

**SAFETY, RELIABILITY, STEWARDSHIP, AND REGRET:
CONTRIBUTIONS TO DEPENDABLE SYSTEM DESIGN FROM
THE STUDY OF HIGHLY RELIABLE ORGANIZATIONS**

**ANDREW KOEHLER, PHD
STATISTICAL SCIENCES, D-1
LOS ALAMOS NATIONAL LABORATORY**

12/16/2005

ABSTRACT:

Part of the Dependable System Design and Engineering project is the idea that a technology's operational experience is determined both by the nature of the physical artifacts comprising the technical system and the organizational environment in which those artifacts function. For example, completion of national policy objectives by the Apollo effort resulted from causal factors beyond those explained by the collection of parts making up the Saturn V system. Rather, system outcome was caused by a complex, highly interconnected set of processes and relationships between social, organizational, and technical components. Through a long, dynamic, development process, rocket designers, mission controllers, flight crews, and a multitude of others had to fabricate an organizational form capable of managing a complex and unforgiving enterprise. With perfect safety of space flight a known impossibility, effective attention at all stages of program life to operational vigilance, to learning from failure, to controlling such risks as could be managed, were consciously and seriously undertaken parts of the organizational "design," woven together with technical system engineering activities.

The objective of the High Reliability Organization (HRO) project, started at UC Berkeley in the mid-1980s and since taken up elsewhere, has been to study how organizations charged with performing activities characterized by high hazard and demanding technical features either manage or fail to meet operational challenges. The HRO project started with the identification of an anomaly: despite general agreement with organizational theory literature seeking to explain the causes of system accidents, researchers were puzzled by the relative success of some organizations at performing better than would be expected by this literature. For example, given the extreme hazards in operating aircraft off a carrier, how is it that the US Navy has successfully created, maintained, and improved this activity? From the standpoint of contemporary organizational theory, any sane bottom-up statistical model, or experience gleaned from failed foreign efforts to create carrier aviation capabilities, landing weapon-laden aircraft onto a rolling, pitching deck in the middle of the night should be unacceptably dangerous. Yet, the US Navy has managed, through a variety of interrelated technical and organizational strategies, to "work in practice, but not in theory," reducing the rate of major (class A) accidents from approximately 50 per 100,000 flight hours in 1950 to less than 2 in 2003.

Initially, the HRO project focused on characterizing the form this anomaly took on, in the hope of posing a challenge to existing organizational theory approaches to explaining technology related performance. However, as study of activities such as operation of nuclear/conventional power plants, air traffic control, and aircraft carriers, broadened out to include waste management, spacecraft, and other types of activities, a body of general regularities characterizing HRO behaviors emerged. While social "design" is far more difficult to prescribe than engineering design, these organizational regularities do provide a means by which the form of technology employed as part of the operation of a hazardous/complex socio-technical system can be linked to the necessary kinds of organizational structures that must be created to explain a particular level of performance.

This paper explores the different threads in the HRO literature that are of particular interest in thinking about and designing systems with dependable operational characteristics. Drawing upon this literature, I argue that the opportunity to become seriously engaged with the system design and engineering community in the creation of dependable systems offers the possibility for the development of new forms of socio-technical system design and management tools. These tools do not yet exist; however in thinking about HRO in the context of the system design task, the outline of these tools may now become visible.

HRO Performance in the Context of Dependable Technical Systems Design

The Highly Reliable Organization (HRO) project, started at UC Berkeley in 1986 offers a body of research useful to the dependable system designer; the degree of utility however, depends upon a clear understanding of the project's scope, what capabilities exist in the social sciences to predict institutional behavior, and a willingness to engage in cross disciplinary dialog.

This essay, written by a social scientist who has “gone native” and is engaged in primarily operations research based system prediction efforts, is an attempt to frame HRO research from the perspective of the system designer interested in understanding the role social science research can (and I think should) play in technical system operations. This frame is one that is not identical to what the HRO project is primarily about or how researchers involved have framed their work themselves. The collection of work produced by scholars involved in the HRO project is extraordinarily rich, diverse and robust; this reframing is an attempt to accentuate its relevance to system designers—hopefully without doing harm to the research in its own terms. I hope to provide engineers interested in dependability with a basic sense for what the HRO literature is about and a sense for where they can go next for more information.

Consider a hypothetical system design problem; perhaps for a new manned spacecraft intended for exploration of Mars. From the standpoint of various publics, policy-makers, and other actors outside of the system design/operations community, there are a set of potential costs and benefits that follow from the pursuit of this activity. On the benefit side, there are factors such as national prestige, pride, science, and perhaps subsidiary benefits such as jobs, technology development, and the growth of complementary capabilities. On the cost side of the ledger, there are factors such as the monetary expense of the program, potential hazards/deaths of both crew and bystanders, environmental pollution from toxic propellant or radioactive elements, or possible future negative externalities such as the emergence of national competition for resources in space, etc.

Whether or not sufficient social support will exist for the project throughout the system's life cycle is the result of an ongoing evaluation process that can be put in terms of a benefit-cost framework. If perceived benefits of the system outweigh the perceived combination of hazards (the intrinsic capacity for harm) and the probability of occurrence of these hazards posed by the system's functioning, the project is likely to be allowed (or provided resources) to continue.

This understanding, or framing, of the creation and functioning of systems as part of a socio-technical bargain is essential to the understanding of why system dependability is important; if no-one cares about the cost of failure, about Astronaut deaths, or whether

the spacecraft is operated successfully there is little reason to be concerned about dependability. There are many systems where dependability is not a major concern; for example most home electronics fail gracefully, without warning, and without major consequences. Still other systems such as the automobile are permitted to endure despite very low dependability and high risk because social benefits are perceived as far outweighing harm.

Indeed, only systems involving organizations which, due to the pressure of maintaining the social-technology bargain, must, “commit to using very powerful, costly technical systems that are inherently dangerous calling for high hazardous, low risk performance as a condition of delivering their benefits” will exhibit a high degree of concern for dependable system features and will be willing to pay the material and organizational costs necessary to obtain dependability (La Porte 1996, p.60). These are the same systems about which HRO study has focused, not from the standpoint of dependability engineering, but rather from the standpoint of organizational context and the management of the central benefit-cost trade-off between society and technologies demanding of operational vigilance and precise operation.

Accordingly, the relationship between dependable design efforts and results from the HRO project can be thought of as being intertwined but separable—dependable system design and HRO research are both focused on understanding how similar kinds of technical activity either succeed or fail within the confines of an operational and social contract. In the case of dependable system design, this is phrased largely in terms of an engineering framework: how can systems be made more manageable, reliable, or effective through improved internal intelligence, graceful failure modes, or better ability for operators to diagnose operational states? (Noor 2004) The unspoken assumption behind this call for dependability is that performance of the system matters; that failure of the system is likely to result in harm or costs of the sort caused by failure of a spacecraft, air traffic control system, power plant, or nuclear weapon. By thinking about the kinds of technical undependability inherent to these types of “reliability challenging systems,” dependable system designers hope to either decrease the hazards/risks of failure, improve operational benefits through efficiency gains, or at very least limit the harm done by surprise outcomes.

