Illusions of explanation: A critical essay on error classification

International Journal of Aviation Psychology, Vol. 13, Issue 2, pp. 95-106

Sidney Dekker
Centre for Human Factors in Aviation, IKP
Linköping Institute of Technology
SE - 581 83 Linköping
Sweden
tel. + 46 13 281646
fax. + 46 13 282579
sidde@ikp.liu.se

Abstract

Error classification methods are used throughout aviation to help understand and mitigate the causes of human error. Many assumptions underlying error classification, however, remain untested. For example, "error" is taken to mean different things, even within individual methods; and a close mapping is uncritically presumed between the quantity measured (errors) and the quality managed (safety). Further, error classifications can deepen investigative biases by merely relabeling error, rather than explaining it. This essay reviews such assumptions and proposes alternative ways forward.

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

Introduction

Why do we want to classify human errors? Classification of observed phenomena is basic to science. It serves to order empirical reality as we encounter it; it creates a causal structure that supports our understanding of phenomena. Classification of human errors has practical reasons along the same lines: it can help managers and engineers understand and presumably manage ways in which people contribute to system reliability and breakdown. Aviation Human Factors has produced categorizations such as Kowalsky *et al.* (1974), who classified decision errors together with the conditions that helped produce them; Billings & Cheaney (1981), who categorized information transfer problems (e.g. instructions; errors during watch change-over briefings; coordination failures); Fegetter (1982), who divided error causes into cognitive, social and situational (physical/environmental/ergonomic) systems; and Rouse & Rouse (1983), who categorized error causes along the lines of a linear information processing/decision making model. Currently, LOSA (Line Oriented Safety Audit) is making its rounds through the aviation industry as error classification and tabulation tool (e.g. Helmreich *et al.*, 1999). One reason for its popularity is the idea that it can help demonstrate how Crew Resource Management (CRM) training helps improve aviation safety (see Croft, 2001). The adoption of CRM (even by regulators) was predicated on the belief that it would lead to better safety (Helmreich & Foushee, 1993). Yet such training is expensive and needs to be justified in the face of doubts whether the link between CRM training and improved safety exists in the first place (cf. Maurino, 1999; JAA, 2001). Counting and tabulating errors in relation to CRM interventions is thought to yield such justification (Croft, 2001).

The aim of error classification tools is as simple in principle as it is difficult in practice: go beyond the superficial "error" and probe the system for underlying reasons why it occurred. Human error cannot be the conclusion of an investigation, it has to be the starting point (Woods *et al.*, 1994). This has long been imperative in (aviation) human factors (Fitts and Jones, 1947): since human error is systematically connected to features of people's tools and tasks, we need to work on those tools and tasks if we want to prevent recurrence (Maurino *et al.*, 1995; Reason, 1990; 1997). Error classification, however, can be seen to run into a number of problems in achieving such an aim in practice. This paper looks at three of the more acute issues:

- Classification of errors is easily mistaken for analysis and deeper understanding.
- Finding deeper reasons for the observed error is often a matter of finding other errors, either inside the heads of the people affected, or by other people;
- Safety is modeled as the absence of "negatives" (errors), misguiding managerial interventions.

It then looks for alternative ways forward.

Mistaking classification for understanding

Croft (2001) explains how the LOSA method of error counting, popular in commercial aviation, asks observers of line practice to rate aircrew errors according to the following five categories: 1) Intentional non-compliance errors (conscious violations against standard operating procedures or regulations); 2) Procedural errors (slips, lapses or mistakes); 3) Communication errors (incorrectly transmitted or interpreted information); 4) Proficiency errors (due to a lack of knowledge or basic flying skills); and 5) Operational decision errors (discretionary decisions not covered by regulation and procedures that unnecessarily increase risk). These categories have been applied to observations of 1426 commercial airliner flights since 1997. Interestingly, more than half the human errors detected by observers were never detected (or classified as errors) by the flight crews themselves. This is chalked up as a success of the method (Croft, 2001) rather than as a warning of potentially misleading data, or an indication of a mismatch in perspectives on what "error" really means. In fact, the first problem in counting errors occurs is that observers have to agree what they mean by "error". In today's aviation error classification systems, error can mean several things:

