Working Paper 51

The Problem of Defining High Reliability Organisations

Andrew Hopkins

Professor of Sociology
Faculty of Arts and National Research Centre for OHS Regulation
Australian National University

January 2007
The National Research Centre for Occupational Health and Safety Regulation is a research centre within the Regulatory Institutions Network (RegNet), in the Research School of Social Sciences, at the Australian National University. The Centre is funded by the Office of the Australian Safety and Compensation Council (OASCC).

The aims of the Centre are to:

• conduct and facilitate high quality empirical and policy-focused research into OHS regulation, consistent with the National OHS Strategy;
• facilitate and promote groups of collaborating researchers to conduct empirical and policy-focused research into OHS regulation in each of the States and Territories;
• facilitate the integration of research into OHS regulation with research findings in other areas of regulation;
• produce regular reports on national and international developments in OHS regulation;
• develop the research skills of young OHS researchers; and
• assist in the development of the skills and capacities of staff of the NOHSC Office.

In order to achieve these aims the Centre undertakes research, supports a consortium of OHS regulation researchers in Australia, assists in the development of skills and capacities of staff of OASCC and collaborates in research programs with other Australian and international researchers and research centres.

Address for correspondence:
National Research Centre for Occupational Health and Safety Regulation
Regulatory Institutions Network (RegNet)
Second Floor, Coombs Extension
The Australian National University
Canberra ACT 0200
Ph (02) 6125 1514 Fax (02) 6125 1507
Email nrcohsr@anu.edu.au
Web http://ohs.anu.edu.au
THE PROBLEM OF DEFINING HIGH RELIABILITY ORGANISATIONS

Introduction

So-called “high reliability organisations” are frequently seen as models for others to follow. For instance, The Presidential Inquiry into the 2003 space shuttle Columbia disaster chose to evaluate NASA (the National Aeronautical and Space Administration) against the way a high reliability organisation would have performed, and it found that NASA fell a long way short of this standard. The report was a clear invitation to NASA, and more generally to all organisations using complex, risky technologies, to examine how high reliability organisations operate and to learn from them.

But if we are to learn from HROs (high reliability organisations) we must first be able to identify them. How do we know one when we see one? What are the criteria by which we can decide whether or not an organisation is an HRO? How do HROs differ from non-HROs? These are the questions that motivate this paper.

The original HRO research

The term HRO originated in the 1980s with a group of researchers at the Berkley campus of the University of California. They observed that there had been much research on organisations that had experienced disaster, but very little on organisations that, despite operating highly hazardous technologies, appeared to function without mishap. Here are their words:

(In 1984) three interdisciplinary faculty members at the University of California Berkley, joined forces because they were in a position to capitalise on an amazing and quite unique opportunity to examine, in depth, three organisations in which errors could have catastrophic consequences, but which seemed to manage their tasks well despite great technical complexity and pressure to perform. The three organisations... were not so much selected as offered to us, by a conjunction of personal contacts and previous research. Although the selection process was far from ‘objective’… the opportunity was not resistible (Mannarelli, 1996:84)

The three organisations concerned were first, the US air traffic control system, second, an electricity company operating both a nuclear power station (Diablo Canyon) and an electricity distribution system, and third, US Navy nuclear aircraft carrier operations. Nearly all the literature on HROs derives from these three research sites.

It was only after the work began that the researchers turned their attention to defining the term HRO more carefully. Perhaps the most considered attempt was the following:
Within the set of hazardous organisations there is a subset which has enjoyed a record of high safety over long periods of time. One can identify this subset by answering the question, “how many times could this organisation have failed, resulting in catastrophic consequences, that it did not?” If the answer is on the order of tens of thousands of times the organisation is “highly reliable” (Roberts, 1990:160).

Critics have pounced on an obvious problem with this definition. Many organisations could fail catastrophically at each and every second of the day. A quick calculation reveals that such an organisation could undergo a major accident every day and still count as highly reliable by the above definition. Indeed, say the critics, by this criterion “it is difficult to think of any low reliability organisations” (Marais et al, 2004:3).

Perhaps in order to avoid this kind of objection HRO researchers have also defined HROs in a less precise way, by saying that they are hazardous systems that produce “nearly accident free performance” (La Porte, 1996), or function in a “nearly error-free fashion” (La Porte and Consolini, 1998:848). Various statistics are then produced to support the claim that the three organisations studied are indeed HROs in this sense.