HRO has approached this management of technical costs and benefits primarily from a framework of organizational structure and the implied demands placed on organizations by technical activities. Rather than focus on the design of systems, HRO is interested in what forms organizations managing systems must take on, and what they must do, in order to manage demanding systems. In terms of the cost-benefit bargain of technology operation, HRO focuses primarily on understanding how organizations maintain an ability to operate high risk technical systems through management practices rather than through system design. For example, how is the US Navy able to pursue carrier flight operations despite, “...operating under the most extreme conditions in the least stable environment, and with the greatest tension between preserving safety and reliability and attaining maximum operational efficiency... with a young and largely inexperienced crew, [and] a ‘management’ staff of officers that turns over half its complement each

year...in a working environment that must rebuild itself from scratch approximately every eighteen months?" (Rochlin, LaPorte et al. 1987)

In the case of our Mars spacecraft, or any similar system demanding of highly dependable system performance, HRO and dependable system concepts intertwine. Questions, such as the following, key to planning such a mission and to devising a spacecraft configuration, combine high reliability and system dependability concerns:

- How can we provide the greatest probability that ground and space based flight operations are vigilant and attuned to the right set of system health indicators over the duration of the mission? What signals should either the system, or system operators, use to diagnose whether operational effectiveness is slipping in ways that should cause concern?
- What diagnostic, communication, and cognitive problems will the proposed spacecraft configuration present, and how can these features be ameliorated by design, training, management or other means?
- What technical choices might designers make that would cause operational regrets later? How well can the spacecraft be maintained during flight?
- What bundle of resources and institutional capabilities represents the minimum below which reasonable likelihood of a desirable mission outcome cannot be professionally justified?

Where HRO does differ, and substantially so, from dependable system research is in the focus of dependability on the system design phase. Dependability design practitioners tend to be participants in the technical creation process of a complex technical artifact. The HRO program is primarily interested in understanding how systems behave during operational phases of existence, from the standpoint of organizational theory.

HRO efforts and interests emerged from responses to a hypothesis critical of technical system development advanced by Charles Perrow labeled Normal Accident Theory (NAT). Looking *ex-post facto* at a number of different modern technical accidents, Perrow distilled a set of characteristics categorizing technologies that seemed to be at the limit of any known organizational form to manage without resulting in catastrophic failures as a result of normal operations—hence the idea of normal accidents (Perrow 1984; Sagan 1994).

These characteristics (rapid system feedback dynamics Perrow calls “tight coupling” and system complexity) are used to identify types of technology that NAT proponents feel cannot be reconfigured to pass the standards of the social-technology bargain. Rather, Perrow and others argue that efforts at providing dependable system features, such as through improved safety diagnostic capabilities instead inevitably decrease overall system understanding and create surprise failure modes outweighing any possible safety benefit. Further, in the absence of a positive general social benefit to extremely NAT-like systems (for example nuclear power, some space exploration, and genetic engineering) coercion is required to explain why such activities exist.

The HRO project, in turn, has attempted to bring greater nuance to the debate over technology in the social sciences. Starting with the observation that despite tight

coupling and system complexity a number of organizations in the modern world do manage to perform effectively, seemingly well enough to pose a basic challenge to NAT or traditional organizational theory explanations, HRO researchers argued:

that the attention being paid to studies and cases of organizational failure was not (and still is not) matched by parallel studies of organizations that were (and are) operating safely and reliably in similar circumstances...From our preliminary observations, and discussions with our original contacts, we thought that the three activities--air traffic control, electric utility grid management and the operation of a US Navy aircraft carrier--had much in common...All had similar challenges to maintain reliability, performance and safety, simultaneously, at very high levels and similar dependencies upon the individual and collective skills and high degrees of responsibility of human operators. They posed similar conundrums for managers seeking to keep operational performance high in the face of continuing pressure to achieve higher levels of performance at lower cost without thereby increasing the risk to the organization or to the public (Rochlin 1996, p55).

The results from these initial studies, and following work in a variety of settings including medical, spacecraft, and other types of system operations, led to a set of findings that I will present in the next section as being relevant to dependable design.

Within the social sciences, however, the subsequent conflict between NAT and HRO has generated controversy; this controversy is inherently of little impact in the use of HRO research as part of dependable design efforts (Weick 1993). However, the debate has left its mark in some of the HRO-NAT literature and obfuscated some of the research points of relevance to external parties (such as the Dependable Design community) amidst a conversation mostly of interest within the social sciences.¹

The vital question that should not be obscured behind a disciplinary squabble is this: what can designers and producers of the benefits of modern existence do to limit the costs associated with potentially high risk technical systems? This is a question that HRO has spent two decades addressing from the perspective of the social sciences and organizational theory. The results are a body of work relevant to dependable system design. From the standpoint of HRO, engagement with the Dependable Systems community offers access to the technical system design process and opportunity to build a focused capability for social science dialog with engineers.

Lessons from the Field: HRO Patterns of Behavior

Gene Rochlin, one of the founders of the project (along with Todd La Porte, Karlene Roberts, and Paul Schulman) has framed the main observations from HRO field research as:

¹ For more about this discussion between NAT and HRO see Pinch, T. (1991). How Do We Treat Technical Uncertainty in Systems Failure? The Case of the Space Shuttle Challenger. Social Responses to Large Technical Systems. T. Laporte. Boston, Kluwer Academic Publishers. **58**: 143-157, also Rijpma, J. A. (1997). "Complexity, Tight Coupling and Reliability: Connecting Normal Accidents Theory and High Reliability Theory." Journal of Contingencies and Crisis Management **5**(1): 15-23.

...they [HRO] operate equipment whose complexity and inherent vulnerabilities are such that technical and/or physical failures will occur, regardless of engineering solutions or methods applied. The role of operators is not just to operate the equipment, but actively seek to anticipate, detect and correct technical failures as they occur. The undetected malfunction of some part (e.g. a bolt in an aircraft landing gear) is therefore seen not as an equipment error but as a *systemic* failure. Technical and anthropologic causes merge...The underlying belief, and the foundation of organizational reliability [of HRO] is...it can continue to operate only if both the probabilities and, to some extent, the risks themselves are effectively managed through its own training, skills, and error-detection and correction mechanisms. The alternative is probably the external imposition of intrusive regulation or increasingly stringent operational requirements, both of which are seen by the organizations as having a decidedly negative impact on safety (Rochlin 1993, p.19).