- Error as the *cause* of failure. For example: This event was due to human error. Classifications rely on this definition when seeking the cause of operator error in, for instance, a supervisor's "failure to provide guidance" (Shappell & Wiegman, 2001, p. 73).
- Error as the *failure itself*. For example: The operator's decision was an error (e.g. Helmreich, 2000). Classifications rely on this definition when categorizing the kinds of observable errors operators can make (e.g. decision errors, perceptual errors, skill-based errors) (Shappell & Wiegmann, 2001).
- Error as a *process*, or, more specifically, as a departure from some kind of standard. This standard may consist of operating procedures. Violations, whether exceptional or routine (Shappell & Wiegmann, 2001), or intentional or unintentional (Helmreich, 2000), are one example of error according to the process definition. Depending on what we use as standard, we of course come to different conclusions about what is an error.

Not differentiating among these different possible definitions of error is a well-documented problem (Dougherty, 1990; Hollnagel, 1998): error classification schemes often display this inability to sort out what is cause and what is consequence; what is genotypical, what is phenotypical. LOSA contains categories of manifestations of error (such as communication errors) as well as causes of error (such as proficiency problems), thus it too mixes causes and consequences. This does not help our understanding of error.

Without clear definitions, or models of error, error counting amounts to pseudoscience or numerology. Measurement is of course fundamental to science. But measuring without an explicit underlying model that directs observations and allows classification is folk science. The measurements typical of error counting methods are products of folk models—commonly held notions about the nature of human work. Folk models are inexplicit connections between the behavioral particulars that are measured and the condition they point to. Folk models encourage observers to measure what can be superficially observed (and thus measured) simply because it *can* be measured. In actual scientific endeavors, the definition of a measurement depends on how the corresponding domain or phenomenon is theoretically described or explained (Hollnagel, 1998). As Einstein said to Heisenberg: 'whether you see a thing or not depends on the theory which you use. It is the theory which decides what can be observed' (quoted in Angell & Straub, 1999, p. 187). The measurement presupposes a clarification of what the model behind the measurement is, where a model is understood as a simplified representation of the salient features of the target situation. The model constrains what can be measured by describing the essential elements of performance and the model parameters thereby become the basis for specifying the measurements (Hollnagel, 1998).

Folk modeling makes error classification fundamentally untestable. For, without an underlying analytic model, who can go back and challenge the conclusions of the observer once the moment of observation has receded into history and the excised "error" has disappeared into the categorization tool, accompanied with its hypothesized mental function (e.g. prioritizing attention) or organizational defect (e.g. oversight failure) as "explanation"? No such testing is possible—denying this proto-scientific activity of error classification even the most basic scientific quality-control since Karl Popper: falsification. Observed practitioners themselves may have an opportunity to object, of course. But if observers claim that their method has succeeded when its error count is twice as high as that of the practitioners themselves (Croft, 2001), there can be little confidence that the practitioner perspective is afforded any relevance at all.

The disembodiment of data

Second, understanding is retarded when what can be known about human performance is replaced by what can be forced into five categorical labels of an error counting method. Attempts to map situated human capabilities such as decision making, proficiency or deliberation onto discrete categories are doomed to be misleading for they cannot cope with the complexity of actual practice without serious degeneration (Angell & Straub, 1999). Error classification disembodies data. It removes the context that helped produce the error in its particular manifestation. This disables understanding because by excising performance fragments away from their context, error classification destroys the local rationality principle. This has been the fundamental concept for understanding—not judging—human performance for the last fifty years: People's behavior is rational, if possibly erroneous, when viewed from the inside of their situations, not from the outside and from hindsight. The local rationality principle also reminds us that the consequences of actions are not well-correlated with intentions, yet this completely evaporates in the wake of error classification. 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 their knowledge, goals, tools and limited resources. For we have to assume that if it made sense to someone (given the background and circumstances), it will make sense to someone else too, and the "error" will repeat itself. Controversial behavior can be made to make sense (read: understood) once resituated in the context that brought it forth (Vaughan, 1996; Snook, 2000; Dekker, 2001). Indeed, Kowalsky's (1974) integration of context in classifying errors anticipates (if crudely) later critiques from for example Dougherty (1990), who decries the insufficient ability of error classifications to describe how context wields an influence on human assessments and actions that get classified as "erroneous".

