of failure. For example: This event was due to human error.
- Error as the *failure itself*. For example: The pilot's selection of that mode was an error.
- Error as a *process*, or, more specifically, as a departure from some kind of standard. This may be operating procedures, or simply good airmanship. Depending on what you use as standard, you will come to different conclusions about what is an error.

The classification systems do not differentiate among these different possible definitions of error. This is actually an old and well-documented problem in human factors (e.g. Hollnagel, 1998) and specifically in error classifications (see e.g. Dougherty's 1990 critique of first-generation HRA (Human Reliability Analysis)): the inability to sort out what is cause and what is consequence. Or, to put it in other words, what is the cause of error and what is its manifestation; what is genotypical, what is phenotypical. As a case in point, Helmreich *et al.* (1999) asks its raters to divide their observations into the following five categories, and tabulate how many they find in each: intentional non-compliance, procedural errors, communication errors, proficiency problems and operational decisions. These are categories of manifestations of error (such as communication errors) as well as causes of error (such as proficiency problems), mixing causes and consequences; or confusing genotypes and phenotypes. Whether a manifested problem can be attributed to proficiency problems is not a classification; it is an inference, one that probably requires a considerable amount of domain knowledge to make. Counting and coarsely classifying surface variabilities is protoscientific at best. Counting does not make science, or even useful practice, since interventions on the basis of surface variability will merely peck away at the margins of an issue. A focus on superficial similarities blocks our ability to see deeper

relationships, deeper patterns, deeper reasons and subtleties. It disconnects performance fragments from the context that brought them forth, from the context that accompanied them; that gave them meaning; and that holds the keys to their explanation. Instead it renders performance fragments denuded: as uncloaked, context-less, meaningless shrapnel scattered across broad classifications in the wake of an observer's arbitrary judgment.

Exit the connection between errors and tools

While the original Fitts and Jones legacy lives on very strongly in human factors (for example in Norman (1994) who calls technology something that can make us either smart or dumb), neither Helmreich nor Shappell & Wiegmann pay much attention to the connection between error and people's tools. According to Helmreich (2000), "errors result from physiological and psychological limitations of humans. Causes of error include fatigue, workload, and fear, as well as cognitive overload, poor interpersonal communications, imperfect information processing, and flawed decision making" (p. 781). Gone are the systematic connections between people's assessments and actions on the one hand, and their tools and tasks on the other. In their place are purely human causes—sources of trouble that are endogenous; internal to the human component. Shappell and Wiegmann, following the original Reason (1990) division between latent failures and active failures, merely list an undifferentiated "poor design" only under potential organizational influences—the fourth level up in the causal stream that forms HFACS. Again, little effort is made to probe the systematic connections between human error and the engineered environment that people do their work in.

The inability of these methods to make a meaningful connection between errors and the tools people work with may reflect a broader oversimplification in aerospace that followed the widespread introduction of flightdeck automation. The decades following Fitts and Jones saw an enormous increase in cockpit complexity—and a slow drift away from the premise of their work. Increased complexity was due in large part to the growing ubiquity of computing power (Billings, 1996). From the early eighties onward, incidents and accidents with highly automated airliners showed that more complex, more powerful technology meant more complex pathways towards failure. Aircraft—by now well-defended against many known operator vulnerabilities—had to be managed into breakdown. Most mishaps with automated airliners shared a basic signature: a series of omissions and commissions; of miscommunications and misassessments was necessary to gradually push a system towards and over the edge of breakdown (FAA, 1996). A number of factors, both pertaining to flight deck design and to the situation, needed to be present to conspire against the crews' ability to coordinate their activities with the automation—high-tempo operations; autonomous actions on part of the machine; and limited feedback about its behavior. While defying attempts to categorize causes, these going sour accidents emphasized once again how the nature of the tools people work with is critical to both the success and failure of work in cockpits. Features of people's tools and tasks systematically shape their performance. In this case, coordination breakdowns between crew and automation were encouraged by the silent strength of, what really was, a bad teamplayer (Sarter & Woods, 1997).

