Human Relations
DOI: 10.1177/0018726709339117
Volume 62(9): 1357–1390

The Tavistock Institute ®
http://hum.sagepub.com

Normal Accident Theory versus High Reliability Theory: A resolution and call for an open systems view of accidents

Samir Shrivastava, Karan Sonpar and Federica Pazzaglia

ABSTRACT

We resolve the longstanding debate between Normal Accident Theory (NAT) and High-Reliability Theory (HRT) by introducing a temporal dimension. Specifically, we explain that the two theories appear to diverge because they look at the accident phenomenon at different points of time. We, however, note that the debate's resolution does not address the non-falsifiability problem that both NAT and HRT suffer from. Applying insights from the open systems perspective, we reframe NAT in a manner that helps the theory to address its non-falsifiability problem and factor in the role of humans in accidents. Finally, arguing that open systems theory can account for the conclusions reached by NAT and HRT, we proceed to offer pointers for future research to theoretically and empirically develop an open systems view of accidents.

KEYWORDS

high reliability ▪ negentropy ▪ normal accident ▪ open system theory ▪ requisite variety

It is hardly possible for organizational theorists to write about accidents without referring to Normal Accident Theory (NAT) and High Reliability Theory (HRT). But this is not to say that the theories enjoy uncritical acceptance. Moreover, the proponents of the two theories cannot seem to agree whether their views complement or contradict each other. The genesis of the NAT-HRT debate can be traced back to Sagan's (1993) book, *The*

limits of safety. Sagan applied the theories to analyse accidents and near-misses in the nuclear weapon systems in the US during the Cold War. While Sagan's decision to explicitly contrast and evaluate the theories won support from Perrow (1994: 212), the founder of NAT, who described the comparison as a 'signal service' to organizational theory; it invited criticism from La Porte (1994: 211), the founder of one of the main branches of HRT, who thought that the 'strawman approach' had pitted 'complementary perspectives against each other as if they were in competition'. So who is correct, and why this unusual disagreement between the protagonists?

Previous attempts to reconcile the NAT-HRT debate have proven inconclusive (Rijpma, 1997), indicating that the matter is not straightforward. Vaughan (1999: 296) attributes 'The Great Divide' in the area to the theories focusing on different things and reaching different conclusions. She points out that while NAT focuses on structure and claims that complex and tightly coupled structures inevitably trigger system-wide accidents, HRT focuses on processes and identifies organizational initiatives that can prevent such accidents. Reviewing the relevant literature, Rijpma (2003) asserts that the NAT-HRT debate has reached a dead-end and efforts to resolve the debate would prove unfeasible.

To complicate matters, the reasoning underpinning NAT and HRT appears to be non-falsifiable (Rosa, 2005). If a tightly coupled complex system were to succeed in avoiding an accident, NAT proponents would attribute the safe outcome to the system in question being not complicated enough. Similarly, in the event of an accident in a highly reliable organization, HRT proponents would argue that the accident occurred because the organization had ceased being reliable in that it had not followed recommended processes. Thus the two theories cannot be tested as they can rationalize any outcome and almost always explain away their failure to make a prediction. Noting this weakness, Rosa (2005) issues a call for integrating NAT and HRT through a falsifiable theory. Rosa's call serves as the motivation of this article.

We begin by briefly reviewing and critiquing NAT and HRT. We then revisit the NAT-HRT debate to highlight contentious issues. Thereafter, we attempt to resolve the debate by introducing a temporal dimension. We point out that while HRT focuses on processes related to a dynamic situation and offers insights about the period leading up to the point of accident, NAT identifies the key elements of organizational structure and circumstances at the point of time of an accident. Having established the vantage points of the theories, we claim that NAT and HRT are not incommensurate. We however note that the debate's resolution fails to solve the non-falsifiability problem. In search of solutions, we turn to open systems theory in the latter half of this article.

We treat organizations as open systems and apply our interpretation of systemic properties to reframe NAT in a manner that makes it possible to test the theory. We point out that the reframed NAT can incorporate the damage potential of accidents, and is able to better account for the role of humans in accidents. Finally, we issue a call for developing an open systems view of accidents by discussing how systemic insights can explain the accident phenomenon in general.

NAT and HRT: A critical review

NAT

Analysing the organizational aspects of the nuclear accident at Three Mile Island, Perrow (1981, 1984) concluded that accidents are inevitable or 'normal' in some types of technological systems. Thus NAT got its name. Summarized below are the key notions from NAT (Perrow, 1984) that are germane to our arguments.

Although Perrow does not formally define the term *system*, his notion of a system is critical to NAT. As he points out, any unintended and untoward event that disrupts the ongoing or future output of a system could be viewed as an accident. Consistent with open systems thinking, Perrow argues that one could draw mental boundaries around one's focal system (also see Leveson, 2004). He then divides a system into four levels. In his division, at the first level of a system lies an individual part – for example, a valve. Functionally related collections of individual parts are said to make a unit at the second level. Arrays of units make a subsystem in the third level, and subsystems combine to make a system in the fourth level. Beyond the system lies the environment. In this scheme, failures at the first two levels, even if they temporarily disrupt the output of the entire system, do not qualify as accidents. Instead, they are called *incidents*. So a failure of a unit or a part (say a component like a valve) would be called an incident. Only disruptions at levels three and four would qualify as *accidents*. Perrow points out that most engineered safety features (ESFs), such as redundant components, emergency shut-offs, suppressors, and so forth, are incorporated in systems to prevent incidents from transitioning into accidents.

Perrow (1984) implies that accidents generally begin with failures of one or more lower level components that escalate to higher levels usually through defeating ESFs. If the failures progress through a system to levels three and four in an anticipated sequence, interacting in a manner that is comprehensible to the designers of the system and to those trained to operate it, then the failures culminate in a component failure accident. So, according

to NAT, one could avoid component failure accidents through better designed ESFs, preventive maintenance, operator training, and so forth.

Armed with data from several industries, Perrow (1984) claimed that on rare occasions, some complex systems suffered an accident owing to multiple failures, which interacted with each other in ways that could neither be anticipated nor comprehended. Perrow called such accidents *normal accidents* or *system accidents*. NAT therefore implies that one cannot, by definition, prevent system accidents. All the same, in his treatise, Perrow, mainly with a view to forewarn societies, identifies two properties – *complex interactions* and *tight coupling* – that make systems susceptible to system accidents. We discuss each in turn.

Complex interactions are interactions that occur in unfamiliar sequences, or unplanned and unexpected sequences, and which are either not visible or not immediately comprehensible (Perrow, 1994). The factors said to drive complex interactions in a system include the presence of components that have multiple functions (multi-functionality means that the components can fail in more than one direction at once); physical proximity of components; specialized knowledge of personnel that limits their awareness of interdependencies; several control parameters with potential interactions; and the need to decipher unfamiliar or unintended feedback loops and make inferences.

A system is said to be *tightly coupled* when there is minimal time lag between the processes it executes; the sequence of processing does not vary; there is only one method available to accomplish a task; little slack is possible in supplies, equipment, and personnel; buffers and redundancies are inbuilt with there being little scope of introducing them at a later stage; substitution of supplies, equipment, and personnel is not readily possible, and where possible, it can be done only in a prescribed manner (Perrow, 1984).

The two systemic properties or dimensions – the nature of interactions within a system, and the degree of coupling amongst its subsystems – described above are integral to NAT's main thesis, which may be stated thus: an odd failure in technological systems that are at once *complexly interactive* and *tightly coupled* can, under peculiar circumstances, lead to system accidents. When circumstances are just right, the failure can trigger other failures that can interact amongst each other in a manner that defies comprehension. To make matters worse, the complex interactions can cascade very rapidly in tightly coupled systems, which, by design, as discussed, afford minimal slack and preclude the possibility of substituting either personnel or material. In effect, recovery from failure under such circumstances is almost impossible. The challenge then, from an organizational perspective, is to acquire the capacity to simultaneously cope with complex interactions and tight coupling.

According to NAT, decentralization aids organizations to cope with complex interactions. The rationale is that an organization can respond to unanticipated interactions in real time only if it empowers those proximal to the processes to improvise. Similarly, centralization aids in coping with tight coupling – the thinking being that only an agency that is privy to the big picture can be expected to sensibly override local considerations to ensure the stability of the entire system. Organizations operating complexly interactive and tightly coupled systems cannot be simultaneously centralized and decentralized. Therefore, as per NAT, organizations are structurally incapable of coping with system accidents.

NAT also identifies the factors that promote linear interactions and loose coupling within systems. Thus one can visualize how the interaction and coupling dimensions might be charted along a continuum to generate a two-by-two matrix (see Figure 1). Perrow (1984) populates all the quadrants of the matrix with various types of systems ranging from a post office to a nuclear power plant. NAT however focuses on tightly coupled and complexly interactive systems that populate Quadrant 2 – the top right quadrant.