Once the observation of some kind of "error" is tidily locked away into some category, it has been objectified, formalized away from its context. Without context, there is no way to re-establish local rationality. And without local rationality, there is no way to understand human error. Error categorization presents an illusion of understanding. It disconnects human agents' performance from the context that brought it forth, from the circumstances that accompanied it; that gave it meaning; and that hold the keys to its explanation. Instead it renders performance fragments disembodied: as uncloaked, context-less, meaningless shrapnel scattered across superficial categories in the wake of the auditors' 1426th assigning frenzy. Error categorization is not equivalent to understanding error, and perhaps not even the beginning of understanding error. In fact, it may be the opposite.

Explaining error by finding errors elsewhere

The common aim of error classification systems is to try to find, through categorization, what lies behind observed errors. Further explanation typically takes one of two forms: it is sought either in shortcomings of hypothesized information processing structures in people's minds (Rouse & Rouse, 1983; Kirwan & Ainsworth, 1992) or it is sought in organizational deficiencies that surrounded people at the time (e.g. Shappell & Wiegmann, 2001). None of this, however, explains error—it simply displaces the interpretative load.

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). This is not very enlightening, because in this "explanation" the errors are simply the result of other errors (e.g. "flawed" decision making). It is not explanation, but re-labeling. Similarly, Shappell & Wiegmann (2001, p. 63) suggest that observed errors can be labeled as "poor decisions", "failures to adhere to brief", "failures to prioritize attention", "improper procedure", and so forth. Such reformulations of error, too, are illusions of deeper understanding. Yet it is typical for error classification methods, especially those dominated (implicitly or explicitly) by information processing approaches to human factors. Performance problems are "explained" by reference to inherent limitations of mental processing mechanisms or other human shortcomings. For example, Rouse & Rouse (1983) seek the source of error somewhere along a hypothesized linear psychic highway, that strings together the head's input and output through a fixed sequence of representational traffic stops: observation of system state; choice of hypothesis; testing of hypothesis; choice of goal; choice of procedure; execution of procedure. They are not alone in seeking enlightenment from such quarters (e.g. Kirwan & Ainsworth, 1992), and thus all share the major shortcomings. Errors are not explained. They merely get displaced from the outside to the inside; from observable actions to psychic misfirings. As a consequence, the classification is totally unverifiable.

Fitts' & Jones' (1947) orientation, as well as that of later human factors work, has always been more ecological: performance problems can be understood by reference to constraints that the world imposes on people's goal-directed behavior. Put crudely: if you want to understand what went on in the mind, look in the world in which the mind found itself, instead of trying to pry open the mind. Constraints in the world can for example arise from the engineered interface (which, by the way, is nowhere to be found in LOSA) or the organizational context (e.g. Maurino et al., 1995). Yet here too, an illusion of explanation slips easily into error classification methods. When looking for organizational contributors to operator error, the reasons are simply sought in errors that occurred higher up in the chain of command. For example, operator errors can be "understood" on the basis of unsafe supervision, which 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 too, simply re-introduces "human error" in a different cloak, or by a different human.

Also, these putative "explanations" only judge people for not doing what they (in hindsight) should have done. Modern human factors concepts of this sort heavily populate the various error classifications, for example loss of effective CRM; complacency, non-compliance; loss of situation awareness. While masquerading as explanations, these labels do little more than saying "human error" over and over again, judging performance instead of explaining it:

- 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. We merely 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 to them in hindsight. But simply labeling these phenomena fashionably, and stopping there because it now fits a category of the error classification, does not explain anything.

Safety as the absence of negatives

Error classification methods assume that safety is a positivistic empirical given. Safety is out there to be discovered by the auditor with the right labeling tool. It can subsequently be forced in certain directions by managers wielding the numbers. For example, pilots who violate rules are 1,4 times more likely to commit operational errors (Helmreich *et al.*, 1999; Klinect *et al.*, 1999). For a manager, such digits suggest not only that violators are less reliable components in a system than pilots who do not violate (sponsoring "the bad apple theory", see Dekker (2001)), but that demotions, stringent training or supervision might take care of the problem.