Yet the immediate response—one that persists in many ways today—appears not to have embraced this insight. Fitts and Jones' 1947 findings were not re-inscribed in the language of automation, to once again explore the systematic connections between assessments and actions and features of people's operating environments. The straightforward link between people's actions and the environment in which they took place—that had once typified

human factors research and design countermeasures—became lost in the complex, twisted pathways to failure, and in the inability of an industry to face or grasp the profound design implications that automated airliner crashes in fact represented (FAA, 1996). Citing "clear misuse of automation in trouble-free aircraft" at the root of what went wrong, one manufacturer concluded that "while you can incorporate all the human engineering you want in an aircraft, it's not going to work if the human does not want to read what is presented to him, and verify that he hasn't made an error" (Woods *et al.*, 1994, p. 87). Problems with automation and system management, in other words, were attributed to human motivational or processing shortcomings—a premise sustained in the Helmreich and HFACS error classification methods. If only the human had tried a little harder (had wanted "to read what is presented to him"), if only she or he had paid a bit more attention to data or shared with other crewmembers that which we now know was critical, then these problems would not have occurred. Yet in reality, aerospace has seen the introduction of more technology as illusory antidote to the plague of human error. Instead of reducing human error, technology changed it, aggravated the consequences and delayed opportunities for error detection and recovery. Technology upset traditional human-to-human coordination, undermining traditional strategies for double-checking and monitoring each other's work.

Judgments instead of explanations

Fitts and Jones remind us that it is counterproductive to say what people failed to do or should have done, since none of that explains why people did what they did. With the intention of explaining why people did what they did, the error methods in question here help investigators reformulate human errors. For instance, they can be labeled as "poor decisions", "failures to adhere to brief", "failures to prioritize attention", "improper procedure", and so forth (Shappell & Wiegmann, 2001, p. 63). These are not explanations, and could not even lead to explanations of performance. They are judgments, made from an after-the-fact, normative position that is based on "unrealistic assumptions of virtual omniscience and unlimited computational power" on part of people inside the situation (Simon, 1969, p. 202). They judge people for not seeing or doing what others, with the benefit of hindsight, would have seen or done. In contrast, the fundamental concept for understanding—not judging—human performance in context is the bounded rationality principle: People use their knowledge to pursue goals in practice, yet people's knowledge is limited, their awareness is finite and multiple goals may compete for their attention. People's behavior is rational, if possibly erroneous, when viewed from the inside of their situations, not from the outside and from hindsight. The point in learning about human error is not to find out where people went wrong. It is to find out why their assessments and actions made sense to them at the time, given how their situation looked from the inside. It is not to say what people failed to do. It is to understand why they did what they did, by probing the systematic, lawful connections between their assessments and actions, and the tools, tasks and environment that surrounded them.

Both error methods also rely on relatively modern human factors concepts: loss of effective CRM; complacency, non-compliance; loss of situation awareness. While masquerading as explanations, these labels too, do little more than saying "human error" over and over again, re-inventing it under a more fashionable guise:

- Loss of CRM (Crew Resource Management) is one name for human error—the failure to invest in common ground, to share data that, in hindsight, turned out to have been significant.

- Complacency is also a name for human error—the failure to recognize the gravity of a situation or to adhere to standards of care or good practice.
- Non-compliance is a name for human error—the failure to follow rules or procedures that would keep the job safe.
- Loss of situation awareness is another name for human error—the failure to notice things that in hindsight turned out to be critical.

Instead of explanations of performance, these labels become judgments. For example, we judge people for not noticing what we now know to have been important data in their situation, calling it their error—their loss of situation awareness. That these kinds of phenomena occur and even help produce trouble is indisputable. People do not coordinate perfectly across workplaces; people adjust their vigilance and their working strategies over time on the basis of their perception of threat; people locally adapt written guidance; and there is always a mismatch between what people observed and what we can show was physically available in their worlds in hindsight. But simply labeling the phenomena fashionably, and stopping there because it now fits a category of some error tool, does not explain the mechanisms that lie beneath the phenomenon—be they psychological, social or otherwise. These labels are an illusion of explanation. It takes a whole different level of analysis (e.g. Vaughan, 1996; Snook, 2000) to begin to understand, for example, how local, contextualized practice can drift from global, centralized guidance. It takes deeper insights to see how people on the line continually face the double bind of needing to adapt rules in the face of unanticipated complexity but risk doing so unsuccessfully, or sticking to the rules but finding that they cannot safely manage the situation with them. Applying procedures is not simple rule-following. It is a substantive cognitive activity. Discovering a mismatch between centralized guidance and local practice is nothing egregious—and not a unique accomplishment of an error tool that counts their instances. The typical countermeasure that emanates from such tabulation of the superficial amounts to admonishing people, exhorting them to follow rules, or introducing even more rules. This will merely serve to tighten the fundamental double bind that people on the line already face and in the end probably increases the gap between procedure and practice.