Consider, first, nuclear aircraft carriers. HRO researchers note that in one typical year the US Navy had 1.93 accidents involving fatalities or property damage of half a million dollars or more, for every 100,000 flight hours. However, such evidence does not mean much till put into comparative perspective. The researchers claim, without data, that compared to “ordinary organisational performance” this is “very safe” (Roberts, 1989:114). On the other hand critics claim, also without data, that “the accident rate in aircraft carrier landings (a commonly cited example of an HRO) is relatively high compared to many other high-risk activities” (Marais et al, 2004:3). These differing interpretations indicate just of difficult it is to use statistics to support any claim of “near accident free performance”.

Second, in relation to the electricity company, HRO researchers cite various figures demonstrating the reliability of the electricity supply to consumers. In one period, for instance, the supply was “99.965% reliable in terms of outages” (Roberts, 1989:114). The researchers also refer to the reliability of the company’s nuclear reactors, noting that in 1991 one of these power plant set a reliability record by running for 481 consecutive days without interruption (Klein, 1995:777).

The problem with these data, however, is that reliability, especially reliability of supply, is not always equivalent to safety. Indeed the two may pull in opposite directions. Reliability of supply over a particular period may be achieved by running plant in a dangerous way, and conversely, safety may sometimes depend on shutting off supply. The fact that the nuclear power station achieved a record run without interruption is not necessarily an indicator of how safely it was being run.

Third, in relation to Air Traffic Control, HRO researchers make statements of the following type:
“While there were 44,450 highway fatalities in 1990 (in the US) there were 827 aviation accidents. In that same year air route traffic control centres handled 37 (million) aircraft and airport towers handled 63 (million) aircraft” (Roberts, 1993:169).

It is difficult to draw any conclusions about air safety from this statement, as a moment’s reflection will make clear. More importantly, it tells us nothing about the safety of the ATC (air traffic control) system, since many of the 827 aviation accidents referred to above would be outside controlled airspace and would have nothing to do with ATC.

Far more relevant is the observation that “in the past ten years there has been no instance of a mid-air collision when both planes were under positive control” (Roberts, 1989:114). This is obviously a carefully worded statement and one is left wondering about the possibility of collisions where only one plane was “under positive control”. Nevertheless, this record supports the status of the US ATC system in an HRO, in that ATC has been accident-free, not just near accident free, for at least ten years. Of course this conclusion leaves open the question of how many mid-air collisions it would take to undermine ATC’s status as an HRO. European ATC was responsible for a disastrous midair collision in 2002 with large scale loss of life (BFU, 2004). If such an event occurred in the US, would ATC still be regarded as an HRO?

It is clear from this discussion that most of the figures provided by the HRO theorists cannot readily be used to justify their decisions to study these particular organisations. They lend plausibility to those decisions once made, but they do not provide the kind of criteria that would enable us to identify other HROs.

Although HRO theorists deploy statistics in the manner discussed above, they do acknowledge that these statistics are not definitive. In an article devoted specifically to this question of definition, Rouchlin (1993:17) admits that “no truly objective measure is possible”. He goes on:

“What distinguishes reliability enhancing organisations is not their absolute error or accident rate, but their effective management of innately risky technologies… There is, therefore, no a priori way to evaluate … the mathematical or statistical performance of the organisation … relative to any theoretical optimal condition.”

Here, then, is a very explicit statement that, strictly speaking, statistical data are of no use in identifying additional HROs beyond those that have already been studied.

Another HRO researcher candidly admits that the three organisations studied cannot be assumed to be a random sample and indeed “no one now knows what the population of HROs might be” (La Porte, 1996:69).

Comments that HRO theorists have made about NASA dramatically demonstrate their uncertainty over the HRO status of organisations beyond the original three. In 1989,
three years after the Challenger space shuttle disaster, NASA was described as an HRO (Roberts and Rousseau, 1989:133,137), while in 2001, before the Columbia space shuttle disaster, NASA was said not to exhibit the characteristics of an HRO (Roberts and Bea, 2001:179).

Another approach to identifying HROs

Fortunately, there is another approach to identifying HROs. Based on their empirical investigations, HRO theorists have made various lists of organisational characteristics that describe the three organisations they have studied. In principle, these lists of characteristics can be used to evaluate the extent to which other organisations qualify as high reliability organisations.