Interestingly, Perrow observes that all the systems in Quadrant 2 (with the apparent exception of military systems) execute *transformation processes*. That is, the systems transform the main raw material that they work with in some fundamental way.[1] We believe that the observation about transformation processes reveals something important about accidents. For

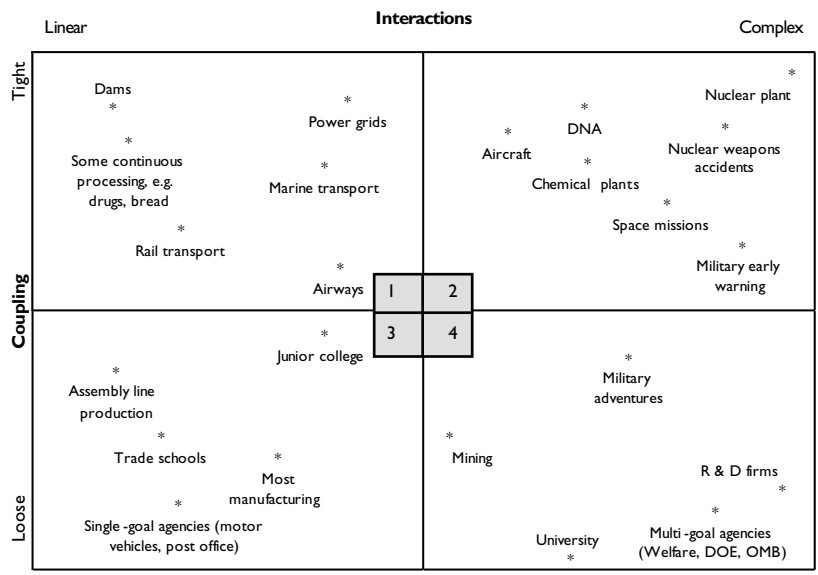

Figure 1 NAT: Interaction/coupling chart

reasons not entirely clear, Perrow does not take his observation to its logical conclusion. We will revisit this issue. For now, we identify one of NAT's major weaknesses that will need to be addressed if the theory is to provide impetus to empirical work.

Our concern pertains to NAT's intra-system levels. Perrow acknowledges that his scheme 'has its ambiguities, since one could argue interminably over the dividing line between part, unit, and subsystem . . .' (1984: 65). All the same, he insists that the scheme has practical value. We argue that while the scheme could help delineate technical systems, it has lesser applicability in an organizational or a socio-technical context. Perrow states that the equivalent of a valve (i.e. a part) in the socio-technical domain could be a human operator. His scheme overlooks the fact that the human mind has the capacity to engage across all of the four arbitrary levels. Operators can deliberately, as Vaughan (1999) warns us, trigger system accidents. And by the same token, they can pre-empt system accidents through timely action. Unless Perrow's scheme can account for such outcomes, it may not be valid to equate an operator failure with merely an incident. Admittedly, shifting the focus away from operator errors and emphasizing organizational properties is one of NAT's major achievements. Nonetheless, we need to ask whether NAT might have swung the pendulum to the other extreme by equating human operators with valves.

We also note that the extant literature contains several examples of disagreements over whether an accident is a component failure accident or a system accident (Hopkins, 1999, 2001). These disagreements will persist because NAT's intra-system levels, being arbitrary, lend themselves to varying interpretations. Going strictly by Perrow's interpretation, only a minuscule percentage of accidents would qualify as system accidents. This fact has led critics to raise legitimate questions about the theory's limited relevance (see Hopkins, 1999).

HRT

Around the time Perrow (1984) articulated his NAT, another stream of research emerged that prima facie had the potential to, irrespective of its avowed aim, falsify NAT's main premise. Scholars from the Berkeley campus of the University of California came together to study how organizations that operate complex hazardous technologies manage to remain accident-free for impressive lengths of time while simultaneously retaining their capacity to meet highly unpredictable and demanding production goals. If these scholars could identify organizational properties or processes that greatly mitigated the risks of operating in a complexly interactive and tightly coupled environment,

then they would apparently undermine NAT. The body of work produced by the Berkeley group, and other scholars who later enhanced the work, has been dubbed HRT.

HRT scholars have studied systems such as naval aircraft carriers (e.g. Rochlin et al., 1987), air traffic control systems (e.g. La Porte, 1988), nuclear power plants (e.g. Bourrier, 1996), submarines (e.g. Bierly & Spender, 1995), and space shuttles (e.g. Vaughan, 1996, 2005). Despite being highly diverse, the organizations studied have something in common. They are all complex technological systems that put a high premium on *reliability* since their operating environments seldom offer them a second chance. Given their focus, the first challenge before HRT scholars was to define reliability. It proved particularly troublesome to do so (see Hopkins, 2007; also see Wolf, 2001).

Although HRT scholars have abandoned attempts to explicitly define reliability, they appear to agree that reliability is the ability to maintain and execute error-free operations. Weick et al. (1999) report that HRT emphasizes the following conditions as being necessary, but not sufficient, for ensuring reliability: a strategic prioritization of safety, careful attention to design and procedures, a limited degree of trial-and-error learning, redundancy, decentralized decision making, continuous training often through simulation, and strong cultures that encourage vigilance and responsiveness to potential accidents. The authors then imply that to move closer towards attaining a sufficient condition of reliability, organizations must also become 'mindful'. Mindfulness entails a unique way of looking at the world. Weick and colleagues argue that organizational scholars might have erred in uncritically importing a notion of reliability from the engineering discipline that ignores processes of cognition.

The engineering notion equates reliability with lack of variance in performance. But Weick et al. (1999), citing Schulman's (1993) analysis of the Diablo Canyon nuclear power plant, argue that in the organizational context, reliability is not the outcome of organizational invariance, instead it results from the management of fluctuations. The emphasis thus shifts from stable routines to stable processes of cognition that must make sense of and reconcile the varying processes of production. Besides, they point out that routines are seldom stable in the sense that each time they are enacted they unfold in a slightly different manner, in a slightly different environment. Only an alert mind that is cognizant of the subtle differences can produce reliable outcomes in an organizational context is the conclusion.

Mindfulness is said to involve: preoccupation with failure (i.e. a suspicion of quiet periods); reluctance to simplify interpretations (i.e. a hesitation to generalize and make assumptions); sensitivity to operations (i.e. a high level of situational awareness of the big picture about what is

happening and what one might expect in the immediate future); commitment to resilience (i.e. a tendency to do whatever it takes to ride out a crisis); and under specification of structure. Weick and colleagues (1999) argue that by under specifying structure, organizations can encourage flexibility and a healthy disregard for formal hierarchy that allows decision-making to migrate with a problem. Migration of decision-making partly explains how highly reliable organizations (HRO) can meet the contradictory requirement of being simultaneously centralized and decentralized. In this context, Weick (1987) also observes that some organizations cope with conflicting requirements by granting considerable decision-making autonomy at lower levels and ensuring buy-in of centrally determined goals, decision premises, and assumptions.

Like mindfulness, *conceptual slack*, a term coined by Schulman (1993), relates to how firms might afford autonomy at lower levels in the face of centrally determined goals. Conceptual slack indicates a 'divergence in analytical perspective among members of an organization over theories, models, or causal assumptions pertaining to its technology or production processes' (p. 364). In an organizational context, the notion is a form of 'redundancy' – it adds to the variety of ways in which an organization can respond. While conceptual slack has the potential to create confusion, on the flip side, as happened in the case of the nuclear power plant that Schulman (1993) studied, it can lead to an engaged workforce that vigorously debates differing viewpoints and negotiates to arrive at an acceptable solution.

While some scholars have studied reliability and introduced notions such as conceptual slack and mindfulness, others have concentrated on non-reliability to draw lessons. For instance, Vaughan (2005) analyses NASA's inability to learn from previous experience and identifies the tendency to normalize deviation as a factor that contributes to accidents. Organizations unwittingly institutionalize practices that encourage gradual erosion of standards. An acceptable outcome of risky behaviour in the immediate past is allowed to set the expectation for risky behaviour on the next occasion. The changes in the harmful direction take place in such small increments and get injected into daily routines through *normalization of deviance* so surreptitiously that it is impossible to detect them until it is too late (Vaughan, 2005). Despite the fact that the reliability literature has introduced several constructs that may have relevance for non-HROs as well, it is puzzling that the area has been unable to connect its work with mainstream organization theory (Scott, 1994). We attribute this state of affairs to the high reliability area lacking a theoretical anchor.

Currently, HRT scholars can claim to have merely produced a list of factors associated with high reliability. Unless systematic empirical

comparisons with non-HROs are made, the area cannot make causality claims. Admittedly, Weick and colleagues (1999) have taken the first steps towards elaborating a predictive theory of reliability by explicating the cognitive processes that might be responsible for producing reliable structures, but we argue that they appear to have made questionable assumptions about the applicability of micro-level cognition processes at the macro-level (also see Weick & Sutcliffe, 2001). We therefore believe that more work is needed if HRT is to integrate with mainstream organization theory. We will later attempt to draw parallels between HRT and the proposed open systems perspective of accidents. In the next section, we highlight the salient aspects of the NAT-HRT debate.

The NAT-HRT debate

Despite differing motivations – HRT looks for organizational factors and processes that contribute to reliability, and NAT focuses on organizational properties that lead to accidents – we believe that both theories have similar implications for practice. NAT implies that organizations can lower the statistical probability of systems accidents (but never lower it to zero) by reducing their complexity and loosening the coupling amongst their subsystems. We argue that the initiatives identified by HRT – strategic concern for safety and safe design, redundancy, simultaneous centralization and decentralization, training, organizational learning, and mindfulness – can all be construed as attempts to either directly or indirectly address the challenges posed by complex interactions and tight coupling, the very dimensions central to NAT.