All of this presents various problems. Error classification methods presume a close mapping between the quantity measured (numbers of errors) and the quality investigated or managed (safety). But safety is more than the measurement and management of negatives (errors), if it is that at all. There is little or no evidence that "safety" is a positivistic given that exists 'out there', independent of operators' minds or their surrounding culture, ready to be measured by an etic probe. What research has shown instead is that it is a "constructed human concept" (Rochlin, 1999, p. 1550). This research in human factors has begun to probe how individual practitioners construct safety, by assessing what they understand risk to be, and how they perceive their difficulty of managing challenging situations (Orasanu, 2001). A substantial part of practitioners' construction of safety turns out to be self-referential, assessing the pilot's own competence or skill in maintaining safety across different situations. Amalberti (e.g. 2001) found this too: pilots will do a lot (in terms of planning, preparation, anticipation, etc.) to not have to make difficult decisions that might get them into trouble. Interestingly, Orasanu discovered a mismatch between risk salience (how critical a particular threat to safety was perceived to be by the practitioner) and frequency of encounters (how often these threats to safety are in fact met in practice). The safety threats deemed most salient were the ones least frequently dealt with.

Given these results, it is no wonder that good empirical indicators of social and organizational definitions of safety are difficult to obtain. Operators of reliable systems "were expressing their evaluation of a positive state mediated by human action, and that evaluation reflexively became part of the state of safety they were describing" (Rochlin, 1999, p. 1550). In other words, the description itself of what safety means to an individual operator is a part of that very safety, dynamic and subjective. "Safety is in some sense a story a group or organization tells about itself and its relation to its task environment" (Rochlin, 1999, p. 1555). Clearly, such aspects of safety can only be captured by a less etic, numerical approach. It requires a more emic one, that probes the interpretative aspect of situated human assessments and actions.

By treating safety as positivistically measurable, error counting breathes the scientific spirit of a bygone era in human factors. It is a holdover of how performance was gauged (by counting errors) in the laboratory, testing limited, contrived task behavior in spartan settings that kept people's cognition in captivity. Because of its apparent simplicity, there is industry enthusiasm and human factors hysteresis to continue with such a forcedly positivistic practice, even as human factors makes ever deeper forays into the natural settings in which people carry out actual complex, dynamic and interactive work. The idea of a positivistic count is compelling to industry for the same reasons that any numerical performance measurement is (e.g. Elg, 2001). Error counting becomes a quantitative basis for managerial interventions. Pieces of data from the operation that have been excised and formalized away from their origin can be converted into graphs and barcharts which are subsequently engineered into interventions. Never mind that the barcharts show comparisons between apples and oranges (causes and consequences of error) that kid managers into believing they have learned something of value. It does not matter because managers, and their airlines, can elaborate their idea of control over operational practice and its outcomes. It is optimistic, positivistic progressivism. It is also illusory. The real world is not so easily fooled: managerial "control" exists only in the sense of purposefully formulating and trying to influence pilots' intentions and actions (Angell & Straub, 1999), which, if done by way of sanctioning pilots, amounts to pre-historic behaviorism. In any case it is not at all the same as being in control of the consequences (by which safety ultimately gets measured industry-wide). Because for that the real world is too complex and operational environments too stochastic (e.g. Snook, 2000). The managerially appealing numbers are quite sterile, inert. They do not reflect any of the nuances of what it is to "be there", doing the work, creating safety on the line (e.g. Sanne, 1999). Yet this is what ultimately determines safety (as outcome): people's local actions and assessments are shaped by their own perspectives—self-referential; embedded in histories, rituals, interactions, beliefs and myths, both of their organization and them as individuals.

In the face of compelling numbers, none of this seems to matter. Recall how Croft (2001) reported completed safety audits on 1426 line flights since 1997. It would seem that the aviation industry has reached a point where progress on safety gets equated with the number of airlines that have undergone such audits, and whether pan-industry organizations have embraced the error count as a way forward (Croft, 2001). Error counting, and a managerial dependence on quantifiable bases for decision making, has created an arena in which the entire aviation industry is becoming drawn into a self-sustaining loop. The problem is that managerial interventions (training, procedures, sanctioning etc.) not only justify themselves on the basis of measurement, but in turn confirm the utility of the error count. If auditors count errors for managers, they, as ("scientific") measurers, have to presume that errors exist. But in order to prove that errors exist, auditors have to measure them. In other words, measuring errors becomes the proof of their existence, an existence that was pre-ordained by their measurement. Angell and Straub (1999, p. 184) call such self-propagation "consensus authority": Everyone agrees that counting errors is a good way forward on safety because almost everyone seems to agree that it is a good way forward. The practice is not questioned because few seem to question it.

