In the system view of human factors, who is accountable for failure and success?

Sidney Dekker
Lund University
Lund, Sweden

Abstract

Human factors research and practice is tentatively exploring exciting developments in science—the embrace of systems thinking and complexity—and what it might mean for our understanding of human-machine systems and their successes and failures. Rather than following reductive logic, and looking for the sources of success and failure in system components, complexity and systems thinking suggests that we see performance as an emergent property, the result of complex interactions and relationships. A central problem today is that this may clash with how we think about accountability. When systems fail, we tend to blame components (e.g. human errors and the "villains" who make them), and when they succeed spectacularly, we tend to think in terms of individual heroism (e.g. the A320 Hudson River landing). In this paper, I lay out the contrast between a Newtonian ethic of failure, followed in much human factors work, and one inspired by complexity and systems thinking.

Introduction

Complexity is a defining characteristic of high-technology, high-consequence systems today (e.g. Perrow, 1984), yet componential explanations that condense accounts of failure down to some individual human action or inaction still reign supreme. An analysis by Holden (2009) showed that between 1999 and 2006, 96% of investigated US aviation accidents were attributed in large part to the flight crew. In 81%, people were the *sole* reported cause. The language used in these analyses has judgmental or even moralistic overtones too. "Crew failure" or a similar term appears in 74% of probable causes and the remaining cases contain language such as "inadequate planning, judgment and airmanship," "inexperience" and "unnecessary and excessive . . . inputs." "Violation" of written guidance was implicated as cause or contributing factor in a third of all cases (Holden, 2009). Single-factor, judgmental explanations for complex system failures are not unique to aviation—they are prevalent in fields ranging from medicine (e.g. Catino, 2008), to military operations (e.g. Snook, 2000), to road traffic (Tingvall & Lie, 2010).

The problem of safety analysis reverting to condensed and individual/componential explanations rather than diffuse and system-level ones (Galison, 2000) was one

In D. de Waard, A. Axelsson, M. Berglund, B. Peters, and C. Weikert (Eds.) (2010). *Human Factors: A system view of human, technology and organisation* (pp. 9 - 28). Maastricht, the Netherlands: Shaker Publishing.

driver behind establishing the fields of human factors and system safety (Fitts & Jones, 1951; Dekker, 2005; Leveson, 2006). A charter of these fields has been to take safety scientists and practitioners behind the label "human error," to more complex stories of how normal system factors contribute to things going wrong (including factors associated with new technology and with the organization and its multiple interacting goals) (Hollnagel, ETTO; Woods et al., 1994; Woods & Cook, 1999). The difficulty in achieving such deeper systems stories of failure has been considered from a variety of angles, including psychological (e.g. Fischhoff, 1975; Reason, 1990), sociological (Vaughan, 1999; Rochlin, 1999), and political (Perrow, 1984). Additional literatures are explicitly trying to depart from the conceptualization of accidents as linear progressions of successive component failures/malfunctions (e.g. Rasmussen, 1997; Leveson, 2006, Hollnagel et al., 2006), supplanting it with notions of complexity, non-linearity and accidents as emergent properties. Despite these initiatives, common discourse about failure in complex systems remains tethered to language such as "chain-of-events", "human error" and questions such as "what was *the* cause?" and "*who* was to blame?" (Catino, 2008).

This paper takes an angle not yet well considered in the literature—the philosophical-historical and ideological bases for sustained linear, componential thinking about failure in complex systems and their implications for what is seen as ethical in response. The search for broken or underperforming components that can carry the explanatory and moral load of accidents in complex systems has deep roots in what Western society finds rational, logical and just. Throughout the past three and a half centuries, the West has equated scientific thinking with a Cartesian-Newtonian worldview which prizes decomposition, linearity, and the pursuit of complete knowledge. In this paper, I lay out how this has given rise to what I call a Newtonian ethic of failure—which makes particular assumptions about the relationship between cause and effect, foreseeability of harm, time-reversibility and the ability to come up with the "true story" of a particular event. With an eye on systems failure, I suggest below that these assumptions animate much human factors work today, including accident investigations, societal expectations, managerial mandates, technical tools and artifacts, and judicial responses to system failures (e.g. Dekker, 2007).

The Cartesian-Newtonian worldview and its implications for system failure

The logic behind Newtonian science is easy to formulate, although its implications for how we think about the ethics of failure are subtle, as well as pervasive. In this section I review aspects of the Newtonian-Cartesian worldview that influence how we understand (and consider the ethics around) failure in complex systems, even today. Until the early 20th century, classical mechanics, as formulated by Newton and further developed by Laplace and others, was seen as the foundation for science as a whole. It was expected that observations made by other sciences would sooner or later be reduced to the laws of mechanics. Although that never happened, other disciplines, including psychology and law, did adopt a reductionist, mechanistic or Newtonian methodology and worldview. This influence was so great that most people still implicitly equate "scientific thinking" with "Newtonian thinking." The

mechanistic paradigm is compelling in its simplicity, coherence and apparent completeness. It was not only successful in scientific applications, but largely consistent with intuition and common sense.

Reductionism and the eureka part

The best known principle of Newtonian science, formulated well before Newton by the philosopher-scientist Descartes, is that of analysis or reductionism. To understand any complex phenomenon, you need to take it apart, i.e. reduce it to its individual components. This is recursive: if constituent components are still complex, you need to take your analysis a step further, and look at their components. In other words, the functioning or non-functioning of the whole can be explained by the functioning or non-functioning of constituent components. Attempts to understand the failure of a complex system in terms of failures or breakages of individual components in it—whether those components are human or machine—is very common. The investigators of the Trans World Airlines 200 crash off New York called it their search for the "eureka part:" the part that would have everybody in the investigation declare that the broken component, the trigger, the original culprit, had been located and could carry the explanatory load of the loss of the entire Boeing 747. But for this crash, the so-called "eureka part" was never found (see Dekker, 2005).