Although NAT seems to recommend that lowering the complexity levels of interactions in systems and decoupling them would lower the probability of accidents, the theory also points out that the very nature of the transformation processes that must be executed precludes this possibility. Furthermore, Perrow (1984) states that even when it is possible to tweak the two dimensions, financial considerations and pressures from the powerful elite interfere. Although NAT does not question the wisdom of doing everything possible to avoid system accidents, it asserts that every once in a while our best may not prove good enough. In contrast, HRT seems to predict safety for organizations that are totally committed to high-reliability practices (Rosa, 2005).

To the proponents of HRT, NAT's pessimism appears to stem not so much from its prediction that accidents are inevitable, as from its belief that it is Utopian of organizations to assume that their processes can potentially

remain foolproof for reasonable lengths of time. Perrow (1994) states that it is around this belief, amongst others, that NAT and HRT diverge.

Reflecting on the enhancements to NAT, Perrow (1994) concedes that he had failed to anticipate the emergence of HRT, which uses a different theoretical model of organizations. He also admits to having glossed over the consequences of his own model choice. His seminal work mentions only in passing that NAT is based on Garbage Can Theory (see Cohen et al., 1988). Perrow (1994) notes that the 'Garbage Can' approach, which draws attention to instability, ambiguity, misunderstanding, mis-learning, happenstance, confusion and so forth, is appropriate where there is high uncertainty about goals, structures, and processes. Given the complex environment that most organizations operate in and must interact with, Perrow claims that NAT is justified in eschewing the HRT-embraced notion of organizations being rational, stable, and closed systems.

HRT proponents could, however, argue that expecting internal processes of organizations to work efficiently in the face of external turbulence does not necessarily mean that one considers organizations to be closed, rational, and stable systems. Of particular importance to HRT is the need to build an organizational culture that puts safety first. It is important to note that culture building exercises are often influenced by external regulatory agencies and increasingly, at least in responsive democracies, by societal concerns. If anything, efforts to foster a high reliability culture are more consistent with an open rather than a closed systems perspective.

Sounding a dissenting note over the importance given to culture in the current context, Perrow (1999) states that a 'focus upon a culture of reliability is a luxury in the world of risky systems' (p. 360). He, instead, focuses on the use of power in the context of safety. But as Weick (2004: 31) forcefully argues, 'culture and power are not opposed explanations . . . culture shapes the way for power, defines power, is shaped by power, masks power, embodies power'. Thus it should be possible to reconcile the differences between HRT and NAT over their treatment of culture and power.

In our opinion, Perrow's most pessimistic observation is that the elite tend to wield power to insulate themselves from the dangers of systems accidents at the expense of exposing the less privileged. If his treatise about the powerful elite is accurate then indeed societies will seldom feel compelled to limit the exposure of all their elements to system accidents. Not surprisingly, Perrow (1999: 378) is dismayed that 'much of the work in the risk area today is systematically detoxing the power aspects' of his book. While we agree that scholars in the area have been guilty of paying too little attention to the role of power, we concur with Weick (2004) who observes that 'NAT is often a pretext for Perrow to make some larger points about which he feels

strongly' (p. 28). In other words, there appear to be no theoretical reasons for Perrow to pontificate about power.

NAT does make references to the systems perspective to justify discussions of power, but the rationale remains weak since at no stage Perrow (1999) makes any ontological commitment to systems theory. Although Perrow (2004) clarifies that NAT takes into account the wider context of a failure that ranges from mental models of individual operators, to group, organizational, and industry structures, he emphasizes that NAT is not based on 'the so called systems theory of the 1960-to-1980 period, grand, inclusive, and virtually circular' (p. 10). Thus we argue that claims that HRT treats organizations as closed systems are as untenable as the claims that NAT treats organizations as open systems. We will later return to this issue. For now, we discuss the 'centralization versus decentralization' dichotomy that will need to be addressed if one is to reconcile NAT with HRT.

NAT holds that systems cannot at once be decentralized and centralized. This is why the theory concludes that it may be best to abandon developing some of the more complex technologies. HRT, without necessarily disputing NAT's conclusion, states that it may be possible to ensure that decision-making migrates to where the action is (through under specification of structures and by affording conceptual slack) while ensuring buy-in for centrally determined goals. Thus the problem of meeting conflicting requirements may not be as intractable as NAT states. The same cannot be said of 'redundancy'.

The two theories hold very different views on the effects of redundancy. This may be attributed to HRT confusing reliability with safety. It is possible for a component to be reliably unsafe (see Rijpma, 1997; Wolf, 2001). For instance, Turner (1978) describes how a reliable distribution system proved very efficient at delivering contaminated fluids to British hospitals. In fact, one could argue that scholars could have used the term HSO (highly safe organizations), rather than HRO, to describe the organizations they have studied. HRT ignores the fact that redundancy (i.e. duplication in systems design to insure against failure) can carry costs by increasing complexities and opportunities for failure (Perrow, 1994), especially when the redundant component is not incorporated into the original design and 'added after problems are recognized' (Perrow, 1999: 368).

Consider how additional pilots in an aircraft could, somewhat counterintuitively, lower reliability. The senior pilots, being aware that there was someone else available to alert them in case of any emergency, could end up becoming careless. La Porte and Rochlin (1994) however, reject this criticism. They agree that incorporating redundancies could increase complexities and opportunities for failure, but argue that some organizations

are willing to 'carry out processes compensating both for the intrinsic hazard of their technical systems *and* the subsequent increase in "secondary complications"' (p. 223; emphasis in original). The auto-pilot feature, for instance, could be regarded as an additional redundancy against human failure. But one can readily imagine Perrow arguing that the auto-pilot feature would add another layer of complexity, and its failure could have disastrous consequences. What one ends up with in this argument is infinite regress with no resolution in sight.

To summarize: from its vantage point, HRT asserts that accidents, even in complex organizations that operate hazardous technologies, are avoidable if the organizations take enough pains to make the workplace safe. NAT, on the contrary, holds that regardless of the intensity of organizational efforts, accidents in complex and tightly coupled systems, because of system characteristics, are inevitable. To support their claim, HRT proponents point to accident-free environments that have existed for long periods. Expectedly, NAT subscribers point to the moment when accidents do eventually take place in order to support their inevitability claims. We argue that while the theories make different claims, they do not contradict each other. This is so because they look at the accident phenomenon at different points of time. We believe that time is a central but ignored factor in the NAT-HRT debate.

Resolution through a temporal dimension

Does time play a role in what is a stochastic process? Since NAT cannot predict as to when exactly the conditions become ripe for a system accident, one can only conclude that a system accident can occur at any time. But some reflection will reveal that the operative phrase in the previous sentence is 'become ripe'. This would point to an incubation period of sorts. Musing over why the US had not had more nuclear power plant accidents, Perrow (1984: 32) states:

> One answer is that the 'defense-in-depth' safety systems have worked, limiting the course of accidents . . . But a more accurate and less reassuring answer is that we simply have not given the nuclear power system a reasonable amount of *time* to disclose its potential.
>
> (emphasis added)

But are there any theoretical reasons for Perrow to implicate time?

One of the reviewers of this article pointed out that 'time of operation is an irrelevant dimension for NAT. Even if the system has run long enough for either experience to develop or for complacency to set in, it is the

complexity and coupling of the system, independent of time of operation, that creates the potential for the rare normal accident.' We acknowledge that NAT does not factor in time as a dimension. But this is not the same as saying that time has no bearing on a system from the point of time it is commissioned to the point of time it suffers a system accident.

In Perrow's quote cited above, he also refers to the possibility of 'defenses-in-depth' having worked to help the US avoid nuclear accidents. Interestingly, we would argue that time plays a crucial role in the collapse of 'defense-in-depth' systems as well. In his Swiss Cheese Model (SCM), Reason (1998) asks us to visualize various defence strategies as slices of cheese that are stacked alongside each other. The slices of cheese have holes, which signify weaknesses in the various defence strategies. In an ideal world, the slices would have no holes. But since perfection does not exist, slices invariably have one or more holes. The holes may close, shut, and move with the passage of time. On occasions, there is a chance, a very remote chance, of all the holes getting aligned and permitting one to look through the cheese stack. When this happens, it presents an opportunity for an accident trajectory to pass through or defeat the system's defence-in-depth. This is akin to 'time being ripe' in NAT.

Although NAT does not directly concern itself with events that contribute to the holes getting aligned, it is possible to reconcile NAT's language with that of Reason's (1998) SCM. NAT seems to imply that the statistical probability of the holes getting aligned increases as the complexity of interactions and coupling within a system increase. Safety features can come under enormous strain when tightly coupled systems must execute increasingly complex operations. Perrow's (1984) rich description of system accidents across industries highlights how system characteristics can render organizational effort and human intentions impotent. Albeit the examples narrated by Perrow constitute anecdotal evidence; collectively, they do make a persuasive case for the inevitability of system accidents. But the issue is not whether system accidents are inevitable or not. Inevitability is immaterial for practical purposes. Even if commissioning tightly coupled complex technological systems is fraught with risk, once such systems start operating, organizations can do little more than strive for error-free operations. The fundamental issue, from our perspective, is that HRT and NAT focus on entirely different stages of an organization's journey towards a system accident.