Taking these findings about 1) the interconnected non-isolatable nature of human and equipment in creating risks, 2) that risk (from whatever source) are dynamic and partially controllable through on-going vigilance and 3) that HRO technical activities function within a framework of a social relationship which can be harmed by operational failure, the social benefit-cost framework of interest to HRO can be rethought of as follows:²

$$\text{Social Benefit-Cost of Activity} = \sum_i ((B_i - \sum_{i,j} (X_{ii}P_{ii} + X_{ji}P_{ji})))$$

where

B_i is the collection of social benefits per unit of time produced by the technical activity

X_i, X_j are specific per time unit hazards from equipment and anthropogenic errors

P_i, P_j are specific per time probabilities of hazard occurrence from equipment and anthropogenic error

However :

P_i, P_j, X_i, X_j are non-independent results of common factor operator organizational actions/characteristics

B_i, X_i, X_j are functions of common factors external to the organization such as public dread, trust/confidence

From the perspective of an interested dependability system designer, HRO research can be categorized in terms of efforts to better understand the variables described above. Through the use of immersive ethnographic and historical system studies, practitioners have sought to understand how to manage the complex dynamic relationship between technical activities posing dependable/highly reliable operational challenges, and the maintenance of sufficient social permission to make continued operation possible.

Inseparability of Systemic Equipment and Anthropologic Hazards (Xi, Xj)

The inseparability of human and equipment causes for system hazards has been explored in a number of different settings; this is one of the most robust findings about technical system management emerging from social science research in the past 40 years (Turner

² Presented with modifications for the following discussion from Rochlin's original in Rochlin, G. (1993). Defining "High Reliability" Organizations in Practice: A Taxonomic Prologue. New Challenges to Understanding Organizations. K. H. Roberts. New York, NY, Macmillan Publishing Company: 11-32.

1976; Wildavsky 1988; Rochlin 1989; Clarke and Short 1993; Heimann 1993; Reason 1995; Kennedy and Kirwan 1998; Weisbecker 1998; McLaughlin, Monohan et al. 2000; Morris and Moore 2000). Dependable system design shares this interest with the social science literature on accident causation. Unfortunately, the conclusiveness of research performed in this area has yet to be translated into improvements in disaster forensics, system reconfiguration mechanisms, or legal assignment of blame. Frequently, efforts to improve performance by “designing out human error” (or assign culpability in the form of “operator error”) endure as the default system risk/hazard management strategy (Rodgers 1992; Rees 1994; Lancaster 1996; Tenner 1996, Turner, 1976; Pool 1997).

HRO has contributed substantially to research on the problem of prematurely assigning sole responsibility for system failures to (often hapless) system operators. These contributions can be thought of as varying along an axis measuring organizational size. At the most micro level of organization, HRO researchers have studied small team interactions with complex systems in an immediate decision-making setting. For example, researchers have worked with flight crews, studying the interaction of fatigue, crew communications, and task complexity (Foushee and Lauber 1993). Other HRO studies looking at the causes disasters such as air crashes, medical equipment failures, marine accidents, and industrial catastrophes, have varied between individuals and large groups as loci of decision-making.

While it is impossible to do this thread of the HRO literature justice in this paper, dependability design may find a couple of conclusions especially notable. The first is the beginning of an ability to explain the ways that systems are likely to fail if attempts are made to naively design hazards “out of the system.” Both Vaughn and Pinkus, in their separate studies of the Space Shuttle Challenger accident, identify complex interactions between previous “hazard” design fixes and with operational misjudgment (Vaughn 1996; Pinkus, Shuman et al. 1997). Gene Rochlin’s examinations of how the unintended consequences of computerization can result in loss of piloting skill and how the removal of friction from complex systems can result in unforeseen tight-coupled failures are especially important (Rochlin 1993; Rochlin 1997). Similar work by Chris Demchak has proved remarkably prescient in light of present US Army difficulties employing aspects of Future Combat System technology in Iraq (Demchak 1996; Talbot 2004). There are many other examples (La Porte 1988; Bea and Moore 1993; Roberts and Moore 1993, La Porte, 1995; Schulman, Roe et al. 2004).

The second observation of special use in understanding the dynamic relationship between X_j and X_i hazards has been developed within the HRO community through study of cross cultural system operations studies (Rochlin and von Meier 1994; Rochlin, Cook et al. 1995; Bourrier 1996). Through study of how hazards and operational patterns differ between technical systems, such as comparison studies of air traffic control systems, or of nuclear power plant operation in Europe and the United States, HRO researchers have been able to explore the interrelation of human and technical hazard generation mechanisms. Holding technology relatively constant (for example because French reactors are based on US Westinghouse PWR designs) these studies very convincingly show how differences in hazard generation cannot be attributed to “social” or “technical”

factors. The continual interaction between the social and technical factors creates different system “meanings,” and although system parts may be the same between French and American reactors, system use and risk management can vary greatly. Accordingly, engineering safety controls and operational patterns are not isolatable during system operation—even “human proof” designs depend on behavioral patterns from system operators in order to reduce system risk. For example, multiply redundant safety-class systems such as diesel generators must be maintained, tested and inspected. Likewise, no hazard-limiting design can withstand a “fault propagating” operational culture in which operators steal parts to make up for a failure to meet the monthly payroll (Zimmermann 1995).

The difficulty in obtaining access to comparable technical systems in multiple countries for purposes of research cannot be overstated. The small set of HRO studies undertaken is unique, in that the same set of scholars was able to spend substantial amounts of time observing complex technical systems, with sufficient cultural/technical interpretive ability to make sense of what they observed. The complex dynamics of technical and anthropologic hazard remain at a basic level of understanding; dependability and the challenge of design for highly reliable performance would be greatly served by greater number of these controlled observations.

Dynamic Management of System Risks (XiPi, XjPj)

The second broad category of HRO results of interest to Dependable System community is concerned with how operators responsible for producing a social benefit try to manage probabilities and costs so that the technical system can survive under the social bargain (Wolf 2005). HRO work in this area is large, and useful for its focus both on organizational successes and failures.

Overall, the attentions of practitioners, in understanding how the risk dynamic must be managed effectively by systems with strong requirements for reliability performance, have been focused on identification rather than explanation. This is an inevitable result of the very long time scales involved in organizational development—even several years of observation cannot provide “the answer” as to how the reliability culture of the US Navy developed. In lieu of explanation, HRO study of management has been driven by first the observation of anomalous patterns of behavior relative to expectations, and then by the attempt to identify what organizational characteristics correlate with those anomalies. These expectations, in turn, have been derived from much of the “standard” literature on the behavior of bureaucracies and other formalized organizational structures—much of which is not greatly interested in technology or technical risks.

Because correlation is not causation, HRO studies of management cannot—and do not—claim to have crafted a set of standard recipes an organization can “cook from” as part of adopting a technical activity demanding of high reliability. Instead, over time practitioners have developed a robust set of characteristics that can be observed in common between “reliability seeking organizations.” Identification of practices in the

real world which suggest how reliability dynamics can (at least for some time) be managed is a matter of recognizing regularities in a constantly changing pattern. HRO researchers do argue these patterns can be grown and fostered through attentive management over time; they cannot be created through any single action.