Shifting blame up the corporate ladder

Both Helmreich's threat and error management model and HFACS try to lead investigators further up the causal pathway, in search of more distal contributors to the failure that occurred. The intention is consistent with the organizational extension of the Fitts and Jones '47 premise (see Maurino *et al.*, 1995) but both models turn it into re-runs of The Bad Apple Theory, if *sub rosa*. A case study used to illustrate Helmreich's aviation error model to medicine (Helmreich, 2000) involved a healthy 8-year old boy choked to death on a clogged endotracheal tube while undergoing minor elective surgery. Unpacking this case of human error through the model leads to the discovery of multiple failures—nothing more than saying "human error" again. The system was basically safe (the boy was healthy), but unreliable people break it. This includes nurses who failed to speak up; a surgeon who continued to operate when he should have stopped; an anaesthetist who failed to intervene. The model simply supplants one failure for the other, while masquerading as explanation. Even when putatively making forays into the organizational antecedents of failure, The Bad Apple Theory prevails. If only the organization had acted earlier on reports about the anaesthetist, this death would not have occurred. There was a bad apple, a weak component, ready to break at any moment in a system with other, ignorant bad apples higher up.

HFACS falls into the same trap. As Shappell & Wiegmann (2001) explain, "it is not uncommon for accident investigators to interview the pilot's friends, colleagues, and supervisors after a fatal crash only to find out that they 'knew it would happen to him some day'." (p. 73) HFACS suggests that if supervisors do not catch these ill components before they kill themselves, then the supervisors are to blame as well. In these kinds of judgments the hindsight bias reigns supreme (see also Kern, 1999). Many sources show how we construct plausible, linear stories of how failure came about once we know the outcome (e.g. Starbuck & Milliken, 1988), which includes making the participants look bad enough to fit the bad outcome they were involved in (Reason, 1997). Such reactions to failure make after-the-fact data mining of personal shortcomings—real or imagined—not just counterproductive (sponsoring The Bad Apple Theory) but actually untrustworthy.

Where front-line operators are dead or non-existent or taken care of, the pursuit of middle managers in safety critical organizations is legendary. From NASA middle managers (Vaughan, 1996) to an AWACS commander (Snook, 2000) to the man responsible for not purchasing ground proximity warning systems for Air Inter's Airbus A320 fleet (METT, 1993), middle management has been identified as responsible for failure and demanded to carry the blame by resigning or being punished otherwise. This relocation of blame corrupts the spirit of going "Beyond Aviation Human Factors" (Maurino *et al.*, 1995). The new view of human error is not an excuse to look for blame higher up in an organization. In fact, Moshansky (1992), Vaughan (1996) and Snook (2000) all remind us that if managers would truly see their people or processes or practices as deficient or unsafe, they would unlikely allow their continuation. Vaughan convincingly dispels the myth that NASA contained "immorally calculating managers" who pushed the production goal in the face of obvious safety flaws, leading to the 1986 Challenger disaster. In its stead she explains how the cumulative normalization of observed O-ring damage was something that connected systematically to features of the organization that people worked in: technical uncertainty, structural secrecy, a culture of production. Managers did not intend the Challenger mission to fail; they did not continuously gamble and finally lost in 1986. They truly believed the accident to be impossible or at least highly improbable or they would not have continued with the operation (see Wagenaar & Groeneweg, 1987).

When studied in the context that brought them forth and that accompanied them, managerial assessments and actions in other cases make similar sense. For example, ground proximity warning systems for an airline that operates short-haul in often mountainous terrain would be a mixed blessing: false alarm rates would severely undermine the system's credibility with pilots and ultimately contribute to an erosion of safety margins instead of bolstering them. The false alarm syndrome is a common and well-documented problem in human factors, and as human factors professionals we should not find it egregious for mid-level supervisors to make managerial decisions on its basis. In the same vein, Snook (2000) explains how mid-level supervisors, in the Airborne Warning and Control (AWACS) aircraft and elsewhere, would have had a hard time preventing the accidental shoot-down of two friendly Black Hawk helicopters over Northern Iraq in 1993, no matter how desperately we want them to take responsibility after the fact.