Figure 2 depicts a probable journey of an organization from the time of its inception to the time it meets with a system accident. In theory, the organization could suffer a system accident on the very first day of its operation (i.e. at point A). But in practice, such occurrences are rare. We

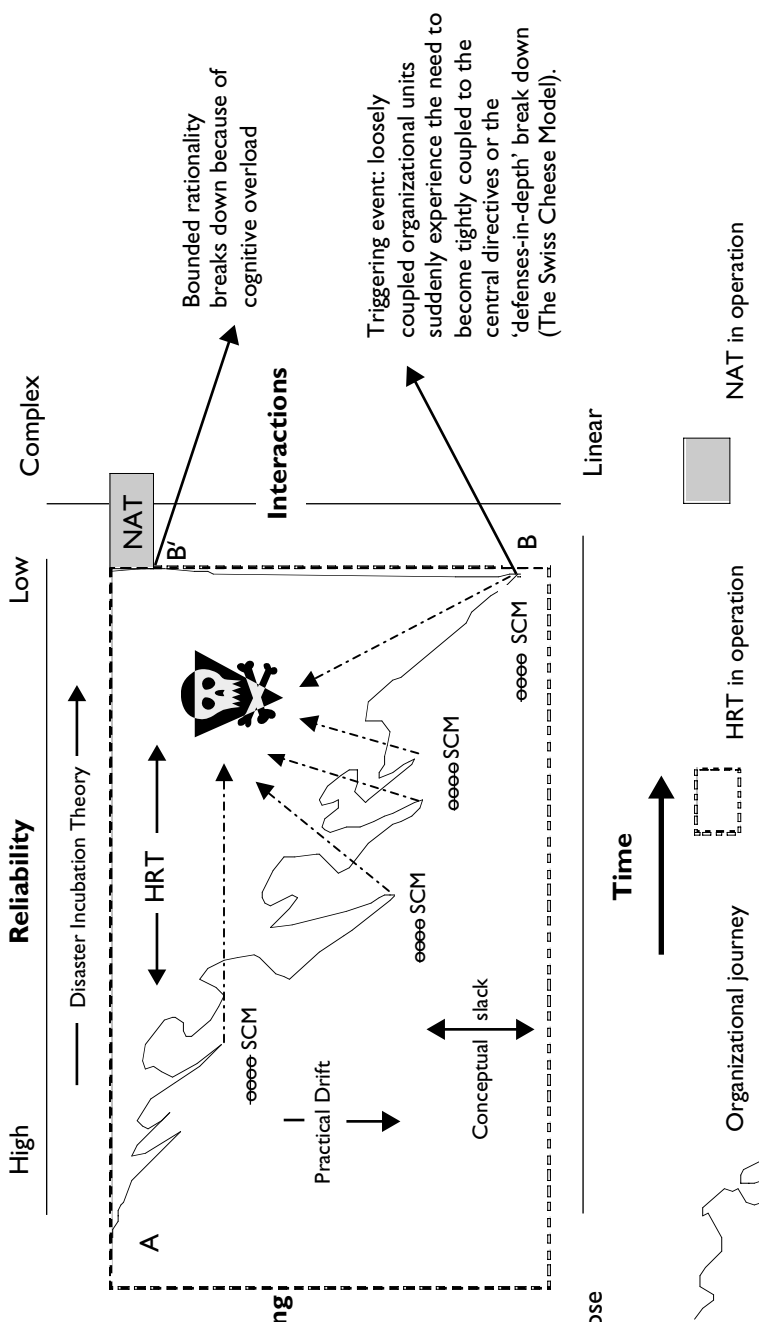

Figure 2 Journey towards a system accident: an example

argue that first day accidents, or accidents in the initial stages, are more likely to be component failure accidents attributable to engineering-related design errors, cost cutting measures, and pre-mature commissioning of projects. It is worth noting that Perrow (1994) emphasizes how difficult it is for all the failures to combine in a manner that defeats all the safety devices to trigger a system accident. Hence his observation about sufficient time not having elapsed to defeat the safety measures in the US nuclear power plants. Thus the journey from point A to point B could, in some instances, take decades.

Note that at point A in Figure 2, the system is shown as being tightly coupled and complexly interactive (i.e. the focal system meets the conditions necessary for a system accident). It would be reasonable to assume that when projects are commissioned after due diligence, there is expertise available to commence operations. We argue that as the implications of tight coupling and high complexity play themselves out, they impact the learning and experience curves of the system concerned. Irrespective of whether a system improves or deteriorates from a safety perspective, time, as we illustrate in Figure 2, is usually available as a resource to those concerned with safety.

No sooner than the focal system commences its journey at point A, it starts incubating for a disaster (Turner & Pidgeon, 1997). Disaster Incubation Theory (DIT) claims that with the passage of time, organizations start ignoring and misunderstanding danger signals; further, those with good safety records become complacent. Lackadaisical management and lack of information also combine to make matters worse – eventually, things give way and organizations suffer a major accident. Note the parallels between DIT and HRT. Complacency as described by DIT probably drives the tendency amongst organizations to normalize deviance (Vaughan, 2005) and to gradually migrate their activities towards the boundary of acceptable performance (Rasmussen, 1997). Also, note that just as HRT concentrates on what is right in accident-free organizations, DIT concentrates (in hindsight, one might add) on what goes wrong in the lead up to organizational accidents (Rijpma, 2003). Perrow (1984) however, argues that it is pointless to talk about warning signals and possible interventions in hindsight. He asserts that it can be very difficult, if not impossible, to decipher the meaning of and attend to dozens of simultaneous signals against the background of noise and false alarms *before* an accident. A study by Snook (2000) examined precisely this issue.

Snook (2000) asked whether a particularly tragic accident in the US military could have been averted. On concluding his research, he answered the question with a 'yes and no'. Things were ignored, but no individual could be blamed for ignoring them. The case study offered a fine-grained

analysis of how HROs, owing to practical exigencies, could drift towards a potentially unstable state.

Adopting a grounded theory approach to study a US Air Force friendly-fire incident, Snook (2000) implicates individual, group, and organizational factors. He theorizes that most organizations, to begin with, are tightly coupled to cater to worst case scenarios. The scenarios, however, seem so remote that people gradually start ignoring them. At the local level, teams start adopting unauthorized practices to complete the task on hand as efficiently as possible. This *practical drift* – defined by Snook (2000: 24) as 'the slow steady uncoupling of local practice from written procedure' – leads to the focal system becoming loosely coupled. If by sheer chance the system becomes tightly coupled again (e.g. when the system must form an ad hoc team for a particular operation that requires knowledge of the formally articulated central logic), and if at that moment a failure occurs, then the chances of a system accident occurring increase exponentially. This is because by then the system has drifted so far from the original state that none amongst those remaining is capable of responding sensibly to the unfolding events.

Let us return to Figure 2. The figure shows the focal system getting loosely coupled through practical drift and reaching point B. At this point, it confronts a contingency that requires it to become tightly coupled. It has lost the capability to do so and cannot make sense of what is happening. Thus to the system, the complexity it must cope with is very high and it, consistent with NAT's prediction, meets with an accident. We wish to emphasize two issues in the context of socio-technical systems: i) being loosely or tightly coupled occurs in the mental realm; and ii) complexity is relative. On both these inter-related issues, NAT is not very clear. Figure 2 shows the focal system oscillating along the two vertical 'Coupling' and 'Interaction' axes. We argue that this happens because the levels of *conceptual slack* (Schulman, 1993) that organizations afford to their employees can vary. The most relevant measure of coupling in the socio-technical domain may well be the level of *conceptual slack*. The higher the levels of conceptual slack enjoyed, the looser the coupling and the greater the ability of the system to cope with complexity. This brings us to the second issue about complexity being relative.

Our explanation implies that complexity is high only because the focal system cannot make sense of the interactions and this eventually leads to cognitive overload. As Perrow (1984: 78) states, 'systems are not linear or complex, strictly speaking, only their interactions are'. When one factors in the human mind, as one must in the case of socio-technical systems, complexity becomes relative. Perrow (1984: 84) seems to recognize this when he

says, 'It is also true that a poorly trained or inexperienced operator may see a system as replete with unsuspected interactions or "traps", but after gaining experience may find it to be more linear . . .' Herein is the answer to the paradox of being simultaneously centralized and decentralized. Physical entities may not be able to cope with this paradox, but mental entities can. As we had mentioned in our critique, NAT needs to make room for the human mind.

One of the reviewers argued that:

> It is the virtue of NAT that the only assumption about humans is that they are boundedly rational, and thus there will be an inevitable error in the design, and comprehension in the face of complexity, and so on. But human errors are not at the heart of NAT, though they are for HRT. NAT is valuable because it focuses upon system characteristics rather than such things as operator errors that may be owing to lack of mindfulness, inadequate training, management's failure to pursue safety goals and all the other things that HRT is concerned with.

We agree with the reviewer, but at the same time argue that system characteristics are a function of human understanding. It is plausible that humans design inelegant complex systems when they do not fully understand the technologies and the underlying processes. In fact, Perrow (1984) states that technologies that populate Quadrant 2 in Figure 1 could eventually migrate to other quadrants. Air traffic control systems are said to be a point in case. Similarly, Perrow notes that technological processes involved in iron and steel production have become less complex as human understanding of the processes has improved.

Note that Perrow (1984) acknowledges that experience and training can help reduce complexity, thus NAT must also make room for HRT. So how, and where, does HRT fit in Figure 2? As per HRT, organizations improve their reliability through training, mindfulness, and so forth. Actions such as preventive maintenance, replacement of worn out parts, and technology upgrades can enhance the life of systems. Thus in a conceptual sense, as shown in Figure 2, a 'renewed' system can be shown as travelling back in time. Hence organizations are also shown as oscillating along the horizontal 'Reliability' and 'Time' axes, even though they inexorably drift towards an accident (as explained by DIT).