Examples of such studies of “reliability seeking organizations” have been found both in military and civilian settings (Rochlin, LaPorte et al. 1987; Roberts 1990; La Porte and Consilini 1991; Weick and Roberts 1993; La Porte 1997; Vogus and Welbourne 2003). Reliability seeking behavior is comprised of a number of interrelated management patterns. Probably the most important is the recognition that risks pose a dynamically manageable, but not controllable, challenge. HROs manage against outcomes that are treated as the outcome of processes rather than as events. Todd La Porte has described this as “prideful wariness” in aircraft carrier culture, a sense of, “high technical/professional competence and technical knowledge of the system and demonstrated high performance and awareness of the system's operating state.” (La Porte 1996, p.63)

From the standpoint of designing sensors and supplying systems with information necessary for dependability, HRO findings on internal exchange of information and the importance of effective external monitoring are especially salient. In response to the pressures of maintaining vigilance under threat from unacceptable failure modes, HRO research has worked at describing how communication patterns and content relate to system demands. HRO operations depend upon, “keen situational awareness for decisive action to be taken.” (La Porte 1996, p.65) Situational awareness has been observed to be maintained differently depending on whether the system is operating in surge or standard states of operation. For example, studies of air traffic control activities have shown how a wide variety of formal and informal information is used by managers and controllers to determine if co-workers are safely operating “in the bubble” and to allocate “extra eyes” when traffic loads increase (Roberts and Rousseau 1989). Studies of operations on board the USS Vincennes during the Iran air-liner shoot down, of glove-box operations at Los Alamos National Laboratory and in other settings have also been undertaken to get at this fascinating issue.

Maintaining effective external oversight of the system, through transparency with a strong overseer provides another key component of HRO information handling characteristics. As Todd La Porte puts it, “HRO performance is centrally associated with extraordinary dense patterns of cooperative behavior within the organization...Continuous attention to both achieving organizational missions and avoiding serious failures requires sustained interaction with elements in the external environment, not only to insure resources, but, as importantly, to support internal resolve to maintain internal relations and sustain the HRO's culture of reliability.” (La Porte 1996, p.65) Among some HRO, these external observers have been used as an important source for management efforts to ensure vigilance is not degraded due to institutional boredom or goal displacement. In the case of large technical systems, the design of dependability features may include consideration for how information generated by the system is shared externally and the role of outside observers. For example, in the case of our Mars spacecraft, dependable

function may well require the use of human or computerized “red team” methods during a multi-year flight to protect against group-think, mission control errors, or over-commitment of flight crew stamina (as has been observed on the ISS & MIR) (Burrough 1998).

Communication content, and type of decision-making structure, is also highly varied depending system characteristics. In this, Paul Schulman’s work on a typology of HRO decision-making is most notable (Schulman 1993). Schulman describes how varied kinds of organizational information processing methods have developed among HRO organizations faced with different technical system characteristics. Dividing reliability system demands between requiring clearance or action focused analysis and between decomposable or holistic action, Schulman (and others) use these patterns to show why pursuit of vigilance requires more localized decision authority for some systems, such as air traffic control centers, versus more formalized risk management methods appropriate in other cases (such as electrical distribution centers).

In related work, these findings have been extended to look at information content and even at how HRO organizations use such typically overlooked tools such as “hero stories” to establish behavioral norms and teach stimuli response patterns within the framework of system risk management (Schulman 1996). Florman has also discussed similar engineering related norms although not in the context of HRO (Florman 1997). These observations suggest that data gathering and sensor effectiveness as a system control tool for dependability will require careful consideration of how the operating organization is structured, the degree to which the organization is capable of a high degree of performance, and how information is communicated internally and externally.

Social Perceptions of Benefits and Hazards (Bt, Xj, Xi)

The third major category of HRO literature examines the relationship between how a technical system is operated and how external audiences perceive the impact of associated hazards and benefits.³ Within the literature, this conversation has been framed around the idea of “institutional stewardship.” By examining the external connections between reliability seeking socio-technical systems and external public and policy-making audiences, the HRO project has identified two key aspects to long term endurance as a steward for the public of a risky technical activity. La Porte describes these dual characteristics of public trust and confidence and institutional constancy as long term organizational burdens requiring organizational evolution of, “institutional properties that...signal the public trustworthiness of the organization, and...if [the activity] is seen as having a social function demanding effectiveness into a far reaching

³ Risk judgments by the external public tend to be more strongly based upon the degree of dread induced by the hazard and the perceived magnitude of the hazard more than estimates of occurrence probability, (Slovic 2000; Morgan 2002). . Accordingly, while system managers may be able to reduce the actual risk of the activity, they may not be able to manage public perceptions of the benefits and of these activity-related hazards. In cases where these assessments of relative benefits and hazards diverge widely, external publics and system operators to a great degree are comparing results from two entirely different benefit-cost frameworks.

future and/or the potential to put that future at risk for many years, to show that as an institution it can assure the public of faithfulness, as well as continuously available highly reliable capacity. Absent these, the political legitimacy of the enterprise is at stake.”(La Porte 2000, p.6) also (Koehler 2003)

Institutional constancy emerged in part from involvement by members of the Berkeley HRO community in a set of studies involving nuclear weapon, power, and waste technologies (La Porte and Keller 1996). The facilities and technical communities studied all faced the direct challenge of both requiring a high degree of operational reliability to prevent high hazard accidents, and requiring that this level of reliability continue for the indefinite future. There are remarkably few organizational models of any kind that exist as examples of enduring entities that can maintain even a basic similarity with their original purpose. The bulk of histories for long-lived organizations end with some degree of public regret for pollution created, lives lost, promises unmet, and structural impediments created.⁴ Religions, a few financial institutions, and national bodies such as the Armed Forces provide only partial models of success—either they fail to require the operation of complex hazardous technologies, or they do so by relying on a discipline, command, and motivational structure unavailable to most socio-technical entities. Over time, it is natural for vigilance to decline as operations become familiar, as mission drift sets in, and through processes Vaughn has labeled the “normalization of deviance.” What organizational mechanisms will prevent “drift into regret-inducing behavior” are not well known—perhaps through the involvement of the dependable system design community, these constancy eroding mechanisms can be better understood and designed against.