One set of characteristics that has been identified has to do with the nature of the technology in use: complex, demanding, tightly coupled (in the sense that sequences of events are rapid and difficult to interrupt) and highly hazardous. Interestingly, HRO theorists have sometimes used this set of technological characteristics to rule out whole industries as potential locations in which HROs might be found. So for example, Rochlin (1993:22) argues that no organisation in the railroad industry can be an HRO, mainly on the grounds that the technology is “reasonably simple and straightforward” and the hazards are considered by the public to be “self-limiting”. Similarly, Roberts and Rousseau argue that petroleum refineries cannot be HROs because

“these and many other continuous processing facilities do not specifically involve compressed time frames or simultaneously critical outcomes. The technology itself has a high degree of predictability, unlike that found in high-reliability organisations” (Roberts, & Rousseau, 1989:133).

This is bad news for organisations in these industries that have been influenced by some of the HRO literature and are aspiring to join the ranks of HROs. Projects have been undertaken in the oil and gas industry to encourage the growth of HRO modes of operation, but HRO theorists appear to be arguing that such projects are misconceived from the outset.

There is, however, an alternative way of describing the organisational characteristics of the three HROs studied that is potentially more useful than focusing on the nature of their technologies. The alternative is to examine how these organisations go about their business. This approach does not limit the possibility of achieving HRO status to certain industries and it broadens the range of organisations that might aspire to HRO status. It has been developed most extensively by Carl Weick, not one of the original HRO researchers. Perhaps because of this, he has been able to free himself more effectively from the original context and to extract from it the ideas of broadest relevance. In so doing he has reconceptualised HROs as “mindful” organisations. This is a useful change of terminology since it gets away from questions of just how safe does an organisation have to be before it can be considered an HRO, and it highlights instead what an organisation needs to do in order to reach the required end state. As Weick says:
“Other people who had examined these organisations were struck by their unique structural features. We saw something else: These organisations also think and act differently” (Weick and Sutcliffe, 2001, xiii).

The following passage both identifies the five characteristics that Weick has distilled from the HRO literature and demonstrates the conceptual shift from HRO to mindful organisation.

“HROs manage the unexpected through five processes: (1) preoccupation with failures rather than successes, (2) reluctance to simplify interpretations, (3) sensitivity to operations, (4) commitment to resilience and (5) deference to expertise, as exhibited by encouragement of a fluid decision-making system. Together these five processes produce a collective state of *mindfulness*” (Weick and Sutcliffe, 2001, v, emphasis in original)

Let us consider these five characteristics in more detail.

1  **Preoccupation with failures rather than successes**

HROs understand that long periods of success breed complacency and they are therefore wary of success. They are preoccupied with the possibility of failure. They hunt for lapses and errors, recognising that these may be the precursors to larger failures. They therefore have well developed systems for reporting near misses, process upsets and small and localised failures of all sorts.

Putting this another way, errors and other small failures amount to warnings of danger, indicators of how things might be about to go disastrously wrong. HROs are alert to the warnings of danger and operate on the basis that if warnings are identified and acted on, disaster can be averted.

Critics, however, have argued that the warning signs are only obvious in retrospect and that it is often not possible to discern their significance beforehand. The point is often put in terms of the signal/noise metaphor. For instance, Perrow has argued that although there were warnings prior to the near disaster at Three Mile Island nuclear power station in 1979, it would have been impossible to distinguish signal from noise beforehand. “Signals are simply viewed as background noise (he says) until their meaning is disclosed by an accident” (1982:175).

Weick alludes to this criticism when he writes: “some experts argue that it is impossible to anticipate the unexpected both because there are almost an infinite number of weak signals in the environment and because the ability to pick up these weak signals is far beyond the existing technological capabilities of most organisations” (Weick and Sutcliffe, 2001:53). In his view, however, the evidence does not support these critics.
To take a particular example, the warnings prior to the near meltdown of the Three Mile Island nuclear power station in 1979 were not weak signals, lost in the background noise; they were explicit letters and memos from engineers foreshadowing exactly what happened (Hopkins, 2001). They were ignored, not because they were indistinguishable from noise, but because the organisations concerned had no capacity to listen. (It was the experience of Three Mile Island which transformed nuclear power stations, at least in the US, into High Reliability Organisations. See Rees, 1994)

The signal/noise metaphor is central in the analysis of HROs. “The key difference between HROs and other organisations in managing the unexpected often occurs at the earliest stages, when the unexpected may give off only weak signals of trouble. The overwhelming tendency is to respond to weak signals with a weak response. Mindfulness preserves the capability to see the significant meaning of weak signals and to give strong responses to weak signals” (Weick and Sutcliffe, 2001:3-4).