Notice that Figure 2 also incorporates Reason's SCM. This captures the stochastic nature of system accidents. We recognize that the probability of an accident can never be zero at any point in time. Figure 2 indicates that the likelihood of the holes aligning would perhaps increase during periods

of sharp discontinuity. It would be reasonable to believe that cultural issues come to the fore when work practices change abruptly and humans must reorient themselves. HRT would suggest that the holes rapidly close in *mindful* environments. Conversely, new ones would open, or existing ones widen, if safety concerns were to diminish for whatever reason. As per Figure 2, the organization could have met with a system accident at any time had the 'defenses-in-depth' been breached. The figure highlights four such occasions. In one sense, the focal system is 'lucky' to have drifted unscathed up to point B in its eventful journey.

One final issue about *practical drift* merits a discussion. Rijpma (2003) notes that Snook's analysis and description of the phenomenon 'underlines the importance of Sagan's extension to NAT: accidents are inevitable not because of the technological complexity, but because of more banal organizational, cultural, and economic reasons' (p. 43). We argue that Snook's (2000) analysis does not discount the importance of coupling and complexity. Instead, practical drift introduces a dynamic element that underscores the fact that organizations tend to travel (for banal reasons perhaps) along the coupling and interaction continuum.

Thus we claim that NAT and HRT are not incommensurate – they refer to the same phenomenon, but at different time frames. We believe that DIT and related notions such as *practical drift*, migration towards the boundaries of acceptable performance, and *normalization of deviance* can help describe a system's behaviour at different points of time. Figure 2 depicts the changing states of a system as it drifts and incubates towards an accident. The system is in perpetual danger of its 'defenses-in-depth' getting breached and must remain *mindful* at all times. HRT, in Figure 2, operates across levels (in that the theory implicates organizational culture, team dynamics, and individual mental models) until such time an accident occurs. It is only when a system accident occurs at point B′ that NAT becomes applicable. In other words, coupling and complexity thresholds are reached at point B′ that result in bounded rationality breaking down.[2]

With HRT becoming inoperative at point B, historical organizational practices cease being the unit of analysis. Instead, the focus shifts to a *situation* at a moment in time, wherein a fault is tackled in a complexly interactive and tightly coupled systemic environment by an agency that has become ill-prepared owing to the travails of everyday existence (as explained by *practical drift*) in a culture that is not as safety conscious or as reliable as it once was (as explained by DIT and other related notions). Although NAT and DIT both appear to suggest that an accident is inevitable, there is a difference. As Hopkins (1999) points out, in DIT, human beings and organizations are assumed to cause disaster and are accorded the power to intervene,

whereas in NAT, human beings are unable to intervene once the chain reaction is set in motion owing to interactive complexity and tight coupling. But more importantly, as we show in Figure 2, DIT, like HRT, refers to a period of time that precedes the brief moment that NAT concerns itself with.

We clarify that it is not our case that all system accidents are preceded by practical drift and the other mechanisms described above. Nonetheless, we argue that the various case studies of aircraft carriers, nuclear power plants and so forth that are available in the HRT literature pertain to what would be the period between points A and B in Figure 2. Similarly, the rich descriptions of system accidents across industries provided by Perrow (1984) in support of NAT pertain to what would be the brief period between B and B′. In this connection though, it bears reiterating that as per Figure 2 a system accident can occur at any point of time. However, to the extent that technology upgrades, training, and mindfulness in general can help systems better cope with complexity, HRT proponents could claim that it may be possible to postpone a system accident.[3] We suggest that the proponents of NAT and HRT have been talking past each other because the boundary conditions of the two theories, in terms of time, are clearly distinct from each other. In its current form, NAT predicts that all tightly coupled and complexly interactive systems would *eventually* meet with a system accident (and until they do not, it is only because the right moment has not arrived). It is hardly surprising that HRT proponents are not impressed when reminded that initiatives such as redundancy, training, and so forth prove effective *until* they fail. The problem, as we claimed in our introduction, is that NAT can always explain away its failure to predict a system accident.

The non-falsifiability problem is reflected in the nature of empirical work in the area (Rosa, 2005). Case studies pertain to either organizational accidents or to organizations that have not had accidents. Systematic comparisons are lacking. Not surprisingly, researchers can always find support for their respective theoretical positions as their sample has zero variability. Perrow (1994), for instance, asserts that a majority of the so-called HROs scrutinized by researchers have been safe over long periods only because the organizations in question have not been operating highly complex and tightly coupled technological systems. Similarly, HRT can explain away its failure to predict error-free operations by finding faults in hindsight and claiming that an accident took place because high-reliability conditions had been violated. This is precisely what Rijpma (2003) does when arguing that the friendly-fire accident analysed by Snook (2000) did not take place in an HRO as the subunits involved were not a well-knit team.

In the process of resolving the NAT-HRT debate, we might have added conceptual clarity and explained why the two theories appear to diverge, but

we have not done much to address the non-falsifiability problem. If anything, our resolution has brought the issue into sharper relief. Recognizing that any theory on accidents must be sensitive to criticisms regarding lack of falsifiability, we propose a new framework of accidents that is based on an interpretation of systemic properties. The view presented is different from the under-developed systems perspective present in NAT and HRT.

Interpreting the properties of open systems

Although Perrow (2004) dismisses systems theory of the 1960s vintage as being virtually circular, we believe that an insightful application of the open systems perspective can help refine NAT. In a generic sense, accidents may be seen as instances of unintended or uncontrolled energy releases. In the current context, the open systems perspective seems promising because it essentially conceptualizes organizations as entities that manipulate energy. Over two decades ago, Ashmos and Huber (1987) had lamented that authors claimed adherence to systems perspective for merely embracing the commonsense idea that external environments affected organizations; seldom did they purposefully design their studies around systemic properties.

Cognizant of Ashmos and Huber's criticism, we begin by discussing some of the properties of open systems[4] that are germane to the proposed view of accidents. A discussion on *permeable boundary*, *energy transformation*, *negentropy*, *homeostasis* and *requisite variety* follows. The property of *negentropy* is examined in detail since the extant systems literature overlooks a vital aspect of the property.

Permeable boundary

An open system is distinguished from its environment by an arbitrary boundary. These boundaries are permeable, indistinct, and 'dynamic rather than spatial' (Bertalanffy, 1972: 422). Even if boundaries are only a mental construct, they must be delineated if a systems perspective is to be applied (e.g. Leveson, 2004; Rasmussen, 1997); for without boundaries, the distinction between a focal system and its environment would disappear (Scott, 1992).

Energy transformation

Open systems receive inputs from the environment; they transform these inputs into outputs, and exchange their outputs for new inputs. The

permeability of the boundary facilitates such an exchange. If the outputs of a system do not satisfy the environment or create 'value' for the environment, the inputs eventually cease. In the case of a business organization, one can visualize how organizations transform raw materials (forms of inputs) into finished goods/services (outputs), which are exchanged for a fresh round of inputs. The input-transformation-output (I-T-O) cycle may be described as a dynamic process that involves conversion of energy from one form into another. Open systems thus create value through transforming energy via I-T-O processes. If a nuclear power plant is to be treated as an open system then the nuclear fuel, coolants, and so forth would all qualify as inputs and the electricity generated by the plant would be output.

Negentropy

It is one of the fundamental laws of nature that energy can neither be created nor destroyed – it can only be made to change its form. Whenever energy is converted by a system from one form into another, there is 'wastage' or some loss of energy. In other words, 100 percent energy conversions are seldom possible. Some energy invariably escapes in the energy conversion process in a manner that renders it non-usable by the system in question. More damagingly, not all of the unusable energy escapes to the external environment, some of it accumulates within the system itself – this accumulation of unusable energy within the parent system is a form of entropy. Thus one may define entropy as a measure of disorder or randomness in energy.[5] In well designed systems, a minimal amount of energy escapes as waste, and accumulation of unusable energy is miniscule. This concept is explained further with the help of an example.

An automobile's engine converts chemical energy of the petrol (input) into kinetic energy of the wheels (output), but in the process of doing so, it wastes some energy. Of course, the more efficient engines emit less heat, give better mileage, and are quieter. But even the best of engines cannot prevent wastage. The wastage takes two forms: while some energy escapes to the external environment (e.g. in the form of fumes and sound), the balance gets accumulated or dissipated within the system (e.g. as soot or as heat owing to friction amongst internal parts). The energy that does not escape constitutes entropy and its accumulation has serious consequences for the long-term health of the parent system. According to the second law of thermodynamics, entropy in any closed part of the universe tends to increase with the passage of time. However, open systems, till such time they are in existence, appear to defy the second law of thermodynamics because the amount of order in them always exceeds the amount of disorder. Thus

open systems are said to have negative entropy (i.e. they are said to be negentropic).

Open systems remain negentropic through constantly exchanging their outputs for inputs (Katz & Kahn, 1978) *and* by expelling whatever entropy accumulates within them while they are engaged in energy transfer (Schrödinger, 1944). Because an automobile engine can neither exchange its outputs for inputs, nor can it remove the soot that accumulates within it without external intervention, it is not an open system. It should, however, be noted that open systems appear to defy the second law of thermodynamics. In reality, no matter how hard they try, they cannot expel all the entropy that accumulates within them. Thus every time an I-T-O cycle gets executed; some amount of entropy gets accumulated within a system. Ageing thus may be described as a process of entropy accumulation.