Based on observation of system failures (the perpetually aborted efforts at Yucca Mountain for example) public trust and confidence plays a major role in determining the operational path of hazardous systems. From both an ethical and pragmatic standpoint, sustained HRO institutional stewardship can not often survive the loss of resources and reevaluation of system hazards/benefits that come with loss of public confidence (La Porte 1996; La Porte and Metlay 1996). Gene Rochlin has described the loss of trust as one that converts the technical system into a “regulatory magnet,” adding, “[when] quasi public socio-technical institutions penetrate deeply into the structure of modern societies, a considerable amount of institutional trust is required if they are to be allowed to continue their tasks without intrusive and potentially damaging micro-management.” (Rochlin 1996)

As in the case of any trust-based contract, trustworthiness of HRO organizations depends upon the nature of communication and transactions between the principal (the public, external overseers involved in the social-technical benefit/cost bargain) and the agent (the operators of the social-technical system). By agreeing to allow the system to exist, at least at some basic level the public expectation is created that system operators will take, “[public] interests into account, even in situations where [the public] is not in a position to recognize evaluate and/or thwart a potentially negative course of action by ‘those

⁴ Such impediments may be outdated worker skills, legal impediments to future types of technical activities, or unintended depletion of resources that “lock-in” or limit future technology choices.

trusted'... [and] the party trusted is competent to act on that knowledge and will go to considerable lengths to keep its/her/his word.” (La Porte 2000, p.7)

The maintenance of trustworthiness is a key problem for organizations, and a variety of case studies have been written within and without the HRO project describing organizations suffering from what La Porte has labeled a trustworthiness “deficit.” Simply because it is far more difficult to locate organizations that are fostering trust than those who have entered into a deficit, HRO researchers have developed a much better understanding of how organizations are damaged by mistrust than how they recover. Hazards which were once seen as acceptable no longer are permitted—from the internal standpoint of the socio-technical organization this can result in seemingly arbitrary external restrictions or periods of enforced “stand-down.”

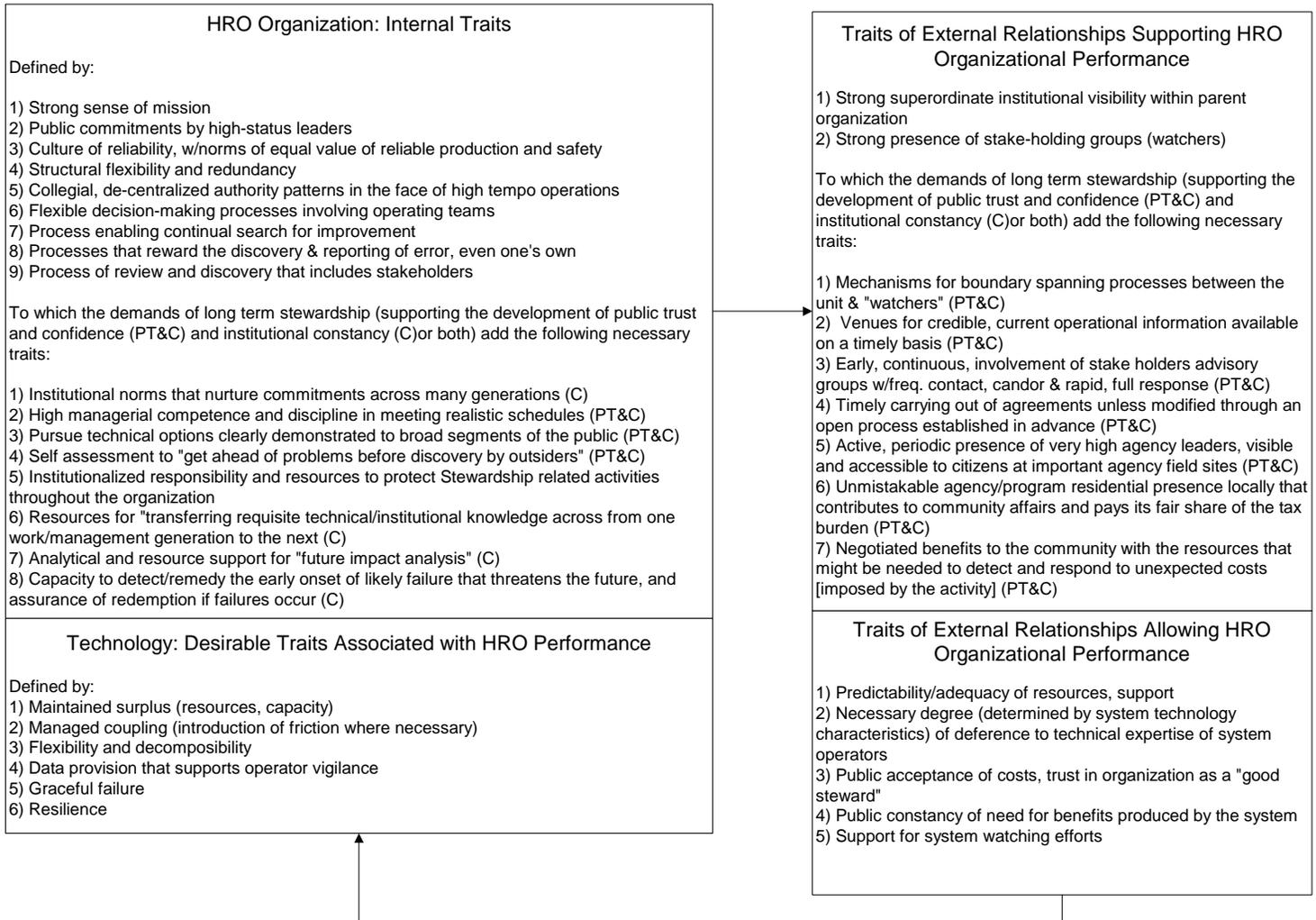
Recovery is a long, asymmetrically difficult process requiring extraordinary prevention of hazards/costs (since external viewers observe any subsequent failure as part of a pattern of deceit) and especially strong adherence to social norms such as allowing information to “boundary span” not just internally, but externally as well, through public acts of contrition by senior managers, and remediation of damage. From the standpoint of dependable systems thinking, HRO work on communications and signaling of trustworthiness between operators and principals is potentially of interest. To date, some work exists characterizing the dynamic interchange of signals between the socio-technical system and the external environment (La Porte 1994; La Porte 1996). However, the dynamics of public loss of confidence remain explanatory rather than predictive. A question for dependable system designers is this: could systems be created so that they provide information about their health to observers in ways that would enhance trust?

Dependable Design, Organizational Behavior and Connections to the HRO Project

So far, this paper has described the major areas of the Highly Reliability Organization project using the basic social benefit-cost “contract” inherent to hazardous system operation as an organizing principle. The question from a system design standpoint is how does HRO research fit into the framework of more predictive effort—if we are faced with designing a spacecraft, or similar, how can we draw upon the HRO body of literature as a whole to create systems with more desirable characteristics?