Linear, componential thinking permeates the investigation of accidents and organizational failures (Holden, 2009) as well as reliability engineering methods (e.g. Dougherty, 1990; Hollnagel, 1998). The defenses-in-depth metaphor, popularized as the "Swiss Cheese Model" (Reason, 1997) and used in event classification schemes (e.g. Shappell & Wiegmann, 2001) relies on the componential, linear parsing-up of a system, so as to locate the layer or part that was broken. The analytic recursion in these methods ends up in categories such as "unsafe supervision" or "poor managerial decision making." Indeed, in technically increasingly reliable systems, the "eureka part" has become more and more the human (e.g. FAA, 1990) and there is a cottage industry of methods that try to locate and classify which errors by which people lay behind a particular problem (see Dekker, 2007). Most or all presume a linear relationship between the supposed error and the parts or processes that were broken or otherwise affected.

Our understanding of the psychological sources of failure is subject to reductive Newtonian recursion as well. In cases where the component failure of "human error" remains incomprehensible, we take "human error" apart too. Methods that subdivide human error up into further component categories, such as perceptual failure, attention failure, memory failure or inaction are now in use in air traffic control (Hollnagel & Amalberti, 2001) and similar linear, reductionist understandings of human error dominate the field of human factors (Dekker & Hollnagel, 2004). The classically mechanistic idea of psychology that forms the theoretical bedrock for such reductionist thinking of course predates human factors (e.g. Freud, 1950; Neisser, 1976; Dekker, 2005). Through analytic reduction, it sponsors an atomistic view of complex psychological phenomena; that understanding them comes from

revealing the functioning or breakdown of their constituent components (e.g. Kowalsky, 1974).

The philosophy of Newtonian science is one of simplicity: the complexity of the world is only apparent and to deal with it you need to analyze phenomena into their basic components. The way in which legal reasoning in the wake of accidents separates out one or a few actions (or inactions) on the part of individual people follows such reductive logic. For example, the Swedish Supreme Court ruled that if one nurse had more carefully double-checked a particular medication order before preparing it (mistakenly at ten times the intended dose) a three-month old baby would not have died (see Dekker, 2007). Such condensed, highly focused accounts that converge on one (in)action by one person (the "eureka part") give componential models of failure a societal legitimacy that keeps reproducing and instantiating Newtonian physics.

Causes for effects can be found

In the Newtonian vision of the world, all that happens has a definitive cause and a definitive effect. In fact, there is a symmetry between cause and effect (they are equal but opposite). The determination of the "cause" or "causes" is of course seen as the most important function of accident investigation today, and assumes that physical effects (a crashed airliner, a dead patient) can be traced back to physical causes (or a chain of causes-effects). That effects cannot occur without causes makes it into legal reasoning in the wake of accidents too. For example, "to raise a question of negligence, harm must be caused by the negligent action" (GAIN, 2004, p. 6). It is assumed that a causal relationship (that negligent action caused harm) is indeed demonstrable and provable beyond reasonable doubt.

The Newtonian ontology that holds all this up is materialistic: all phenomena, whether physical, psychological or social, can be reduced to matter, that is, to the movement of physical components inside three-dimensional Euclidean space. The only property that distinguishes particles is where they are in that space. Change, evolution, and indeed accidents, can be reduced to the geometrical arrangement (or misalignment) of fundamentally equivalent pieces of matter, whose interactive movements are governed exhaustively by linear laws of motion, of cause and effect. A visible effect (e.g. a baby dead of lidocaine poisoning) cannot occur without a cause (a nurse blending too much of the drug). The Newtonian assumption of proportionality between cause and effect can in fact make us believe that really bad effects (the dead baby) have really bad causes (a hugely negligent action by an incompetent nurse). The worse the outcome, the more "negligent" its preceding actions are thought to have been (Hugh & Dekker, 2009). In road traffic, talking on a cell phone is not considered illegal by many, until it leads to a (fatal) accident (Tingvall, 2010). It is the effect that makes the cause bad.

The Newtonian model has been so pervasive and coincident with "scientific" thinking, that if analytic reduction to determine cause-effect relationships (and their material basis) cannot be achieved, then either the method or the phenomenon is not considered worthy of the label "science." This problem of scientific self-confidence

has plagued the social sciences since their inception (Flyvbjerg, 2001), inspiring not only Durkheim to view the social order in terms of an essentialist naïve Newtonian physics, but also for example to have Freud aim "to furnish a psychology that shall be a natural science: that is, to represent psychical processes as quantitatively determinate states of specifiable material particles, thus making those processes perspicuous and free from contradiction" (1950, p. 295). Behaviorists like Watson reduced psychological functioning to mechanistic cause-effect relationships in a similar attempt to protect social science from accusations of being unscientific (Dekker, 2005).

The foreseeability of harm

According to Newton's image of the universe, the future of any part of it can be predicted with absolute certainty if its state at any time was known in all details. With enough knowledge of the initial conditions of the particles and the laws that govern their motion, all subsequent events can be foreseen. In other words, if somebody can be shown to have known (or should have known) the initial positions and velocities of the components constituting a system, as well as the forces acting on those components (which in turn are determined by the positions of these and other particles), then this person could, in principle, have predicted the further evolution of the system with complete certainty and accuracy. A system that combines the physiology of a three-month old baby with the chemical particles diethylamino-dimethylphenylacetamide that constitute lidocaine will follow such lawful evolution, where a therapeutic dose is less than 6 mg lidocaine per gram serum, and a dose almost ten times that much will kill the baby (see Dekker, 2007).

If such knowledge is in principle attainable, then the harm that may occur if particles are lined up wrongly is foreseeable too. Where people have a duty of care (like nurses and other healthcare workers do) to apply such knowledge in the prediction of the effects of their interventions, it is consistent with the Newtonian model to ask how they failed to foresee the effects. Did they not know the laws governing their part of the universe (i.e. were they incompetent, unknowledgeable)? Were they not conscientious or assiduous in plotting out the possible effects of their actions? Indeed, legal rationality in the determination of negligence follows this feature of the Newtonian model almost to the letter: "Where there is a duty to exercise care, reasonable care must be taken to avoid acts or omissions which can reasonably be foreseen to be likely to cause harm. If, as a result of a failure to act in this reasonably skillful way, harm is caused, the person whose action caused the harm, is negligent" (GAIN, 2004, p. 6).