Examples of entropy accumulation would include not only obvious examples like unacceptable wear-and-tear of machinery and exposure of the workforce to radiation in a nuclear power plant, but also negative affect experienced by the workforce during the value creation process. Demoralizing one's employees in the wake of one's I-T-O cycle is an example of entropy accumulation in the socio-technical domain. One could, in fact, capture some of the phenomena from the dark side of organizations (see Vaughan, 1999) as examples of entropy accumulation. Conversely, one could argue that an empowered workforce enjoying autonomy (and afforded conceptual slack), would be less susceptible to entropy accumulation.

The foregoing analysis brings the focus back on the workforce, including human operators, in the accident context. Our aim here is not to make a case to justify the tendency to assign blame on human operators. Rather, by emphasizing the central role of the operators, we argue that everything possible should be done to make things easier for them. This argument is consistent with the human factors and systems engineering movement (see Leveson, 2004; Rasmussen and Svedung, 2000).

Homeostasis

Open systems rely on feedback loops to maintain equilibrium with an ever-changing external environment. In every feedback loop, as the name suggests, information about the result of a transformation or an action is sent back to the system in the form of input data. If these new data facilitate and accelerate the transformation in the same direction as the preceding results, they are positive feedback – their effects are cumulative. If the new data produce a result in the opposite direction to previous results, they are negative feedback – their effects stabilize the system. Positive feedback loops left alone can lead only to the destruction of the system, through explosion

(e.g. hyperinflation) or through implosion (e.g. economic depression). The wild behaviour of positive loops – a veritable death wish – must be controlled by negative loops. This control is essential for a system to maintain itself over time. In a negative loop, every variation towards a plus triggers a correction towards the minus, and vice versa. There is tight control; the system oscillates around an ideal equilibrium that it never attains (deRosnay, 1997).

A thermostat uses negative feedback to attain its goal of maintaining a room's temperature within an acceptable range. Open systems such as human bodies maintain their temperatures and blood sugar content levels through a similar mechanism. Such self-regulating goal-seeking behaviour induced by negative feedback ensures survival of a system even as the system in question continues to grow. This property that maintains equilibrium and allows for stable expansion is called homeostasis.

Requisite variety

Systems evolve to become more complex. The highly complex sense organs and the nervous system of higher organisms have evolved from primitive nervous tissues. Katz and Kahn (1966) observed almost four decades ago that the number of medical specialists in the US outnumbered general practitioners. Since then, not only has the number of specialists mushroomed, but so has the number of medical specialties. This inexorable movement towards increasing differentiation or complexity can be explained by Ashby's (1956) law of requisite variety. Variety is the number of states in which a system can exist. An electric switch can occupy two states (on and off) and therefore has a variety of two; a consumer in today's market who is spoilt for choice has a variety that is enormous. Just as entropy is a measure of disorder, variety is a measure of complexity. A complicated system has a large variety, meaning it can occupy a large number of states. Ashby's law essentially claims that a system needs variety to combat variety. The law of requisite variety tells us that a system can insulate itself from the complexity of the external environment by making itself complex.

That one must embrace complexity to combat complexity is counter-intuitive and not always understood by organizations (see Heylighen & Joslyn, 2001). Citing Cooper's (1973) fascinating account of Apollo 13's aborted moon landing, Perrow (1984: 278) describes how NASA pressed into service:

> four complete teams, each with dozens of experts, available to staff its ground system on a 24-hour basis . . . (and) about forty experts to concentrate on working out solutions to get the astronauts home safely; they were free of routine flight management duties. In addition,

> almost every step they devised could be quickly and realistically tested in a very sophisticated simulator before it was tried out in the capsule . . .

Although it is inconceivable that such resources would ordinarily be available to high-risk systems, the Apollo 13 recovery effort demonstrates how NASA had to acquire requisite variety (i.e. become more complex) in order to successfully cope with a situation of unprecedented complexity. This concludes the discussion on relevant systemic properties. In the next section, we apply the insights gained from this discussion to reframe NAT so as to address its non-falsifiability problem.

NAT reframed

Organizations as open systems must receive, manipulate, and exchange energy to create value. Therefore, all accidents must essentially take place at some point during energy reception, manipulation, or exchange (i.e. at some point during an I-T-O cycle). NAT argues that all systems that transform raw materials in some fundamental way must, per force, be tightly coupled and complexly interactive. It is not as if organizations have a choice. The decision to transform raw materials brings control issues to the fore. It follows that the higher the ability to control transformation processes, the higher the safety levels in the workplace (Leveson, 2004; Rasmussen, 1997). Also, the greater the levels of energy used, discharged, or stored during transformation processes, the greater the potential for damage in the event of an accident, hence the higher the need for controlling such processes.

The above argument suggests that the level of interaction complexity and the degree of coupling in an organization is a function of the amounts of energy levels involved in the organization's transformation processes and the gaps in its knowledge about the processes. Thus one could argue that the coupling- and interaction-related independent variables of NAT are, in fact, dependant variables. Indeed, it might even be possible to reframe NAT as shown in Figure 3. Note that the implications of the reframed NAT with respect to organizational characteristics remain identical to what they were in Perrow's original formulation. Also, note that by incorporating levels of energy, the reframed NAT can factor in the damage potential to society in general and humans in particular. This is something NAT ignores.

The accident literature classifies human victims as first-party victims (operators); second-party victims (non-operating personnel or systems users such as passengers); third-party victims (innocent bystanders); and fourth-party victims (foetuses and future generations). Although Perrow discusses

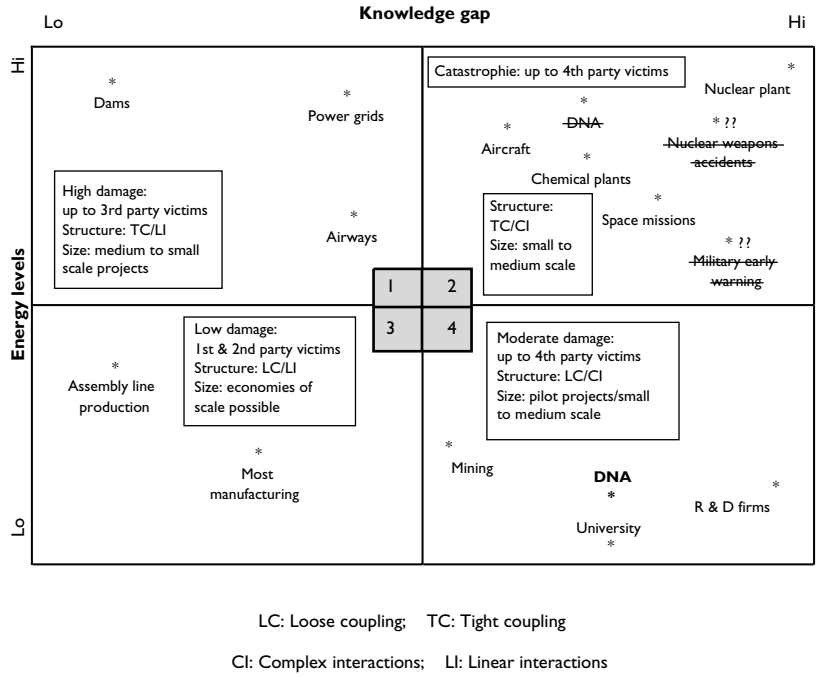

Figure 3 Reframed NAT: interpreting transformation processes

the potential for damage to human victims, he does so in the context of justifying his definition of accident, which focuses exclusively upon system characteristics. Perrow (1984: 66) states that a fixation with human victims could make us 'lose our focus upon the kinds of systems that business and government leaders decide should be built'. Our reframing of NAT shows that it may be possible to factor in damage to human victims without shifting one's focus from system properties.[6]

The damage potential of technologies and the difficulties associated with controlling and transforming large amounts of energy (particularly in instances where knowledge gaps are high) suggest that societies should in general eschew large-scale projects. As Figure 3 indicates, from a safety perspective, societies may exploit economies of scale only in the case of Quadrant 3 technologies. We point out that our reframing enables NAT to make a stronger case for exercising caution while commissioning high risk technologies. Also note the slight differences between the reframed NAT and Perrow's NAT over systems that populate Quadrant 2.

Military early warning systems and nuclear weapon systems do not involve transformation processes, and DNA recombinant technologies do not involve high levels of energy. We cannot possibly capture these three

systems under Quadrant 2 as Perrow does. Does this call into question the rationale behind our reframing? We think not. DNA recombinant technology might be highly complex, but there is no apparent reason as to why the organizations (or the socio-technical systems) handling the technology have to be tightly coupled themselves in the structural sense. As such, Quadrant 4 in Figure 3 would capture the knowledge intensive work involved in DNA recombinant systems. Knowledge workers generally work in autonomous teams (Alvesson, 2004). In other words, knowledge organizations are loosely coupled; but because the technologies they use are tightly coupled, knowledge work's impact can be global (think of the sub-prime mortgage crisis!).

While capturing various systems under the four quadrants, Perrow appears to have been influenced by technological characteristics as opposed to organizational characteristics. Insofar as military organizations are concerned, they are a special case because they do not adhere to democratic norms and at times it can be misleading to reach conclusions about their organizational properties during peace time operations (Schulman, 1993). Arguing that they do not lend themselves to being captured under any quadrant, we have not included them in Figure 3 in our reframed NAT.