Adapting a synthesis of HRO internal and external traits, and those observed to be associated with long term operation of hazardous systems (requiring public trust and confidence and institutional constancy) developed by Todd La Porte, the following description can be pulled together⁵:

⁵ Diagram redrawn from La Porte, T. (2000). Highly Reliable Operations and the Rigors of Sustained Legitimacy: Matters of Public Trust and Institutional Constancy. Printed in French as *Fiabilité et Legitimaite Soutenable [Reliability and Sustainable Legitimacy] in Mathilde Bourrie, Organiser la Faibilité*. Paris, France: 20., with addition of “Technology Desirable Traits” derived from La Porte, T. and



Looking at the above diagram from the perspective of an engineer, what are we to make of this? It is not a defined requirements set as we might wish to have from the standpoint of integrating HRO and dependability features into our hypothetical spacecraft design. This is not the fault of the HRO project. As Gene Rochlin explains the social scientists in the team did not initiate their work as part of an engineering effort:

Although we have, from time to time, been urged to generalize or adapt our work for the purpose of original design, many of the things we learned from working with these organizations makes us very cautious. The experience is such that there were several instances where our credibility depended upon our explaining that "making things work better" was not the reason for our being there, nor the purpose of our work. We were not engaged in a search for excellence, although we saw much that was indeed excellent, nor for a prescriptive set of rules or procedures for avoiding errors and failures. We did not attempt to span the universe of possible organizations, or technical systems, to generate

P. M. Consilini (1991). "Working in Practice but Not in Theory: Theoretical Challenges of "High Reliability Organizations"." Journal of Public Administration Research and Theory 1(Winter): 23-49.

comparative studies or speculate on the relative frequency of organizational successes and failures, but worked with a very special set for the explicit purpose of trying to determine how and why they performed so well, and why others considered that performance to be so special. (Rochlin 1996, p.55)

Yet, while not a set of requirements, the HRO team has gathered, at great expense and difficulty, a set of regularities: observed traits that describe patterns of use in crafting systems with more desirable characteristics. Not all of these traits were observed by the HRO effort in all of the systems they studied, nor could they point to specific values that let them know the degree to which these systems clustered along axes of measurement. Social science is the study of immensely complex multi-person constructs (organizations, cultures, nations) comprised of immensely complex individual entities. Because the behavior of willful agents can only sometimes be predicted by physical laws (often these degenerate cases ultimately involve fatalities) social science practice in many cases must depend on appeals to patterns rather than predictions, to correlation rather than causation. Furthermore, there is far too much evidence that organizational factors matter: the same spacecraft operated by two different entities will not perform identically. If dependability is an important, or even an ethical, consideration given the possible loss of taxpayer dollars and crew-member lives—then we want organizational form and goals to support spacecraft capabilities and dependability.

How then can we include these regularities into engineering efforts of dependable systems? This is a question for research between the HRO and Dependable Design communities. Some paths this research could take that the author has been exploring will be discussed in the following section. Improved socio-technical system design and prediction tools are essential as the capabilities and hazards of such systems increase. Although the HRO effort has not in the past had opportunity to become more involved in prediction related activities, past focus on social science methods does not imply a lack of willingness to engage with engineering concerns. As Gene Rochlin continues:

To extend our work, to a more general survey of other organizations similar along one or more of the several dimensions we explored would have become a major project far exceeding the time and resources of the original core group. What we hoped to do instead was to stimulate others to test our hypotheses and framework on other organizations, performing under similar circumstances, to which they had or could obtain their own access. To adopt a metaphor used by Roberts, the original HRO work was a fountainhead from which issued many streams of possible research and inquiry, to a variety of purposes.

Possible Trajectories for Organizational Analysis in the Design of Dependable Systems

Where, then, the future of HRO in contributing to the sort of system improvements suggested by the Dependability Design program? Operational and organizational context matters in determining system outcome. Whether system operators are mindful, or institutionally incapable, can tarnish otherwise well designed and crafted technical

systems—at the cost of lives, needed capabilities and, not incidentally, the waste of engineering careers.⁶ From the standpoint of Dependability Design, organizational form represents a source of tremendous operational and performance variance, potentially impacting all levels of system prediction.

Looking at HRO from the perspective of an engineer, there is a strong temptation to see the regularities developed through the study of reliability seeking organizations as a set of primary requirements—which with further quantification could be translated into verifiable and validated derived requirements, and then written into system specifications. For example, why not translate the observation of “strong sense of mission” into “95% of operators correctly know why they go to work in the morning?” Unfortunately, the materials at hand for physical engineering are far more plastic than what any similar “social engineer” might have to work with. While, in some operational and system contexts, some of these regularities might be further derived and quantified as part of performance measurement, in general such efforts will fail because they are based on a mistaken model of organizations. Organizations are only crafted in part; they also have lives, and life-cycles of their own, created by complex internal dynamics, personalities, and histories.

Accordingly, benefits to Dependable Design from use of HRO observations will probably result from a multi-faceted effort. Human behavior is very hard to design; it is far more profitable to design systems in ways that seek to bring out the kinds of human behavior designers wish to see (Rasmussen 1994). For example, if a system has tightly coupled, potentially harmful failure modes, HRO literature suggests that Dependable Design criteria should suggest configurations, information gathering, and control structures that help to maintain operator vigilance. New airport baggage X-ray machines intentionally superimpose false images of weapons at random times to encourage operator attention in the face of an otherwise mind-numbingly boring job. Rather than caretakers, systems are best operated by engaged participants who have “the feel” for when the system is either running “sweet” or “acting hinkey.”⁷

Instead of pushing operators out of the system by “idiot proofing” a control room down to one dial and a button, care should be taken to require that operators are forced to interact with the system. For example, well designed maintenance access and telemetry allows for workers to maintain skills at diagnosis, while improving operational transparency. The US Navy has deliberately relied upon “outmoded” control systems in nuclear sub reactors in part for this reason. Likewise, if systems are given high degrees of operational autonomy, dependability designers may want to balance that autonomy with consideration of how well human actors will be able to modify, or understand system function, when unforeseen operational modes and combinations of external stimuli and events are (inevitably) encountered.

⁶ For example the professional life of many who went into aeronautical engineering in the 1980s, or nuclear engineering in the 1970’s, has probably not been entirely as hoped.

⁷ In the words of one shift manager I met at Diablo Canyon NPP.

Beyond efforts to improve human operational oversight and reliability seeking behavior without requiring a redesign of “the human subsystem,” the reliability seeking organizational patterns identified by the HRO project suggest that as in the case of organizational reliability, system engineered dependability is a process not an event. Beyond describing a few basic technical characteristics relating to reliability, there likely exist patterns of dependable system design that should be developed more completely. Certainly, traits such as system decomposability and graceful failure modes are desirable—however how can the Dependable System community draw these characteristics out and understand their interactions in the context of the design of competing system architectures?