These analyses demonstrate that we must try to see how people—supervisors and others—interpreted the world from their position on the inside; why it made sense for them to continue certain practices given their knowledge, focus of attention and competing goals. The error methods do nothing to elucidate any of this, instead stopping when they have found the next responsible human up the causal pathway. "Human error", by any other label and by any other human, continues to be the conclusion of an investigation, not the

starting point. This is the old view of human error, re-inventing human error under the guise of supervisory shortcomings and organizational deficiencies. HFACS contains such lists of "unsafe supervision" that can putatively account for problems that occur at the sharp end of practice. For example, unsafe supervision includes "failure to provide guidance, failure to provide oversight, failure to provide training, failure to provide correct data, inadequate opportunity for crew rest" and so forth (Shappell & Wiegmann, 2001, p. 73). This is nothing more than a parade of judgments: judgments of what supervisors failed to do, not explanations of why they did what they did, or why that perhaps made sense given the resources and constraints that governed their work. Instead of explaining a human error problem, HFACS simply re-locates it, shoving it higher up, and with it the blame and judgments for failure. Substituting supervisory failure or organizational failure for operator failure is meaningless and explains nothing. It sustains the fundamental attribution error, merely directing its misconstrued notion elsewhere, away from front-line operators.

Bad Apple Theory Redux

Helmreich's and the HFACS error classification were intended to help investigators understand human error according to the new view. How are the errors observed symptoms of trouble deeper inside a system? Instead, both error classification systems become retreads of the old view—pursuing culprits by any other name or at any other organizational level for their failures to this or the other thing. As we have seen, the misconceptions that put attempts to new view understanding on the road to the old view, to The Bad Apple Theory, are severe and deep. Among them:

- The myth that we can meaningfully count errors and assign them to categories. This would supposedly give us insight into the size and nature of our human error problem. In reality, human error in the wild resists counting and tabulation because (1) it is too diverse and context-dependent; (2) the definition of what is an error is highly problematic; (3) causal networks are intractable and infinite, and the construction of cause arbitrary; and (4) the boundary between system and human failures becomes blurred once we acknowledge the complexity of failure.
- The idea that human error is caused by inherent processing limitations or motivational shortcomings. This notion, of human error as purely endogenous, would reduce the contribution of human factors to system safety to something decorative, marginal and irrelevant. In reality, human error hinges on the tools and other resources we give people and the tasks we ask them to carry out. Human factors aims to influence the circumstances (engineered, social, organizational) under which people work in order to improve their performance.
- The illusion that judgments, under fashionable guises, can double as explanations of human error. In reality, saying what people failed to do or should have done does not explain why they did what they did. The starting point, and continued empirical regularity, of human factors work is the local rationality principle: people do reasonable things given their knowledge, their focus of attention and their (competing) goals. Understanding people's performance means seeing their situations from the inside; trying to grasp why it made sense for them to do what they did.
- The idea that we should seek sources of failure higher up, away from the local sharp end operators. This idea, while consistent with the spirit of human factors over the past decades, quickly becomes a retread of the old view of human error, slipping into judgments instead of explanations of supervisory or organizational performance.

> Instead of dispelling the notion of blame and final responsibility for failure, the classification methods help move blame, shift it around and find a place to park it higher up.

Classification of errors is not analysis of errors. Categorization of errors cannot double as understanding of errors. The gaps that these classification tools leave in our insight into human performance are daunting. In addition, they reinforce and entrench the misconceptions, biases and errors that we always risk making in our dealings with failure, while giving us the illusion we have actually embraced the new view to human error. The step from classifying errors to pursuing culprits appears a small one, and as counterproductive as ever. In aviation, we have seen The Bad Apple Theory at work and now we see it being re-treaded around the wheels of supposed progress on safety. Yet we have seen the procedural straight-jacketing, technology-touting, culprit-extraditing, train-and-blame approach be applied, and invariably stumble and fall. We should not need to see this again. For what we have found is that it is a dead end. There is no progress on safety in the old view of human error.