Warning signs are usually ambiguous and may well have innocent or unproblematic explanations. The important point is not to default to the assumption of normalcy but to investigate the signals which are appearing until they are either demonstrated to have an innocent explanation or, alternatively, are confirmed as unambiguous indicators of danger. This is exactly what mindful organisations do. “Mindfulness involves interpretative work directed at weak signals” (Weick, Sutcliffe, Oldfield, 1999:90). It is the interpretive work which reveals their true significance.

Before moving on it should be noted that there is another line of research that converges on this same conclusion, namely, research on organisational safety cultures. Reason (1997) identifies various aspects of a developed safety culture, but above all else, he says, a safety culture is a reporting culture, in which people are prepared to report errors, near-misses, unsafe conditions, inappropriate procedures and any other concerns they may have about safety. These are the warning signs of ways in which things might go disastrously wrong. Of course these reports will be to no avail unless the organisation has some way of analysing and responding to them, or to use Weick’s term, doing the interpretative work. Both the HRO and safety culture research, therefore converge on the need to identify warning signs, analyse their significance and act on the analysis.

Neither the HRO researchers (Weick and his associates) nor those working on safety culture (Reason and his followers) have devoted much attention to the issue of how to distinguish signal from noise, that is, how to decide whether reported events are insignificant glitches or symptoms of deeper and potentially disastrous problems. Given the prevalence of the signal/noise metaphor and the general scepticism about the possibility of distinguishing between the two beforehand, there is surprisingly little social science literature on this issue. One of the aims of the broader project of which this paper is a part is shed light on just how an HRO goes about analysing and prioritizing the warning signs it receives.

2 Reluctance to simplify
All organisations must simplify the data which confront them in order to make decisions and move forward. Simplification means discarding some information as unimportant or irrelevant. But this is inherently dangerous, for the discarded information may be the very information necessary to avert disaster. “Simplifications increase the likelihood of eventual surprise” (Weick and Sutcliffe, 2001:94). HROs are therefore reluctant to discard information. HROs take deliberate steps to create more complete and nuanced pictures. They simplify less and see more. “...they position themselves to see as much as possible” (Weick and Sutcliffe, 2001:11). They socialise their workforces to notice more and they employ more people whose job it is to explore complexity and to double check on claims of competency and of success. Cost cutting organisations regard such people as redundant and work on the assumption that redundancy is the enemy of efficiency. HROs treat redundancy as vital for the collection and interpretation of information that is necessary to avert disaster.

Lawson (2001:125) puts this another way:

“Organisational slack, in terms of time and human resources that are not constantly subject to measures of short-term efficiency, is important for organisations coping with the challenges of the 21st century”

For Lawson, slack can be defined as “the pool of resources in an organisation that is in excess of the minimum necessary to produce a given level of organisational output” (2001:126). She goes on, “learning organisations require slack in the form of time to develop, and time for learning must be part of the organisation’s design” (2001:131).

The essential point here is that this is a matter of organisational design. In practice this means that there must be staff whose exclusive job is the collection and analysis of relevant information. Indeed HROs employ whole departments of people to carry out this function. These are the people who do the analysis of weak signals and determine the significance of warning signs that are picked up. In short, HROs embrace the principle of organisational slack as a vital ingredient of organisational mindfulness.

3 Sensitivity to operations

A third feature of HROs is that their front line operators strive to maintain situational awareness, or sensitivity to operations, that is, they strive to remain as aware as possible of the current state of operations. Moreover, they strive to understand the implications of the present situation for future functioning. All this presupposes front line operators who are highly informed about operations as a whole, about how operations can fail and about strategies for recovery.