Further, if dependability is thought of as a process desirable for system technical components, just as reliability is a process desirable for organizations faced with hazardous operations, system flexibility becomes a great concern. As a process, akin to the intermeshed organizational processes of reliability seeking, dependability implies activities that will persist for the life of the system rather than simply ending at the design life-cycle phase. Flexibility of design will therefore be essential in allowing dependability efforts to re-hone system configuration as design problems, or as the form and degree of reliability seeking by operating organizations becomes apparent.

This suggests two simultaneous trajectories for further interactions between the HRO project and Dependable System efforts. The first is that greater work needs to go into the study of how analog types of systems are operated by different organizations; for example several branches of the US military operate similar or identical systems, however they do so with different missions and in distinct operational contexts. By observing the performance differences a great opportunity exists to improve an understanding of how technical characteristics match up with organizational patterns (Woo and Vicente 2003). Such case studies would also provide both the Dependability and HRO communities with an improved sense of how features enhancing system flexibility relate to organizational behavior.

The second research trajectory suggested by HRO observations of reliability seeking patterns is greater focus on recursive system performance simulation. To date, some basic work in the aeronautical engineering community, primarily, has been done to explore how robust alternative spacecraft designs are against changes in production schedule, funding, or subcomponent reliability (Dillon, Pate-Cornell et al. 2002; Hastings and Weigel 2004). With development, this work holds promise in the exploration of how systems can be made more flexible against the kinds of issues described by HRO patterns of behavior. For example, it would be of great use to be able to explore how competing designs differ in their ability to take on missions beyond those presently tasked, or to understand what it would take to continue operation for far longer than presently intended, or to determine failure modes should key skills and resources become unavailable. As ability to design and predict performance of physical systems improves, the need for matching development of methods to study system-organizational interactions is becoming clear. Dependability design and HRO are useful frameworks that should jointly play a role in improving future system development efforts.

Bibliography

- Bea, R. G. and W. H. Moore (1993). Operational Reliability and Marine Systems. New Challenges to Understanding Organizations. K. H. Roberts. New York, NY, Macmillian Publishers: 199-231.
- Bourrier, M. (1996). "Organizing Maintenance Work at Two American Nuclear Power Plants." Journal of Contingencies and Crisis Management **4**(2): 104-113.
- Burrough, B. (1998). Dragonfly: An Epic Adventure of Survival in Outer Space. New York, NY, Harper Collins.
- Clarke, L. and J. F. Short, Jr. (1993). "Social Organization and Risk: Some Current Controversies." Annual Review of Sociology **19**: 375-399.
- Demchak, C. C. (1996). "Tailored Precision Armies in Fully Networked Battlespace: High Reliability Organization Dilemmas in the "Information Age"." Journal of Contingencies and Crisis Management **4**(2): 93-104.
- Dillon, R. L., E. Pate-Cornell, et al. (2002). "Programmatic Risk Analysis for Critical Engineering Systems Under Tight Resource Constraints." Operations Research **51**(3): 354-370.
- Florman, S. C. (1997). Technology and the Tragic View. Technology and the Future. A. H. Teich. New York, NY, St. Martin's Press: 93-106.
- Foushee, C. H. and J. K. Lauber (1993). The Effects of Flight Crew Fatigue on Performance: A Full Mission Simulation Study. New Challenges to Understanding Organizations. K. H. Roberts. New York, NY, Macmillan Publishing Company: 151-173.
- Hastings, D. E. and A. L. Weigel (2004). "Measuring the Value of Designing for Uncertain Future Downward Budget Instabilities." Journal of Spacecraft and Rockets **41**(1): 111-119.
- Heimann, C. F. L. (1993). "Understanding the Challenger Disaster: Organizational Structure and the Design of Reliable Systems." American Political Science Review **87**(2): 421-435.
- Kennedy, R. and B. Kirwan (1998). "Development of a Hazard and Operability-based Method for Identifying Safety Management Vulnerabilities in High Risk Systems." Safety Science **30**: 249-274.
- Koehler, A. (2003). Defining Risk and Safety in a High Security Organization: "Bunkering" at the Los Alamos Plutonium Handling Facility. Constructing Risk and Safety in Technological Practice. J. Summerton and B. Berner. New York, NY, Routledge: 106-119.
- La Porte, T. (1988). The United States Air Traffic Control System: Increasing Reliability in the Midst of Rapid Growth. The Development of Large Scale Technical Systems. T. Hughes and R. Mayntz. Boulder CO, Westview Press: 215-244.
- La Porte, T. (1994). "Large Technical Systems, Institutional Surprises, and Challenges to Political Legitimacy." Technology in Society **16**(3): 269-288.
- La Porte, T. (1996). "Hazards and Institutional Trustworthiness: Facing a Deficit of Trust." Public Administration Review **56**(4): 341-347.
- La Porte, T. (1996). "High Reliability Organizations: Unlikely, Demanding and at Risk." Journal of Contingency and Crisis Management **2**(4): 60-72.

- La Porte, T. (1996). "High Reliability Organizations: Unlikely, Demanding and at Risk." Journal of Contingencies and Crisis Management 4(2): 60-72.
- La Porte, T. (1996). Large Technical Systems as a Source of Social/Institutional Strain: A Conceptual Note. Berkeley, CA, Center for Nuclear and Toxic Waste Management.
- La Porte, T. (1997). Evolving High Reliability Organizations and Institutional Strain in Elements of the US Nuclear Future, Final Report. Berkeley, University of California, Berkeley, Center for Nuclear and Toxic Waste Management and Los Alamos National Laboratory.
- La Porte, T. (2000). Highly Reliable Operations and the Rigors of Sustained Legitimacy: Matters of Public Trust and Institutional Constancy. Printed in French as Fiabilité et Legitimaite Soutenable [Reliability and Sustainable Legitimacy] in Mathilde Bourrie, Organiser la Faibilite. Paris, France: 20.
- La Porte, T. and P. M. Consilini (1991). "Working in Practice but Not in Theory: Theoretical Challenges of "High Reliability Organizations"." Journal of Public Administration Research and Theory 1(Winter): 23-49.
- La Porte, T. and C. W. Thomas (1995). "Regulatory Compliance and the ethos of quality enhancement: Surprises in Nuclear Power Plant Operations." Journal of Public Administration Research and Theory 5(1): 109-137.
- La Porte, T. and A. Keller (1996). Assuring Institutional Constancy: Requisite for Managing Long-Lived Hazards. Berkeley, CA, Department of Political Science.
- La Porte, T. and D. S. Metlay (1996). "Facing a Deficit of Trust: Hazards and Institutional Trustworthiness." Public Administration Review(July): 13.
- Lancaster, J. (1996). Engineering Catastrophes: Causes and Effects of Major Accidents. Cambridge, UK, Abington Publishing.
- McLaughlin, T. P., S. P. Monohan, et al. (2000). A Review of Criticality Accidents: 2000 Revision. Los Alamos, NM, Los Alamos National Laboratory.
- Morgan, G., B. Fischhoff, et al. (2002). Risk Communication: A Mental Models Approach. Cambridge, UK,
- Morris, M. W. and P. C. Moore (2000). "The Lessons We (Don't) Learn: Counterfactual Thinking and Organizational Accountability after a Close Call." Administrative Science Quarterly 45(4): 737-765.
- Noor, A. K. (2004). Perspectives on Nondeterministic Approaches. Engineering Design Reliability Handbook. E. Nikolaidis, D. M. Ghiocel and S. Singhal. New York, NY, CRC Press: 2-1 to 2-19.
- Perrow, C. (1984). Normal Accidents. New York, NY, Basic Books, Inc.
- Pinch, T. (1991). How Do We Treat Technical Uncertainty in Systems Failure? The Case of the Space Shuttle Challenger. Social Responses to Large Technical Systems. T. Laporte. Boston, Kluwer Academic Publishers. 58: 143-157.
- Pinkus, R. L., L. J. Shuman, et al. (1997). Engineering Ethics: Balancing Cost, Schedule, and Risk--Lessons Learned from the Space Shuttle. Cambridge, MA, Cambridge University Press.
- Pool, R. (1997). Beyond Engineering: How Society Shapes Technology. Oxford, Oxford University Press.
- Rasmussen, J., A. M. Pejtersen, et al. (1994). Cognitive Systems Engineering. New York, NY, John Wiley & Sons Inc.