Reinventing Human Error II: Breakdowns in the creation of safety

We can make progress on safety once we acknowledge that people themselves create it, and we begin to understand how. Safety is not inherently built into systems or introduced via isolated technical or procedural fixes. Safety is something that people create, at all levels of an operational organization (e.g. AMA, 1998; Sanne, 1999). Safety is the emergent property of a system of people who invest in their awareness of potential pathways to breakdown and devise strategies that help forestall failure. Take, for instance, the case of the "failure" to arm the spoilers as described above: One pilot explained how he, after years of experience on a particular aircraft type, figured out that he could safely arm the spoilers 4 seconds after "gear down" was selected, since the critical time for potential gear compression was over by then. He had refined a practice whereby his hand would go from the gear lever to the spoiler handle slowly enough to cover 4 seconds—but it would always travel there first. He then had bought himself enough time to devote to subsequent tasks such as selecting landing flaps and capturing the glide slope. This is how practitioners create safety: they constantly invest in their understanding of how systems can break down; invest in their understanding of where they themselves are vulnerable to error, and then they devise strategies that help forestall failure. At the lowest operational level—in direct, close contact with the engineered system that mismatches the workflow, people introduce extra slack by re-timing subtasks. By doing so they de-couple the potential slip of not arming the spoilers from the circumstances that make that error most likely. They entrench a strategy of arm-movement, inserting a layer of standard double-checking that did not exist before. People, in other words, create safety.

This provides additional insights into the myth of the error count. It would seem that 70% human errors represents the distance we have to go before we reach full safety. Full safety lies somewhere on, or beyond, the horizon, and the 70% human errors is what is between us and that goal. This assumption about the location of safety is an illusion, and efforts to measure the distance to it are little more than measuring our distance from a mirage. Safety is right here, right now, right under our feet—not across some 70%. Look back at the spoiler example above. People in complex systems create safety. They make it their job to anticipate forms of, and pathways toward, failure, they invest in their own resilience and that of their system by tailoring their tasks, by inserting buffers, routines, heuristics, tricks,

double-checking, memory aids. The 70% human contribution to failure occurs because complex system need an overwhelming human contribution for their safety. Human error is the inevitable by-product of the pursuit of success in an imperfect, unstable, resource-constrained world. To try to eradicate human error (to depress or reduce the 70%) would mean to eradicate or compromise human expertise—the most profound and most reliable investment in system safety and success we could ever hope for. In order to understand the 70% human error, we need to understand the 70% (or more) contribution that human expertise makes to system success and safety.

Many other examples of the creation of safety, at many other levels, exist. The decision of an entire airline to no longer accept NDB approaches (Non-Directional Beacon approaches to a runway, in which the aircraft has no vertical guidance and rather imprecise lateral guidance) (Collins, 2001) is one example; the reluctance of airlines and/or pilots to agree on LASHO—Land And Hold Short Operations—which put them at risk of traveling across an intersecting runway that is in use, is another. In both cases, goal conflicts are evident (production pressures versus protection against known or possible pathways to failure). In both, the trade-off is in favor of safety. In resource-constrained systems, however, safety does not always prevail. RVSM (Reduced Vertical Separation Minima) for example, which will make aircraft fly closer together vertically, will be introduced and adhered to, mostly on the back of promises from isolated technical fixes that would make aircraft altitude holding and reporting more reliable. But at a systems level RVSM tightens coupling and reduces slack, contributing to the risk of interactive trouble, rapid deterioration and difficult recovery (Perrow, 1984).

Another investment in system resilience that is gaining a foothold in the aviation industry is the automation policy, first advocated by Wiener (e.g. 1989) but still not adopted by many airlines. Automation policies are meant to reduce the risk of coordination breakdowns across highly automated flight decks, their aim being to match the level of automation (high, e.g. VNAV (Vertical Navigation, done by the Flight Management System); medium, e.g. heading select; or low, e.g. manual flight with flight director) with human roles (pilot flying versus pilot not-flying) and cockpit system display formats (e.g. map versus raw data) (e.g. Goteman, 1999). This is meant to maximize redundancy and opportunities for double-checking, capitalizing on the strengths of available flightdeck resources, both human and machine.