In other words, people can be construed as negligent if the person did not avoid actions that could be foreseen to lead to effects—effects that would have been predictable and thereby avoidable if the person had sunk more effort into understanding the starting conditions and the laws governing the subsequent motions of the elements in that Newtonian sub-universe. Most road traffic legislation is founded on this Newtonian commitment to foreseeability too. For example, a road traffic law in a typical Western country might read how a motorist should adjust speed so as to be able stop the vehicle before any hazard that might be foreseeable,

and remain aware of the circumstances that could influence such selection of speed (this from the Swedish *Trafikförordning* 1996:1276, kap.3, §14 and 15, see Tingvall & Lie, 2010). Both the foreseeability of all possible hazards and the awareness of circumstances (initial conditions) as critical for determining speed are steeped in Newtonian epistemology. Both are also heavily subject to outcome bias: if an accident suggests that a hazard or particular circumstance was not foreseen, then speed was surely too high. The system's user, as a consequence, is always wrong (Tingvall & Lie, 2010).

Time-reversibility

The trajectory of a Newtonian system is not only determined towards the future, but also towards the past. Given its present state, we can in principle reverse the evolution to reconstruct any earlier state that it has gone through. The Newtonian universe, in other words, is time-reversible. Because the movement of, and the resulting interactions between, its constituent components are governed by deterministic laws of cause and effect, it does not matter what direction in time such movements and interactions are plotted. Such assumptions, for example, give accident and forensic investigators the confidence that an event sequence can be reconstructed by starting with the outcome and then tracing its causal chain back into time (Dekker, 2006). The notion of *re*construction reaffirms and instantiates Newtonian physics: our knowledge about past events is nothing original or creative or new, but merely the result of uncovering a pre-existing order. The only thing between us and a good reconstruction are the limits on the accuracy of our representation of what happened. We then assume that this accuracy can be improved by "better" methods of investigation, for example (e.g. Shappell & Wiegmann, 2001; cf Dekker, 2005).

Completeness of knowledge

The traditional belief in science is that its facts have an independent existence outside of people's minds: they are naturally occurring phenomena "out there," in the world. The more facts a scientist or analyst or investigator collects, the more it leads, inevitably, to more, or better, science: a better representation of "what happened." The belief is that people create representations or models of the "real" out there, models or representations that mimic or map this reality. Knowledge is basically that representation. When these copies, or facsimiles, do not match "reality," it is due to limitations of perception, rationality, cognitive resources, or, particularly for investigators or researchers, due to limitations to methods of observation. More refined methods and more data collection can compensate for such limitations.

Newton argued that the laws of the world are discoverable and ultimately completely knowable. God created the natural order (though kept the rulebook hidden from man) and it was the task of science to discover this hidden order underneath the apparent disorder. As mentioned before, founding sociologist Emile Durkheim took the same position for social science in the nineteenth century. Underneath a seemingly disordered, chaotic appearance of the social world, there is a social order governed by particular laws (of institutions, obligations and constraints). Newtonian

epistemology is based on the reflection-correspondence view of knowledge: our knowledge is an (imperfect) mirror of the particular arrangements of matter outside of us (Heylighen, 1989). The task of investigations, or science, is to make the mapping (or correspondence) between the external, material objects and the internal, cognitive representation (e.g. language) as accurate as possible. The starting point is observation, where information about external phenomena is collected and registered (e.g. the gathering of "facts" in an accident investigation), and then gradually completing the internal picture that is taking shape. Ultimately, this can lead to a perfect, objective representation of the world outside (Heylighen et al., 2005).

The consequence for the ethics of failure is that there can be only one true story of what happened. In Newtonian epistemology, the "true" story is the one in which there is no more gap between external events and their internal representation. Those equipped with better methods, and particularly those who enjoy greater "objectivity," (i.e. those who, without any bias that distorts their perception of the world, will consider *all* the facts) are better poised to achieve such a true story. Formal, government-sponsored accident investigations enjoy this aura of objectivity and truth—if not in the substance of the story they produce, then at least in the institutional arrangements surrounding its production. First, putative objectivity is deliberately engineered into the investigation as a sum of subjectivities. All interested parties (e.g. vendors, the industry, the operator, the legal system, unions, and professional associations) can officially contribute (though their voices are easily silenced or sidelined). Second, those other parties often wait until a formal report is produced before publicly taking either position or action, legitimating the accident investigation as arbitrator between fact and fiction, between truth and lie. It supplies the story of "what really happened." Without first getting that "true" story, no other party can credibly move forward.

Newtonian bastardization of the response to failure in complex systems

Together, taken-for-granted assumptions about decomposition, cause-effect symmetry, foreseeability of harm, time reversibility, and completeness of knowledge give rise to a Newtonian ethic in the wake of failure. It can be summed up as follows:

- To understand a failure of a system, we need to search for the failure or malfunctioning of one or more of its components. The relationship between component behaviour and system behaviour is analytically non-problematic.
- Causes for effects can always be found, because there are no effects without causes. In fact, the larger the effect, the larger (e.g. the more egregious) the cause must have been.
- If they put in more effort, people can better foresee outcomes. After all, they would better understand the starting conditions and are already supposed to know the laws by which the system behaves (otherwise they wouldn't be allowed to work in it). With those two in hand, all future system states can be predicted. If they can be predicted, then harmful states can be foreseen and should be avoided.