- Reason, J. (1995). "A Systems Approach to Organizational Error." Ergonomics **38**(8): 1708-1721.
- Rees, J. V. (1994). Hostages of Each Other: The Transformation of Nuclear Safety Since Three Mile Island. Chicago, IL, University of Chicago Press.
- Rijpma, J. A. (1997). "Complexity, Tight Coupling and Reliability: Connecting Normal Accidents Theory and High Reliability Theory." Journal of Contingencies and Crisis Management **5**(1): 15-23.
- Roberts, K. H. (1990). "Some Characteristics of One Type of High Reliability Organization." Organization Science **1**(2): 160-176.
- Roberts, K. H. and W. H. Moore (1993). Bligh Reef Dead Ahead: The Grounding of the Exxon Valdez. Challenges to Understanding Organizations. K. H. Roberts. New York, NY, Macmillan Publishers: 231-249.
- Roberts, K. H. and D. M. Rousseau (1989). "Research in Nearly Failure Free, High Reliability Systems: "Having the Bubble"." IEEE Transactions **36**(2): 132-139.
- Rochlin, G. (1989). The Case for Experiential Knowledge. Second International Workshop on Safety Control and Risk Management, Karlsbad, Sweden, Institute for Governmental Studies
Energy and Resources Group.
- Rochlin, G. (1993). Defining "High Reliability" Organizations in Practice: A Taxonomic Prologue. New Challenges to Understanding Organizations. K. H. Roberts. New York, NY, Macmillan Publishing Company: 11-32.
- Rochlin, G. (1993). Essential Friction: Error Control in Organizational Behavior. The Necessity of Friction: Nineteen Essays on a Vital Force. N. Akerman. New York, Springer-Verlag: 196-231.
- Rochlin, G. (1996). "Reliable Organizations: Present Research and Future Directions." Journal of Contingencies and Crisis Management **4**(2): 55-60.
- Rochlin, G. (1997). Trapped in the Net. Princeton, Princeton University Press.
- Rochlin, G., T. LaPorte, et al. (1987). "The Self Designing High Reliability Organization: Aircraft Carrier Flight Operations at Sea." Naval War College Review **90**(Autumn): 76-90.
- Rochlin, G. and A. von Meier (1994). "Nuclear Power Operations: A Cross Cultural Perspective." Annual Review Energy and the Environment **19**: 153-187.
- Rochlin, G. I., N. G. W. Cook, et al. (1995). A Cross Disciplinary Inquiry into Nuclear Power Futures for the U.S. Berkeley, CA, CNTWM.
- Rodgers, R. (1992). Antidotes for the Idiot's Paradox. Technological Innovation and Human Resources. U. E. Gattiker. New York, Walter de Gruyter. **3**: 227-271.
- Sagan, S. D. (1993). The Limits of Safety: Organizations, Accidents and Nuclear Weapons. Princeton, NJ, Princeton University Press.
- Sagan, S. D. (1994). "Toward a Political Theory of Organizational Reliability." Journal of Contingencies and Crisis Management **2**(4): 228-240.
- Schulman, P. R. (1993). The Analysis of High Reliability Organizations: A Comparative Framework. New Challenges to Understanding Organizations. K. H. Roberts. New York, NY, Macmillan Publishers: 33-54.
- Schulman, P. R. (1996). "Heroes, Organizations and High Reliability." Journal of Contingencies and Crisis Management **4**(2): 72-83.

- Schulman, P. R., E. Roe, et al. (2004). "High Reliability and the Management of Critical Infrastructures." Journal of Contingencies and Crisis Management **12**(2): 14-28.
- Slovic, P. (2000). The Perception of Risk. London, UK, Earthscan Press Ltd.
- Talbot, D. (2004). "How Technology Failed in Iraq." Technology Review(November).
- Tenner, E. (1996). Why Things Bite Back: Technology and the Revenge of Unintended Consequences. New York, NY, Alfred A. Knopf.
- Turner, B. (1976). "The Organizational and Interorganizational Development of Disasters." Administrative Science Quarterly **21**(September): 378-396.
- Vaughn, D. (1996). The Challenger Launch Decision. Chicago, University of Chicago Press.
- Vogus, T. J. and T. M. Welbourne (2003). "Structuring for High Reliability: HR Practices and Mindful Processes in Reliability-Seeking Organizations." Journal of Organizational Behavior **24**: 877-903.
- Weick, K. E. (1993). The Vulnerable System: An Analysis of the Tenerife Air Disaster. New Challenges to Understanding Organizations. K. H. Roberts. New York, NY, Macmillan Publishing Company: 173-199.
- Weick, K. E. and K. H. Roberts (1993). "Collective Mind in Organizations: Heedful Interrelating on Flight Decks." Administrative Science Quarterly **38**(3): 357-381.
- Weisbecker, P. (1998). "The Lessons of ValuJet 592." The Atlantic Monthly(March): 81-98.
- Wildavsky, A. (1988). Searching For Safety. New Brunswick, Transaction Books.
- Wolf, F. (2005). "Resource Availability, Commitment and Environmental Reliability & Safety: A Study of Petroleum Refineries." Journal of Contingencies and Crisis Management **13**(1): 11.
- Woo, D. M. and K. J. Vicente (2003). "Sociotechnical Systems, Risk Management, and Public Health: Comparing the North Battleford and Walkerton Outbreaks." Reliability Engineering and System Safety **80**: 253-269.
- Zimmermann, T. and A. Cooperman (1995). Beryllium Deal, Russian Mafia. US News & World Report. New York, NY.