The success of failure

People are not perfect creators of safety. There are patterns, or mechanisms, by which their creation of safety can break down—mechanisms, in other words, by which failure succeeds. Take the case of a DC-9 that got caught in windshear while trying to go around from an approach to Charlotte, NC, in 1994 (NTSB, 1995). Noting the dynamic weather situation in front of them, the crew invests continually in their awareness of the potential threat ahead. They ask for pilot reports from aircraft on the approach in front of them, and consistently receive information (e.g. "smooth ride all the way down") that confirms that continuing the approach makes sense. Given the ambiguity of the situation (a storm is visible, yet pilot reports are all smooth), effective feedforward is difficult. What do you do—when do you go around? As an additional investment in safety, the crew plans to turn right in case they will have to go around. Adapting a procedure to deal better with local circumstances, the planned missed approach route will take them away from the storm to their left. Once close to the runway the DC-9 enters rain and experiences airspeed

variations, and the crew decides to go around, turning right while doing so. What they are flying in at that moment is a microburst—a huge package of colder air falling down and hitting the ground, spreading rapidly in all directions. Where the crew flies in during their go-around turns out to be the worst place: it gives them the most tailwind. This is the stochastic element (Snook, 2000) or bad luck (Reason, 1990) almost always present in breakdown scenarios. While trying to buy more slack, more buffer, the decision to go right instead erodes opportunities to recover because of the nature of the microburst (a nature which nobody can really foresee). Once inside that situation, effective feedback becomes very difficult. Airspeed decreases rapidly and in the end the DC-9 cannot remain airborne. It starts to hit trees and breaks apart on a residential street off the airport. Charlotte is a case where people are in a double bind: first, things are too ambiguous for effective feedforward. Not much later things are changing too quickly for effective feedback. While approaching the airport, the situation is too unpredictable, the data too ambiguous, for effective feedforward. In other words, there is insufficient evidence for breaking off the approach (as feedforward to deal with the perceived threat). However, once inside the situation, things change too rapidly for effective feedback. The microburst creates changes in winds and airspeeds that are difficult to manage, especially for a crew whose training never covered a windshear encounter on approach or in such otherwise smooth conditions.

Charlotte is not the only pattern by which the creation of safety breaks down; it is not the only mechanism by which failure succeeds. For progress on safety we should abandon the construction of cause—in error classification methods or any other investigation of failure. Once we acknowledge the complexity of failure, and once we acknowledge that safety and failure are emerging properties of systems that try to succeed, the selection of causes—either for failure or for success—becomes silly and pointless. Instead of constructing causes, we should try to document and learn from patterns of failure. Snook (2000) anticipates my argument, calling big accidents "treasures" that scientists can mine for behavioral mysteries:

"If we could put together a library of such treasures, thick behavioral descriptions of complex untoward events, I'm confident that such studies will move us closer to unlocking the fundamental design mysteries of hyper-complex organizations." (p. 236)

Have we already unlocked some of the mysteries? What are the mechanisms by which failure succeeds? Can we already sketch some? What patterns of breakdown in people's creation of safety do we already know about? Charlotte, the case described above—too ambiguous for feedforward, too dynamic for effective feedback—is one mechanism by which people's investments in safety are outwitted by a rapidly changing world. Understanding the mechanism means becoming able to retard it or block it, by reducing the mechanism's inherent coupling; by disambiguating the data that fuels its progression from the inside. The contours of many other patterns, or mechanisms of failure, are beginning to stand out from thick descriptions of accidents in aerospace, including the normalization of deviance (Vaughan, 1996), the going sour progression (Sarter & Woods, 1997), practical drift (Snook, 2000) and plan continuation (Orasanu *et al.,* in press). Investing further in these and other insights will represent progress on safety. There is no efficient, quick road to understanding human error, as the error classification methods discussed in this paper make us believe. Their destination will be an illusion, a retread of the old view. Similarly, there is no quick safety fix for systems that pursue multiple competing goals in a resource-constrained, uncertain world. There is, however, percentage in opening the black box of human performance—understanding how people make the

systems they operate so successful, and probing the patterns by which their successes are defeated. In fact, such understanding is our only hope for real progress on safety.

Acknowledgements

The work for this paper was supported by a grant from the Swedish Flight Safety Directorate and its Director Mr. Arne Axelsson.

References

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.

Billings, C. E. (1996). *Aviation automation: The search for a human-centered approach*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

Collins, R. L. (2001). No NDB. *Flying, 128*(1), 18.