Real ways forward: De-emphasizing the construction of cause

Error classifications formalize, or aid, the search for causes (of human errors). They make this search more efficient, more speedy. But cause is not something we find—with or without the help of error classifications. Cause is something we construct. "The cause" is simply "that factor" where we stop looking any further—for whatever reason (perhaps because the error classification method did not provide any further labels). Error classification promise snappy insight into the reasons for error and suggest that there is a quick safety fix. But systems that pursue multiple competing goals in a resource-constrained, uncertain world resist quick fixes. The construction of cause is our final illusion of understanding. Practice that goes wrong in these systems 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 *really* trace "the cause" of failure, the causal network would fan out immediately, like cracks in a window, with only we determining when and where 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 ultimately arbitrary.

Several recent investigations of human error (all consistently huge and still incomplete) have drifted away from the construction of cause, and moved towards the identification of patterns, or genotypical mechanisms of failure that appear to recur in different manifestations. For example:

- Vaughan (1996) describes the normalization of deviance, by which signals of increasing danger are normalized because the organization's repetitive exposure to them continually creates new norms. Every new instance of deviance is only a marginal departure from the new norm, and has no clear safety consequences;
- Sarter & Woods (1997) describe the going sour progression, where a series of misassessments and miscommunications between people and automated systems are necessary to push a system over the edge of breakdown (see also FAA, 1996);
- Snook (2000) describes practical drift, the mechanism by which a mismatch between procedures and practice grows undetected over time as a result of normally loosely coupled operations that both encourage and allow local procedural adaptations in the face of pressures to succeed;
- Orasanu *et al.* (in press) describe plan continuation, where practitioners continue with a plan of action in the face of cues that, in hindsight, warranted changing the plan.

These investigations (cf. Moshansky, 1992—the hugest of them all) attempt to understand how people cannot be perfect creators of safety in operational worlds where resources (e.g. time, computational capacity) are limited and multiple goals compete for attention. The investigations uncover systematic patterns whereby people's normally successful creation of safety appears to break down. This may be where real progress on safety lies. Not in counting positivistic "negatives" in as many episodes of practice as possible (i.e. even more than 1426), but in beginning to understand how operational people see and create safety in practice themselves, and how universal patterns of breakdown occur repeatedly across operational particulars. This requires human factors to pursue the accumulation of research experience, and the building of theories, that emphasize the constructivist nature of safety; that encapsulate not just people's activities but their (and their organizations') self-reference, histories, rituals, interactions, beliefs and myths.

Acknowledgments

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

References

Amalberti, R. (2001). The paradoxes of almost totally safe transportation systems. *Safety Science*, 37, 109-126.

Angell, I. O., & Straub, B. (1999). Rain-dancing with pseudo-science. *Cognition, Technology and Work*, 1, 179-196.

Billings, C. E., & Cheaney, E. S. (1981). *Information transfer problems in the aviation system* (NASA Technical Paper 1875). Moffett Field, CA: NASA Ames Research Center.

Croft, J. (2001, July 16). Researchers perfect new ways to monitor pilot performance. *Aviation Week & Space Technology*, pp. 76-77.

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.

Fegetter, A. J. (1982). A method for investigating human factors aspects of aircraft accidents and incidents. *Ergonomics, 25*, 1065-1075.

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.

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.

Klinect, J. R., Wilhelm, J. A., & Helmreich, R. L. (1999). Threat and error management: Data from line operations safety audits. In R. S. Jensen (Ed.), *Proceedings of the tenth international symposium on aviation psychology*. Columbus, OH: The Ohio State University.

Kirwan, B., & Ainsworth, L. K. (1992). *A guide to task analysis*. London: Taylor & Francis.

Kowalsky, N. B., Masters, R. L., Stone, R. B., Babcock, G. L., & Rypka, E. W. (1974). *An analysis of pilot error related to aircraft accidents*. (NASA CR-2444) Washington, DC: NASA.

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

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.

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.

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.

Rouse, W. B., & Rouse, S. H. (1983). Analysis and classification of human error. *IEEE Transactions on Systems, Man and Cybernetics, SMC-13*, 539-549.

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.

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

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

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