- An event sequence can be reconstructed by starting with the outcome and tracing its causal chain back into time. Knowledge thus produced about past events is the result of uncovering a pre-existing order.
- There can be only one true story of what happened. Not just because there is only one pre-existing order to be discovered, but also because knowledge (or the story) is the mental representation or mirror of that order. The *truest* story is the one in which the gap between external events and internal representation is the smallest. The *true* story is the one in which there is no gap.

Like in many other fields of inquiry, these assumptions remain largely transparent and closed to inquiry in safety work precisely because they are so self-evident and common-sensical. The way they get retained and reproduced is perhaps akin to what Althusser (1984) called "interpellation." If people involved in safety work are expected to explain themselves in terms of the dominant assumptions; they will make sense of events using those assumptions; they will then reproduce the existing order in their words and actions. Organizational, institutional and technological arrangements surrounding their work don't leave plausible alternatives (in fact, they implicitly silence them). For instance, investigators are mandated to find the probable cause(s) and turn out enumerations of broken components as their findings. Technological-analytical support (incident databases, error analysis tools) emphasizes linear arrangements and the identification of malfunctioning components. Also, organizations and those held accountable for failure inside of them, need something to "fix," which further valorizes condensed accounts, focuses on localization of a few single problems and re-affirms a pre-occupation with components. If these processes fail to satisfy societal accountability requirements, then courts deem certain practitioners criminal, lifting uniquely bad components out of the system (Dekker, 2007).

Newtonian hegemony, then, is maintained not by imposition but by interpellation, by the confluences of shared relationships, shared discourses, institutions, and knowledge. Foucault called the practices that produce knowledge and keep knowledge in circulation an epistemé: a set of rules and conceptual tools for what counts as factual. Such practices are exclusionary. They function in part to establish distinctions between those statements that will be considered true and those that will be considered false. Or factual rather than speculative. Or just rather than unjust (Foucault, 1980). The true statement will be circulated through society, reproduced in accident reports, for example, and in books and lectures about accidents. These true statements will underpin what is taken to be common-sensical knowledge in a society—i.e. the Newtonian physical order. The false statement will quickly fade from view as it not only contradicts common sense, but also because it is not authorized by people legitimated and trusted by society to furnish the truth.

A naïve socio-technical Newtonian physics is thus continuously read into events that could yield much more complexly patterned interpretations. Newtonian assumptions not only support but also reproduce their own ideas, values, sentiments, images, and symbols about failure, its origin and its appropriate ethical consequences. Collectively, they assert that action in the world can be described as a set of casual

laws, with time reversibility, symmetry between cause and effect, and a preservation of the total amount of energy in the system. The only limiting points of such an analysis are met when laws are not sufficiently rigorous or exhaustive, but this merely represents a problem of further methodological refinement in the pursuit of greater epistemological rigor and more empirical data produced because of it. This Newtonian commitment all but excludes even a common awareness of alternatives. It is not that more complex readings would be "truer" in the sense of corresponding more closely to some objective state of affairs. But they could hold greater or at least a different potential for safety improvement, and could help people reconsider what is ethical in the aftermath of failure.

The complexity and systems worldview and its implications for system failure

There is something really important that analytic reduction cannot tell us, and that is how a number of different things and processes act together when exposed to a number of different influences at the same time. Discontent with the Newtonian worldview has been brewing since at least Pioncaré in the 19th century and exploded into fuller view over the last forty years in Western science. General Systems Theory (von Bertalanffy, 1973) helped establish a serious scientific foundation for the alternative which collectively became known as complexity- or systems theory. Von Bertalanffy recognized how most systems of interest are not closed (like the planetary system which was the basic model for Newton's ideas) but open—interactive with and dependent on an environment much larger and complex than the system itself. This goes for all living systems, including socio-technical organizations. Cybernetics, an approach closely related to systems theory (Ashby, 1964) demonstrated the simplest of that principle: using feedback from the environment, systems actively compensate perturbations in order to maintain their equilibrium.

Systems science aims to capture the patterns of relationships, of interactions, of organization, at a level of description independent of the domain specifics (a unifying language, in other words, like Newtonian science offered). The central paradigm of system science is the multi-agent system. Agents are intrinsically subjective and uncertain about their environment and future, but global organization emerges out of their local interactions with each other and the environment (Heylighen et al., 2005). Systems thinking moves traditional (social) science to a much more contested postmodernist critical theory. Agency, or human acts, occurs not along some pre-determined linear order, but within a vast and complex and non-linear network of relationships, processes, and systems that are as ecological as they are cultural (Cronon, 1992). What each of these agents knows, or can know, has little or nothing to do with some objective state of affairs, but is produced locally as a result of those interactions. Below is an outline of the implications for (the ethics of) system failure, mirroring the topics used in the section on Newtonian science.

Synthesis and holism

The Newtonian focus on parts as the cause of accidents has sponsored a belief that redundancy is the best way to protect against hazards. Safety-critical systems usually

have multiple redundant mechanisms, safety systems, elaborate policies and procedures, command and reporting structures as well as professional specialization—all to protect or warn them against the effects of component failures or malfunctions. The downside is that barriers, as well as professional specialization, policies, procedures, protocols, redundant mechanisms and structures, all add to a system's complexity. With the introduction of each new part or layer of defense, there is an explosion of new relationships (*between* parts and layers and components) that spreads out through the system. Increasing complexity, particularly interactive complexity, has thus given rise to a new kind of accident: the system accident (Perrow, 1984). This type of accident results from the relationships between components, not from the workings or dysfunctioning of any component part.

This insight grew out of what became known as systems engineering, pioneered in part by 1950's aerospace engineers who were confronted with increasing complexity (i.e. the interconnectivity and interactivity between system components) in aircraft and ballistic missile systems at that time. Greater complexity led to more possible interactions than could be planned, understood, anticipated or guarded against. The enormous increase in the use of software has greatly added to this interactive complexity. With software, the possible states that a system can end up in become mind-boggling and entirely opaque to any meaningful prediction (Leveson, 2006). And with software, no part has to be broken—in fact, nothing *can* logically "break." It behaves as programmed (though not necessarily always as foreseen). As a result, system accidents involve the unanticipated interaction of a multitude of events in a complex system—events and interactions, often very normal, whose combinatorial explosion can quickly outwit people's best efforts at predicting and mitigating trouble.