Dekker, S. W. A. (2001). The disembodiment of data in the analysis of human factors accidents. *Human Factors and Aerospace Safety, 1*(1), 39-57.

Dougherty, E. M. Jr. (1990). Human reliability analysis: Where shouldst thou turn? *Reliability Engineering and System Safety, 29*(3), 283-299.

Federal Aviation Administration (1996). *The interface between flightcrews and modern flight deck systems*. Washington, DC: FAA.

Fitts, P. M., & Jones, R. E. (1947). Analysis of factors contributing to 460 'pilot error' experiences in operating aircraft controls. *Memorandum Report TSEAA-694-12*, Aero Medical Laboratory, Air Material Command, Wright-Patterson Air Force Base, Dayton, Ohio, July 1, 1947.

Goteman, O. (1999). Automation policy or philosophy? Management of automation in the operational reality. In S. W. A. Dekker & E. Hollnagel (Eds.), *Coping with computers in the cockpit,* pp. 215-222. Aldershot, UK: Ashgate Publishing Co.

Helmreich, R. L., Klinect, J. R., & Wilhelm, J. A. (1999). Models of threat, error and response in flight operations. In R. S. Jensen (Ed.), *Proceedings of the tenth international symposium on aviation psychology*. Columbus, OH: The Ohio State University.

Helmreich, R. L. (2000). On error management: Lessons from aviation. *BMJ*, 320, 745-753.

Hollnagel, E. (1998). *Cognitive reliability and error analysis method (CREAM)*. Oxford UK: Elsevier Science.

Kern, T. (1999). *Darker shades of blue: The rogue pilot*. New York, NY: McGraw-Hill.

Maurino, D. E., Reaon, J. T., Johnston, N., & Lee, R. B. (1995). *Beyond aviation human factors*. Aldershot, UK: Ashgate Publishing Co.

Ministère de l'equipment, des transports et du tourisme (1993). *Rapport de la commission d'enquete sur l'accident survenu le 20 Janvier 1992 près du Mont Saint Odile (Bas Rhin) a l'Airbus 320 immatriculé F-GGED exploité par la compagnie Air Inter*. Paris, France: METT.

Moshansky, V. P. (1992). *Commission of Inquiry into the Air Ontario accident at Dryden, Ontario* (Final Report, Vol. 1-4). Ottawa, ON: Ministry of Supply and Services, Canada.

National Transportation Safety Board (1995). *Flight into terrain during missed approach. USAir Flight 1016, DC-9-31, Charlotte/Douglas International Airport, Charlotte, NC, 7/2/94* (NTSB Rep. No. AAR/95/03). Washington, DC: NTSB.

Norman, D. A. (1994). Things that make us smart: Defending human attributes in the age of the machine. New York: Perseus Press.

Orasanu, J., Martin & Davison (in press). Sources of decision error in aviation. In G. Klein and E. Salas (Eds.), *Applications of naturalistic decision making*. Mahwah, NJ: Lawrence Erlbaum Associates.

Perrow, C. (1984). *Normal accidents: Living with high-risk technologies*. New York, NY: Basic books.

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.

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.

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

Sarter, N. B., & Woods, D. D. (1997). Teamplay with a powerful and independent agent: Operational experiences and automation surprises on the Airbus A-320. *Human Factors, 39*(4), 553-569.

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. A. (2000). *Friendly fire: The accidental shootdown of US Blackhawks over Northern Iraq*. Princeton, NJ: Princeton University Press.

Starbuck, W. H., & Milliken, F. J. (1988). Challenger: Fine-tuning the odds until something breaks. *Journal of Management Studies, 25*(4), 319-340.

Vaughan, D. (1996). *The Challenger lauch decision: Risky technology, culture and deviance at NASA*. Chicago, IL: University of Chicago Press.

Wagenaar, W. A., & Groeneweg, J. (1987). Accidents at sea: Multiple causes and impossible consequences. *International Journal of Man-Machine Studies, 27*, 587-589.

Weick, K. E. (1990). Organizational culture as a source of high reliability. *California Management Review, 29*, 112-127.

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.

Woods, D. D., Johannesen, L. J., Cook, R. I., & Sarter, N. B. (1994). *Behind human error: Cognitive systems, computers and hindsight*. Dayton, OH: CSERIAC.