The reframed NAT needed to replace interaction and coupling dimensions in order to address the non-falsifiability problem. Perrow (1994) concedes that NAT needs a metric that can measure the frequency with which errors might interact to defeat or bypass safety systems. But we believe that NAT precludes the development of such a metric because it insists that complex interactions cannot be anticipated, and in any case, are unfathomable. According to Perrow (1994), if processes are well-understood then they are probably not complexly interactive. Thus NAT perhaps cannot rely on measures of complexity for support. The theory needs to identify other variables if it is to offer testable propositions. We contend that knowledge gaps and the levels of energy involved during transformation processes are the two variables that can help NAT make falsifiable predictions. As argued earlier, the open systems perspective provides theoretical reasons to link these two variables to the probability of system accidents. Thus far, we have relied on the property of *energy transformation* to reframe NAT. In the next section, we draw from the other four properties to strengthen our case for developing an open systems view of accidents.

A call for developing an opens systems view of accidents

Developing a systems view of accidents will no doubt need further theoretical work and empirical testing. A starting point could be to theoretically

develop and test the relationship between levels of energy manipulated and the occurrence of accidents in general and system accidents in particular.[7] Gaps in knowledge could be measured by seeking the opinions of scientists, managers, and human operators. Although one could link coupling and interaction dimensions with gaps in knowledge, developing a metric for the two dimensions, as stated earlier, is likely to prove challenging. However, if such a metric is considered indispensable, then scholars could apply insights from open systems theory and measure complexity by quantifying the number of major interfaces involved in the energy transformation processes.

Open systems exchange their inputs for outputs through permeable boundaries at 'interfaces'. All interfaces engaged in energy exchange need to interpret feedback loops in order to ensure that adjustments are made in the correct direction to maintain stability (refer back to the property of *homeostasis*). In the context of system accidents, the interfaces of salience are those that are involved in major energy exchange processes. Arguably, lowering the number of major interfaces in a system would lower the statistical probability of systems accidents because fewer interfaces would make the system less complex. However, since organizations operate in an increasingly complex environment, they must continually, as dictated by *requisite variety*, increase their own complexity. Initiatives that can help organizations increase variety include, but are not limited to, raising ad hoc teams, forming committees, re-training and selecting employees with non-typical skills, and encouraging job rotation (Weick et al., 1999).

Paradoxically, the need to introduce new interfaces in order to survive in a complex environment carries with it the risks of a system accident. But when one factors in the human mind, one need not increase a system's structural complexity in order to increase its variety. As described by Schulman (1993), and as supported by the Apollo 13 example cited earlier, a work force enjoying *conceptual slack* can contribute to the requisite variety levels in the mental realm and enable organizations to cope with complexity.[8] A metric that measures requisite variety in the mental realm is likely to prove particularly useful in the context of system accidents.

The need to maintain requisite variety makes it imperative for organizations to ensure that their managers and technical personnel across levels are capable of coping with the level of complexity that they are likely to encounter in worst case scenarios. This conclusion resonates with the work of Jaques and Cason (1994) that equates human capability to information processing abilities and recommends matching human capability with expected complexity levels at a given hierarchical level. Although developed in the context of comparative managerial worth, scholars could consider using the notion of human capability to predict the quality of human

responses at critical interfaces in the event of an accident. Figure 2 predicts a systems accident when bounded rationality breaks down – or in other words, when personnel at interfaces fail to cope with complexity and decipher feedback loops. Organizations, to some extent, could lower the probability of bounded rationality breaking down by matching human capability with expected complexity levels.

Adding a redundancy to safeguard against accidents also amounts to increasing a system's complexity because it entails introducing a new interface. Hence, redundancies can pose challenges to the workforce. The open systems perspective therefore concurs with the notion that redundancies can simultaneously increase and lower the probability of a system accident and one needs to evaluate the net risk to determine the efficacy of a redundancy (see Sagan, 1994). There are other parallels with NAT. Just as NAT highlights the tension between the need for centralization (brought about by tight coupling) and decentralization (brought about by unanticipated complexly interactive failures that require localized responses), the open systems perspective underlines the tension between the need to increase its own complexity (brought about by the need to maintain requisite variety) and the need to lower the number of interfaces (brought about by the fact that accidents often occur at interfaces as outputs and inputs get exchanged and as feedback loops are deciphered).

Interfaces (whether automated or human) can find it difficult to decipher feedback loops when entropy levels are high. The presence of entropy, or disorder, can make systems unpredictable. Entropy accumulation, as pointed out earlier, is inevitable – ageing involves entropy accumulation. The theoretical explanation for the gradual erosion of reliability as asserted in DIT (Turner & Pidgeon, 1997) may thus lie in the concept of gradual entropy accumulation. It should however be noted that open systems can remain negentropic for lengthy periods. Their ability to remain so can perhaps be taken as support for HRT. Organizations, as envisaged by HRT, retire legacy systems, retrain their employees, upgrade technology, carry out preventive maintenance, and so forth. Such initiatives can be framed as efforts undertaken by organizations to expel entropy. While these initiatives can certainly increase organizational longevity, eventually the second law of thermodynamics must prevail.

Further, some organizations can become more vulnerable to accidents when their efforts to discharge entropy are afoot (i.e. during off-nominal operations) because discharging entropy is essentially an unnatural act for organizations. Unlike natural systems, which can expel entropy through sweating, radiation, and so forth, social systems do not have an inbuilt mechanism to discharge entropy. Again, one may be able to test through

secondary data whether organizations become more vulnerable as predicted by open systems theory.

Organizations can also become vulnerable to accidents during periods of excessive growth. As a system grows, it moves away from its stable state and the probability of its feedback loops breaking down increases. The system usually struggles to make sense of what for it is a new state of equilibrium. Further, periods of excessive growth may often call for quick responses. It is plausible that loose coupling amongst sub-systems during periods of growth aids improvisation and enables prompt responses. However, loose coupling could at the same time increase the propensity amongst sub-systems to pull in different directions and interpret feedback loops locally at the expense of the parent system. To the extent that periods of growth require greater buy-in from sub-systems, organizations need to be tightly coupled.

Excessive growth can thus place conflicting requirements on organizations. There is a need for organizations to be loosely coupled so that they may respond quickly to events in their new state of equilibrium, but to keep the larger picture in mind, the organizations also need to be tightly coupled. Open systems theory thus indirectly supports NAT, which identifies this need through a different premise but concludes that organizations cannot meet it. HRT scholars have however suggested several measures to cope with the centralization-decentralization tension. As we have argued, the concept of mindfulness (Weick et al., 1999) that subsumes the notion of conceptual slack holds promise in this context. Questionnaires developed by Weick and Sutcliffe (2001) to help organizations gauge their mindfulness could inform future empirical studies in the area.

A cross-comparative study of organizations within the same industry could be conducted to test whether greater growth is associated with greater probability of system accidents. Also, fine-grained qualitative inquiry into whether and how human responses lead the system to get tightly coupled could offer insights into the role of humans in systems accidents. From the above discussion, it is clear that the open systems view of accidents has the potential to theoretically account for the apparently disparate accident-related constructs and help advance the area.

Implications and conclusions

In this article, we argued that HRT and NAT are not incommensurate with each other – they merely look at the same phenomenon at different points of time. DIT (Turner & Pidgeon, 1997), and notions such as conceptual slack

(Schulman, 1993), normalization of deviance (Vaughan, 2005), and practical drift (Snook, 2000) were used to resolve the NAT-HRT debate. Nonetheless, the resolution did not circumvent the non-falsifiability problem inherent in NAT and HRT. To address this problem, we invoked the open systems perspective and discussed five systemic properties. In particular, we applied the property of energy transformation to reframe NAT in a manner that obviated the need to measure the levels of complexity and coupling, and enabled the theory to account for the role of human operators and factor in the damage potential to humans. The reframed NAT recommends commissioning small projects particularly when transformation processes involve high energy levels and are poorly understood.

Finally, we offered pointers on how the area might be advanced through applying the insights gained from open system theory. We argued that the open systems perspective could account for the conclusions reached by NAT, HRT, and DIT and indirectly explain Snook's (2000) notion of practical drift. Systemic properties suggest that organizations should exercise caution during the unnatural act of entropy expulsion, and as they grow and move away from their stable state. Our discussion also brought to the fore how organizations might, without increasing their structural complexity, increase their requisite variety. The HRT literature suggests that they can do so through encouraging their employees to share and challenge existing mental models. We believe that our effort reconciles the viewpoints contained in NAT and HRT and makes a case for developing an open systems view of accidents. We exhort our colleagues to respond to our call.

Acknowledgement

The authors are grateful to the guest editors, Nick Turner and Garry Gray, and the three anonymous reviewers; their guidance and feedback proved invaluable in helping the authors to improve the quality of this article.

Notes

1 Perrow appears to ignore the fact that military early warning and nuclear weapons do not transform raw materials in any fundamental way.

2 The problem, as we have pointed out, is that NAT in its current form prevents one from determining what the threshold levels might be. This makes the theory untestable.

3 As was pointed out earlier, one of Perrow's main motivations was to make a case for abandoning high risk technologies given the inevitability of system accidents. HRT is essentially silent on this issue. Nonetheless, La Porte (1994: 210) points out

that HRO scholars are not 'in the business of showing operators and manager how they could be perfect – and therefore feel they could further deploy hazardous systems supposing that they could be run with minimum risk'. He, in fact, insists that HRT sounds a note of caution by highlighting the great difficulties and costs associated with ensuring high reliability.