Newtonian analytic reduction of complex systems not only fails to reveal any culprit part (as broken parts are not necessary to produce system failures) but would also eliminate the phenomenon of interest—the interactive complexity of the system itself. The traditional view is that organizations are Newtonian-Cartesian machines with components and linkages between them. Accidents get modelled as a sequence of events (actions-reactions) between a trigger and an outcome. But such models can say nothing about the build-up of latent failures, about a gradual, incremental loosening or loss of control that seems to characterize system accidents (Dekker, 2005; Leveson, 2006). The processes of erosion of constraints, of attrition of safety, of drift towards margins, cannot be captured because reductive approaches are static metaphors for resulting forms, not dynamic models oriented at processes of the formation, the evolution of relationships.

Emergence

Since ideas about systemic accident models were first published and popularized, system safety has been characterized as an emergent property, something that cannot be predicted on the basis of the components that make up the system. Accidents have also been characterized as emergent properties of complex systems (Hollnagel, 2004). They cannot be predicted on the basis of the constituent parts; rather, they are one emergent result of the constituent components doing their (normal) work. A

systems accident is possible in an organization where people themselves suffer no noteworthy incidents, in which everything looks normal, and everybody is abiding by their local rules, common solutions, or habits.

Emergence means that the behaviour of the whole cannot be explained by, and is not mirrored in, the behaviour of constituent components. No part needs to be broken for the system to break. Instead, the behaviour of the whole is the result—the emergent, cumulative result—of all the local components following their local rules and interacting with each other in numerous ways, cross-adapting to each others' behaviour as they do so. Snook (2000) expresses the realization that bad effects can happen with no clear causes in his study of the shoot-down of two U.S. Black Hawk helicopters by two U.S. fighter jets in the no-fly zone over Northern Iraq in 1993:

> "This journey played with my emotions. When I first examined the data, I went in puzzled, angry, and disappointed—puzzled how two highly trained Air Force pilots could make such a deadly mistake; angry at how an entire crew of AWACS controllers could sit by and watch a tragedy develop without taking action; and disappointed at how dysfunctional Task Force OPC must have been to have not better integrated helicopters into its air operations. Each time I went in hot and suspicious. Each time I came out sympathetic and unnerved... If no one did anything wrong; if there were no unexplainable surprises at any level of analysis; if nothing was abnormal from a behavioural and organizational perspective; then what…?"

Snook's impulse to hunt down the broken components (deadly pilot error, controllers sitting by, a dysfunctional Task Force) is tempered by the lack of results. In the end he comes out "unnerved," because there is no "cause" that preceded the effect. The most plausible story of the incident defies Newtonian logic. It would be here that a Newtonian narrative of failure achieves its end only by erasing its true subject: human agency and the way it is configured in a hugely complex network of relationships and interdependencies. The Newtonian identification of the broken part becomes plausible only by obscuring all those interdependencies, only by isolating, mechanizing, or de-humanizing human agency.

In complexity and system thinking, not only is there no clear line from cause to effect, there is also no longer any symmetry between them as in a Newtonian system. In a complex system, an infinitesimal change in starting conditions can lead to huge differences later on (indeed—having an accident or not having one). This *sensitive dependence on initial conditions*, or butterfly effect, removes both linearity and proportionality from the relationship between system input and system output. This asymmetry between "cause" and "effect" (the Newtonian terms are used even here for lack of alternatives) has implications for the ethical load distribution in the aftermath of complex system failure. Consequences cannot form the basis for an assessment of the gravity of the cause. Trivial, everyday organizational decisions, embedded in masses of similar decisions and only subject to special consideration with the wisdom of hindsight, cannot be meaningfully singled out for purposes of exacting accountability (e.g. through criminalization) because their relationship to

the eventual outcome is complex, non-linear, and was probably impossible to foresee.

Foreseeability of probabilities, not certainties

In a complex system, the future is uncertain. Knowledge of initial conditions is not enough, and total knowledge of the laws governing the system is unattainable. This is true for the non-linear dynamics of traditional physical systems (e.g. the weather) but perhaps even more so for social systems. Social systems composed of individual agents and their many cross-relationships, after all, are capable of internal adaptation as a result of their experiences with each other and with the system's dynamic environment. This can make the possible landscape of outcomes even richer and more complexly patterned. As a result, a complex system only allows us to speculate about probabilities, not certainties. This changes the ethical implications of decisions, as their eventual outcomes cannot be foreseen. Decision makers in complex systems are capable only of assessing the probabilities of particular outcomes, something that remains ever shrouded in the vagaries of risk assessment *before*, and always muddled by outcome and hindsight biases *after* some visible system output. With an outcome in hand, its (presumed) foreseeability suddenly becomes quite obvious, and it may appear as if a decision in fact *determined* an outcome; that it inevitably and clearly led up to it (Fischhoff, 1975).

The evaluation of damage caused by debris falling off the external tank prior to the 2003 Space Shuttle Columbia flight can serve as an example. Always under pressure to accommodate tight launch schedules and budget cuts (in part because of a diversion of funds to the international space station) it became more sensible to see certain problems as maintenance issues rather than flight safety risks. Maintenance issues could be cleared through a nominally simpler bureaucratic process, which allowed quicker Shuttle vehicle turnarounds. In the mass of assessments to be made between flights, the effect of foam debris strikes was one. Gradually converting this issue from safety to maintenance was not different from a lot of other risk assessments and decisions that NASA had to do as one Shuttle landed and the next was prepared for flight—one more decision, just like tens of thousands of other decisions. While any such decision can be quite rational given the local circumstances and the goals, knowledge and attention of the decision makers, interactive complexity of the system can take it onto unpredictable pathways to hard-to-foresee system outcomes.