4 In his seminal article, Boulding (1956) presented a hierarchy of classifying systems based on increasing complexity that extended across nine levels. The first three levels, labelled frameworks, clockworks, and thermostats, comprised closed systems. The next six – cells, plants, animals, human beings, social systems, and transcendental systems – were said to be open systems. Transcendental systems represented the unknowable and were included by Boulding to cater for future advances.

5 Entropy in cybernetics, the science of control and communication in machine and animals, is taken to mean uncertainty or 'ignorance'; conversely, it is held that information fights entropy.

6 The aim is not re-invent the wheel. We are aware of hazard classification schemes such as MIL-STD-82D available at [http://safetycenter.navy.mil/instructions/osh/milstd882d.pdf] (as on 1 July 2008). The terminology of the classification scheme tentatively suggested in Figure 3 could be reconciled with extant schemes. Our main point is that any systems view of accidents can, and should, factor in human operators and victims in a more explicit manner.

7 Regardless of the form of energy consumed – thermal, nuclear, electrical, chemical (potential or kinetic) – it is possible to measure energy in a manner that enables comparisons across systems.

8 One of the reviewers felt that we were perhaps overstating the importance of conceptual slack. We, however, maintain that it is a construct that accurately reflects how mental entities display requisite variety.

References

Alvesson, M. *Knowledge work and knowledge-intensive firms*. Oxford: Oxford University Press, 2004.

Ashby, R.W. *An introduction to cybernetics*. London: Chapman & Hall, 1956.

Ashmos, D.P. & Huber, G.P. The systems paradigm in organizational theory: Correcting the record and suggesting the future. *Academy of Management Review*, 1987, 12, 607–21.

Bertalanffy, L. von. The history and status of general systems theory. *Academy of Management Journal*, 1972, *15*, 407–26.

Bierly, P.E. & Spender, J.C. Culture and high reliability organizations: The case of the nuclear submarine. *Journal of Management*, 1995, *21*, 639–56.

Boulding, K.E. General systems theory – the skeleton of science. *Management Science*, 1956, *2*, 197–208.

Bourrier, M. Organizing maintenance work at two nuclear power plants. *Journal of Contingencies and Crisis Management*, 1996, *4*, 104–12.

Cohen, M.D., March, J.G. & Olsen, J.P. A garbage can model of organizational choice. In J.G. March (Ed.), *Decisions and organizations*. London: Basil Blackwell, 1988, pp. 19–47.

Cooper, H.S.F. *Thirteen: The flight that failed*. New York: Dial Press, 1973.

deRosnay, J. Feedback. In F. Heylighen, C. Joslyn & V. Turchin (Eds), *Principia Cybernetica Web* (Principia Cybernetica, Brussels), 1997, available online at: [http://pespmc1.vub.ac.be/FEEDBACK.html], accessed 15 May 2007.

Heylighen, F. & Joslyn, C. The law of requisite variety. In F. Heylighen, C. Joslyn &

V. Turchin (Eds), *Principia Cybernetica Web* (Principia Cybernetica, Brussels), 1997, available online at: [http://pespmc1.vub.ac.be/FEEDBACK.html], accessed 15 May 2007.

Hopkins, A. The limits of normal accident theory. *Safety Science*, 1999, *32*, 93–102.

Hopkins, A. Was Three Mile Island a 'normal accident'? *Journal of Contingencies and Crisis Management*, 2001, *9*, 65–72.

Hopkins, A. The problem of defining high reliability organizations. Working paper 51, Canberra, ANU, 2007.

Jaques, E. & Cason, K. *Human capability: A study of individual potential and its application*. Falls Church, VA: Cason Hall & Co, 1994.

Katz, D. & Kahn, R.L. *The social psychology of organizations*. New York: John Wiley & Sons, 1966.

Katz, D. & Kahn, R.L. *The social psychology of organizations*, 2nd edn. New York: Wiley, 1978.

La Porte, T.R. The United States air traffic system: Increasing reliability in the midst of rapid growth. In R. Mayntz & T. Hughes (Eds), *The development of large scale systems*. Boulder, CO: Westview Press, 1988, pp. 215–44.

La Porte, T.R. A strawman speaks up: Comments on *The Limits of Safety*. *Journal of Contingencies and Crisis Management*, 1994, *2*, 207–11.

La Porte, T.R. & Rochlin, G. A rejoinder to Perrow. *Journal of Contingencies and Crisis Management*, 1994, 2, 221–7.

Leveson, N.G. A new accident model for engineering safer systems. *Safety Science*, 2004, *42*, 237–70.

Perrow, C. Normal accident at Three Mile Island. *Society*, 1981, *18*, 17–26.

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

Perrow, C. The limits of safety: The enhancement of a theory of accidents. *Journal of Contingencies and Crisis Management*, 1994, *2*, 212–20.

Perrow, C. *Normal accidents: Living with high risk technologies*, 2nd edn. Princeton, NJ: Princeton University Press, 1999.

Perrow, C. A personal note on normal accidents. *Organization & Environment*, 2004, *17*, 9–14.

Rasmussen, J. Risk management in a dynamic society: A modelling problem. *Safety Science*, 1997, *27*, 183–213.

Rasmussen, J. & Svedung, I. Proactive risk management in a dynamic society. Swedish Rescue Services Agency, 2000.

Reason, J. Achieving a safe culture: Theory and practice. *Work & Stress*, 1998, *12*, 293–306.

Rijpma, J.A. Complexity, tight-coupling and reliability: Connecting normal accidents theory and high reliability theory. *Journal of Contingencies and Crisis Management*, 1997, *5*, 15–23.

Rijpma, J.A. From deadlock to dead end: The normal accidents-high reliability debate revisited. *Journal of Contingencies and Crisis Management*, 2003, *11*, 37–45.

Rochlin, G.I., La Porte, T.R. & Roberts, K.H. The self-designing high-reliability organization: Aircraft carrier operations at sea. *Naval War College Review*, 1987, *40*, 76–90.

Rosa, E.A. Celebrating a citation classic – and more. *Organization & Environment*, 2005, *18*, 229–34.

Sagan, S.D. *The limits of safety: Organizations, accidents, and nuclear weapons*. Princeton, NJ: Princeton University Press, 1993.

Sagan, S.D. Toward a political theory of organizational reliability. *Journal of Contingencies and Crisis Management*, 1994, 2, 228–40.

Schrödinger, E. *What is life? The physical aspect of the living cell*. Cambridge: Cambridge University Press, 1944.

Schulman, P.R. The negotiated order of organizational reliability. *Administration and Society*, 1993, *25*, 353–72.
Scott, W.R. *Organizations: Rational, natural and open systems*. Englewood Cliffs, NJ: Prentice Hall, 1992.
Scott, W.R. Open peer commentaries on 'Accidents in high-risk systems'. *Technology Studies*, 1994, *1*, 23–5.
Snook, S.A. *Friendly fire*. Princeton, NJ: Princeton University Press, 2000.
Turner, B.A. *Man-made disasters*. London: Wykeham, 1978.
Turner, B.A. & Pidgeon, N.F. *Man-made disasters*, 2nd edn. Oxford: Butterworth-Heinemann, 1997.
Vaughan, D. *The Challenger launch decision*. Chicago, IL: University of Chicago Press, 1996.
Vaughan, D. The dark side of organizations: Mistake, misconduct, and disaster. *Annual Review Sociology*, 1999, *25*, 271–305.
Vaughan, D. System effects: On slippery slopes, repeating negative patterns, and learning from mistake? In W. Starbuck & F. Moshe (Eds), *Organization at the limit: Lessons from the Columbia disaster*. Oxford: Blackwell, 2005, pp. 41–59.
Weick, K.E. Organizational culture as a source of high reliability. *California Management Review*, 1987, *29*, 112–27.
Weick, K.E. Normal accident theory as frame, link, and provocation. *Organization & Environment*, 2004, *17*, 27–31.
Weick, K.E. & Sutcliffe, K.M. *Managing the unexpected: Assuring high performance in an age of complexity*. San Francisco, CA: Jossey-Bass, 2001.
Weick, K.E., Sutcliffe, K. & Obstfeld, D. Organizing for high reliability. *Research in Organizational Behavior*, 1999, *21*, 81–123.
Wolf, F.G. Operationalizing and testing normal accident theory in petrochemical plants and refineries. *Production and Operations Management*, 2001, *10*, 292–305.

Samir Shrivastava (PhD, Swinburne University of Technology, Australia) is a Lecturer in HRM & Organization Studies at Swinburne University of Technology. His primary research interests include systems thinking, organizational learning, managerial competence, and knowledge management. He has published in *Human Resource Management, Journal of Management Development* and *Australasian Marketing Journal*.
[E-mail: sshrivastava@swin.edu.au]

Karan Sonpar is a Lecturer of Management at University College Dublin, Ireland. He earned his PhD in 2008 from the University of Alberta, Canada. He was a Captain in the Indian Army prior to his career move to academia. His research interests include institutional theory, top managers, using qualitative methods for theory development, ethics and sociological approaches to risk. His work has been published in the *Journal of Management* and *Organizational Research Methods*.
[E-mail: karan.sonpar@ucd.ie]

Federica Pazzaglia is an Assistant Professor of Finance at the University of Manitoba in Canada. She earned her PhD in 2008 from the University of Alberta, Canada. Her main research interests are corporate governance, IPOs, diversification, ethics and risk.
[E-mail: pazzagli@cc.umanitoba.ca]