That does not mean that such decisions are not singled out in retrospective analyses. That they are is but one consequence of Newtonian thinking: accidents have typically been modeled as a chain of events. While a particular historical decision can be cast as an "event," it becomes very difficult to locate the immediately preceding "event" that was *its* cause. So the decision (the human error, or "violation") is cast as the aboriginal cause, the root cause (Leveson, 2006).

Time-irreversibility

The conditions of a complex system are irreversible. This stands in contrast to a linear Newtonian system where the relationships between causes and consequences can be traced out either forward or backward in time without analytic difficulty. This cannot work in a complex system because, for one, it is never static. Complex systems continually experience change as relationships and connections evolve internally and adapt to their changing environment. As discussed above, complexity also means that there are no straightforward relationships between causes and effects. Rather, results emerge in ways that cannot be traced back to the behaviour of constituent components. The precise set of conditions that gave rise to this emergence is something that can never be exhaustively reconstructed. This means that the any predictive power of retrospective analysis of failure is severely limited (Leveson, 2006). Decisions in organizations, for example, to the extent that they can be excised and described separate from context at all, are not single beads strung along some linear cause-effect sequence. Complexity argues that they are spawned and suspended in the messy interior of organizational life that influences and buffets and shapes them in a multitude of ways. Many of these ways are hard to trace retrospectively as they do not follow documented organizational protocol but rather depend on unwritten routines, implicit expectations, professional judgments and subtle oral influences on what people deem rational or doable in any given situation (Vaughan, 1996; Rochlin, 1999; Snook, 2000).

*Re*constructing events in a complex system, then, is nonsensical: the system's characteristics make it impossible. Our own psychological characteristics make it so too. As soon as an outcome has happened, whatever past events can be said to have led up to it, undergo a whole range of transformations (Fischhoff & Beyth, 1975; Dekker & Hugh, 2009). Take once again the idea that it is a sequence of events that precedes a bad outcome (e.g. an accident). Who makes the selection of the "events" and on the basis of what? The very act of separating important or contributory events from unimportant ones is an act of construction, of the creation of a story, not the reconstruction of a story that was already there, ready to be uncovered. Any sequence of events or list of contributory or causal factors already smuggles a whole array of selection mechanisms and criteria into the supposed "re"-construction. There is no objective way of doing this—all these choices are affected, more or less explicitly, by the analyst's background, preferences, experiences, biases, beliefs and purposes. "Events" are themselves defined and delimited by the stories with which the analyst configures them, and are impossible to imagine outside this selective, exclusionary, narrative fore-structure (Cronon, 1992).

Perpetual incompleteness and uncertainty of knowledge

A central pre-occupation of critical perspectives on science and scientific rationality (as offered by Nietzsche, Weber, Heidegger, Habermas, to name a few) for at least a century-and-a-half is that our (scientific) knowledge is not an objective picture of the world as it is. The idea of an objective representation of a pre-existing order is, indeed, the Newtonian position. In it, the job of a scientist (or of the operators we study, for that matter) is to create representations or constructs that mimic or map the

world—their "knowledge." When these copies, or facsimiles, do not match "reality," it is due to limitations of an operator's perception, rationality, or cognitive resources (e.g. mental workload), or, if we are researchers, due to limitations to our methods of observation (Flach, Dekker & Stappers, 2008). More data (and more lines of evidence, cleverer experiments) mean better science: better copies, better facsimiles.

But do the constructs we use *reflect* what we see, or do they *create* what we see? Does knowledge reflect a state of the world, or does it shape, or construct, the world in ways that both facilitates and constrains action (Healy, 2003)? Take memory as an example. Constructs of what memory is and how it works have historically been inscribed on metaphors derived from contemporary technology, ranging from wax tablets and books to photography, radios and computers and even holograms (Draaisma, 2000; see also Leary, 1995). Similarly, the information processing metaphor, dominant in human factors in the 1970's (Neisser, 1976; Hollnagel & Woods, 1983) turned attention into an issue of capacity, and memory into one of storage and retrieval. As technological developments (wax tablets, transistors, computers) influence our models and language, we change the way we think about and therefore study for example memory or attention.

Which constructs, or words, we use to denote certain observations, then, seems infinitely renegotiable. There is what Quine called "an indeterminacy of reference," an "enormous vagueness concerning the referents of our words" (Gergen, 1999, p. 21), and the other way around: a huge flexibility and continuous evolution in the words we use for our referents. We change what we accept as "evidence" and change what constructs we find interesting to publish and read about. This would suggest that words are not pictures of the world. Words are choices, consensual agreements, for how to see the world. A technical vocabulary of constructs creates a particular empirical world which would not even exist without those words. The position taken by most social sciences over the last forty years is that it is consensus that cements word to world (Gergen, 1999). This observation robs those words (or constructs) of any inherent claim to truthfulness in the sense of describing the world as it is.

Still, there is a stubborn belief that researchers have a possibility to access and represent the objective state of the world, and that they can thus *really* know what is going on there. Take situation awareness, for example. Because of the way in which knowledge about situation awareness is gathered in typical human factors experiments, "there is a "ground truth" against which its accuracy can be assessed (e.g., the objective state of the world or the objective unfolding of events that are predicted)" (Parasuraman et al., 2008, p. 144). In other words, an operator's understanding of the world can be contrasted against (and found more or less deficient relative to) an objectively available state of that world. This occurs in error assessment and error counting work too (observers define what is "erroneous," otherwise they could not observe or tabulate errors).

The Newtonian belief that is both instantiated and reproduced is that there is a world that is objectively available and apprehensible. This epistemological stance represents a kind of a-perspectival objectivity. It assumes that we, as researchers, are able to take a "view from nowhere" (Nagel, 1989), a value-free, background-free,

position-free view that is true (the "ground truth"). This re-affirms the classical or Newtonian view of nature (an independent world exists to which we as researchers, with proper methods, can have objective access). It rests on the belief that observer and the observed are entirely separable. Knowledge is nothing more than a mapping from object to subject. Scientific discovery is not a creative process: it is merely an "uncovering" of distinctions that were already there and simply waiting to be observed (Heylighen, 1989). Most social science (and for the last hundred years, natural science too) does not believe this. The inseparability of observer and observed has been a consistent theme in some of the most acute analyses of both the social and natural sciences (e.g. Wallerstein et al., 1996). The observer is not just the contributor to, but in many cases the creator of, the observed. If there is an objective world, then we couldn't know it.

Heylighen (1989) explains how cybernetics introduced this idea to complexity and systems thinking early on. For the rules of cybernetics, knowledge is intrinsically subjective; it is an imperfect tool used by an intelligent agent to help it achieve its personal goals. Not only does the agent not need an objective reflection of reality, it can actually never achieve one. Indeed, the agent does not have access to any "external reality": it can merely sense its inputs, note its outputs (actions) and from the correlations between them induce certain rules or regularities that seem to hold within its environment. Different agents, experiencing different inputs and outputs, will in general induce different correlations, and therefore develop a different knowledge of the environment in which they live. There is no objective way to determine whose view is right and whose is wrong, since the agents effectively live in different environments (Heylighen, 1989; Heylighen et al., 2005)

This means that claiming a position of "objectivity" introduces a sort of normativism or prescriptivism. Researchers can claim, on the basis of their knowledge, that they can adjudicate their subjects' "truth." The researchers are right, and their subjects are wrong, or only partially right. This may be unproblematic within a research paradigm that does not consider its ethical implications a whole lot, but important aspects do get lost in such normativism. Critical studies of for example error counting (Hollnagel & Amalberti, 2001) showed how an error count could achieve its end only by erasing its true subject: adaptation and expertise expressed by practitioners. Post-count interviews revealed how actions or "omissions" previously rated as errors were explained by practitioners as for example anticipation of workload fluctuations. If we adjudicate an operator's understanding of an unfolding situation against our own "ground truth," we may learn little of value about why people saw what they did, and why taking or not taking action made sense to them. What is an error to one is perfectly rational to somebody else (particularly to the one actually doing the work). This should give some pause for thought about what is ethical to do in the aftermath of such an "error." Who said some action or inaction was an "error?" Who decided that the "error" belonged to a sequence of "events" that led up to the outcome? Imposing one normative view onto everybody else could easily be seen as unethical, and unjust (Dekker, 2007).

A post-Newtonian ethic for failure in complex systems

Complexity and systems thinking denies us the comfort of one objectively accessible reality that can, as long as we have accurate methods, arbitrate between what is true and what is false. This has far-reaching implications for what we can consider ethical in the aftermath of failure (see Cilliers, 1998):

- There is never one true story of what happened. In fact, we should aim for diversity, and respect otherness and difference as values in themselves. That people have different accounts of what happened in the aftermath of failure should not be seen as somebody being right and somebody being wrong, or as somebody wanting to dodge or fudge the "truth." In fact, if somebody claims to have the true story of what happened, it turns everybody else into a liar. Diversity of narrative can be seen as an enormous source of resilience in complex systems (it is when it comes to biodiversity, after all), not as a weakness. The more angles, the more there can be to learn.
- Gather as much information on the issue as possible, notwithstanding the fact that it is impossible to gather "all" the information.
- Consider as many of the possible consequences of any judgment in the aftermath of failure, notwithstanding the fact that it is impossible to consider all the consequences. This impossibility does not discharge people of their ethical responsibility to try, particularly not if they are in a position of power; where decisions get made and sustained about the fate of people involved, or about the final word on a story of failure.
- Make sure that it is possible to revise any judgment in the wake of failure as soon as it becomes clear that it has flaws, notwithstanding the fact that the conditions of a complex system are irreversible. Even when a judgment is reversed, some of its consequences (psychological, practical) will probably remain irreversible.

Conclusion

A post-Newtonian ethic of failure could emerge from a human factors that truly embraces systems thinking. In such an ethic, there is no longer an obvious relationship between the behaviour of parts in the system (or their dysfunctioning, e.g. "human errors") and system-level outcomes. Instead, system-level behaviours emerge from the multitude of relationship and interconnections deeper inside the system, but cannot be reduced to those relationships or interconnections. In such an ethic, human factors stops looking for the "causes" of failure or success. System-level outcomes have no clearly traceable "causes" as their relationships to "effects" are neither simple nor linear. In fact, the selection of "causes" (or "events" or "contributory factors") is always an act of construction. It is the creation of a story, not the reconstruction of a story that was already there, ready to be uncovered.

There is no objective way of doing this—all analytical choices are affected, more or less explicitly, by the human factors researchers' background, preferences, language, experiences, biases, beliefs and purposes. The categories by which human factors parses its world ("events, causes, errors, memory, awareness, attention") are

themselves defined and delimited by the stories within which the researcher configures them, and are impossible to imagine outside this exclusionary, narrative fore-structure. This also means that there can never be one true story of what happened. Truth, if there is such a concept, lies in diversity, not singularity. The pursuit of "truth" in human factors, then—or accident investigation or any other related endeavour—is not the pursuit of a single narrative or answer. In fact, achieving "truth" and creating meaningful, or constructive ways of improving systems is helped much more by engaging a diversity of narratives and perspectives. Safe systems, according to High Reliability Theory (HRT) are "complexly sensitized:" they remain attuned to a multitude of channels about the operation of their system and its likelihood of continued safety. They take minority opinion seriously, celebrate dissent, and welcome bad news. It is that kind of diversity that might have provided a different outcome in cases where dissent was instead squelched, where bad news was converted or rationalized, and where one grand narrative of a basically safe system became the only legitimate and dominant story that the system could tell about itself (Vaughan, 1996; Rochlin, 1999).

As human factors tentatively explores a system view of performance and, ultimately, its ethical consequences, the differences between an original Newtonian perspective on failure and a systems perspective should become ever clearer. Rather than seeking to affirm one interpretation (e.g. one story of what happened), human factors could start celebrating multiple dissenting, smaller narratives (including those of lay communities) that can place things in a new language. Of course, while liberating, this could be potentially unsettling. If human factors relinquishes its Newtonian vision of the world and the research that it performs inside of it, then nothing can be taken as merely, obviously, objectively or unconstructedly "true" any longer.

The systems view has no answer to who is accountable for failure and success in complex systems, just as the Newtonian view only has fraudulently oversimplified, extremely limited, and probably unjust answers. What the systems view allows us to do, however, is dispense with the notion that there *are* easy answers, supposedly within reach of the one with the best method or most objective viewpoint. It allows us to invite more voices into the conversation, and to celebrate their diversity and contribution. That, if anything, should spell a rich future for any scientific field of inquiry.

References

Althusser, L. (1984). *Essays on ideology*. London: Verso.

Ashby, W.R. (1964). *An Introduction to Cybernetics*. London: Methuen.

Catino, M. (2008). A review of literature: Individual blame vs. organizational function logics in accident analysis. *Journal of contingencies and crisis management, 16*, 53-62.

Cilliers, P. (1998). *Complexity and postmodernism: Understanding complex systems*. London: Routledge.

Cronon, W. (1992). A place for stories: Nature, history and narrative. *The Journal of American History, 78*, 1347-1376.

Dekker, S.W.A. (2005). *Ten questions about human error: A new view of human factors and system safety*. Mahwah, NJ: Lawrence Erlbaum Associates.

Dekker, S.W.A. (2006). *The field guide to understanding human error*. Aldershot, UK: Ashgate Publishing Co.

Dekker, S.W.A. (2007). *Just culture: Balancing accountability and safety*. Aldershot, UK: Ashgate Publishing Co.

Dekker, S.W.A., & Hollnagel, E. (2004). Human factors and folk models. *Cognition, Technology & Work, 6*, 79-86.

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

Draaisma, D. (2000). Metaphors of memory: A history of ideas about the mind. Cambridge, UK: Cambridge University Press.

Federal Aviation Administration (1990). *Profile of operational errors in the national airspace system*: *Calendar Year 1988*. Washington, DC.

Feyerabend, P. (1993). *Against method* (3rd ed.). London: Verso.

Fischhoff, B. (1975). Hindsight is not foresight: The effect of outcome knowledge on judgment under uncertainty. *Journal of Experimental Psychology: Human Perception and Performance, 104*, 288-299.

Fischhoff, B., & Beyth, R. (1975). "I knew it would happen": Remembered probabilities of once–future things. *Organizational Behavior and Human Performance, 13*, 1-16.

Flyvbjerg, B. (2001). *Making social science matter: Why social inquiry fails and how it can succeed again*. Cambridge, UK: Cambridge University Press.

Foucault, M. (1980). Truth and power. In C. Gordon (ed.), *Power/Knowledge* (pp. 80-105). Brighton: Harvester.

Freud, S. (1950). Project for a scientific psychology. In The standard edition of the complete psychological works of Sigmund Freud, Vol I. London: Hogarth Press.

GAIN (2004). *Roadmap to a just culture: Enhancing the safety environment*. Global Aviation Information Network (Group E: Flight Ops/ATC Ops Safety Information Sharing Working Group).

Galison, P. (2000). An accident of history. In P. Galison and A. Roland (Eds.), *Atmospheric flight in the twentieth century* (pp. 3-44). Dordrecht, the Netherlands: Kluwer Academic.

Gergen, K. (1999). *An invitation to social construction*. London: Sage Publications.

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.

Heylighen, F. (1989). Causality as distinction conservation: A theory of predictability, reversibility and time order. *Cybernetics and Systems*, 20, 361-384.

Heylighen, F., Cilliers, P., & Gershenson, C. (2005). *Complexity and philosophy*. Vrije Universiteit Brussel: Evolution, Complexity and Cognition.

Hollnagel, E. & Amalberti, R. (2001). The emperor's new clothes: Or whatever happened to "human error"? In S.W.A. Dekker (Ed.), *Proceedings of the 4th International Workshop on Human Error, Safety and Systems Development* (pp. 1-18). Linköping Sweden: Linköping University.

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

Hollnagel, E. (2004). *Barriers and accident prevention*. Aldershot, UK: Ashgate Publishing Co.

Hollnagel, E., & Woods, D.D. (1983). Cognitive systems engineering: New wine in new bottles. *International Journal of Man-machine studies, 18*, 583-600.

Hugh, T.B., & Dekker, S.W.A. (2009). Hindsight bias and outcome bias in the social construction of medical negligence: A review. *Journal of Law and Medicine, 16*, 846-857.

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.

Leary, D.E. (Ed.) (1995). *Metaphors in the history of psychology*. Cambridge, UK: Cambridge University Press.

Leveson, N.G. (2006). *System Safety Engineering: Back to the future*. Cambridge, MA: Aeronautics and Astronautics, Massachusetts Institute of Technology.

Neisser, U. (1976). *Cognition and reality: Principles and implications of cognitive psychology*. San Francisco: Freeman Press.

Parasuraman, R., Sheridan, T.B., & Wickens, C.D. (2008). Situation awareness, mental workload, and trust in automation: Viable, empirically supported cognitive engineering constructs. *Journal of Cognitive Engineering and Decision Making, 2*, 140-160.

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

Rasmussen, J. (1997). Risk management in a dynamic society: A modeling problem. *Safety Science, 27*, 183-213.

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

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

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

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

Tingvall, C. & Lie, A. (2010). The concept of responsibility in road traffic (Ansvarsbegreppet i vägtrafiken). Paper presented at *Transportforum*, Linköping, Sweden, 13-14 January.

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

Vaughan, D. (1999). The dark side of organizations. Mistake, misconduct, and disaster. *Annual Review of Sociology, 25*, 271-305.