The significance of this can be seen by contrasting this with the situation in many organisations where “silo” thinking prevails, that is, where employees operate within their own small sphere of influence without thought of the more remote impact of their activities. A culture of silos has been implicated in many organisational accidents (Hopkins, 2005, part 2)
It is not only front line operators who must be sensitive to operations. Managers must be sensitive to the experience of their front line operators, in particular, by encouraging them to report on their experiences. Weick and Sutcliffe note that “people who refuse to speak up out of fear enact a system that knows less than it needs to know to remain effective. People in HROs know that you can’t develop a big picture of operations if the symptoms of those operations are withheld” (2001:13). Here we see again the crucial importance of reporting systems backed up by an organisational capacity to learn from what is reported. Interestingly, as Weick observes “the big picture in HROs is less strategic and more situational than is true for most other organisations” (2001:13). More so than other organisations, HROs “are attentive to the front line, where the real work gets done” (2001:13).

4 Commitment to resilience

According to Weick, mindful organisations show a commitment to resilience, by which he means that they are not disabled by errors or crises but mobilise themselves in special ways when these events occur so as to be able to deal with them. “The signature of an HRO is not that it is error-free, but that errors don’t disable it” (2001:14). HROs work on the assumption that errors will occur and they put in place back-up systems to catch and correct errors. A commitment to resilience is actually a commitment to learn from error.

“To learn from error (as opposed to avoiding error altogether) and to implement that learning through fast negative feedback, which dampens oscillations, are at the forefront of operating resiliently (Weick & Sutcliffe, 2001:69, quoting Wildavsky)

The commitment to resilience, then, is clearly an aspect of the preoccupation with failure discussed above.

5 Deference to expertise

The final characteristic is deference to expertise. When operations are being carried out at very high tempo, decisions “migrate” to the people with the greatest expertise or knowledge about the events in question. These people may be relatively low in the hierarchy, but at such times more senior managers will defer to their expertise. Researchers have identified this as a consistent pattern in flight operations on aircraft carriers, for example. They note that even the lowest level seaman can abort a landing, without reference to higher authority. When the tempo returns to normal, the locus of decision making moves back up the hierarchy.

Although high reliability researchers have emphasized this final characteristic, critics have questioned it significance. They point out that decisions must inevitably be made by people at the front line in time-critical situations. In such situations there is no possibility of referring matters up the chain of command. However these front line decision makers are highly trained and they make their decisions in accordance with their training.
Typically, too, these decisions are in one direction only, in the case mentioned above, to abort landings, and the decision making process is relatively simple (Marais et al, 2004:8). Air traffic controllers routinely abort landings and order aircraft to go around for another attempt, but it is hard to see how this contributes to ATC’s HRO status in the United States.

The principle of deference to expertise has rather more substance in situations that are not quite so time critical. The decision to launch the Challenger space shuttle was made against the advice of the expert engineers. This was a case where decision making did not migrate to the experts; had it done so, seven lives would not have been lost. Or again, the Piper Alpha platform fire was fed by fuel from a neighbouring platform. Managers on this platform were aware of what was happening but did not shut down production because they had not been authorized to do so. As these examples make clear, the principle of transferring decision-making power to those who are most knowledgeable has much to be said for it, particularly where that decision is about whether to shut down or abort in the interests of safety.

But it is not clear just how generalisable this principle is. There is now much evidence that accidents may be the result of decisions made in many parts of an organisation by people who are unaware of the full implications of their decisions. As the critics have noted, “the type of bottom-up decentralized decision making advocated for HROs can lead to major accidents in complex socio-technical systems” (Marais et al, 2004, 9).

Whatever the judgment we end up making about this fifth characteristic of HROs, it is in some respects the odd man out. The other four characteristics hang together in an obvious way; they all in one way or another about organisational learning. The fifth characteristic is about the locus of decision making. Leaving aside this final characteristic, one thing is clear from Weick’s account: above all else HROs are learning organisations.

Implications

Weick’s elaboration of what it means to be an HRO or a mindful organisation can be treated as a definition: an HRO is an organisation that exhibits these five characteristics. Moreover, if we think in these terms it is clear that being an HRO is not an all or nothing matter. Organisations may exhibit the characteristics of an HRO to varying degrees. Indeed the concept of an HRO is best regarded as an ideal type, to use Max Weber’s famous concept (Bendix, 1966): real organisations may hope to approximate this ideal, but never to achieve it in its entirety, not even the organisations studied by the original HRO researchers. The advantage of this definition is that it gets away from disputes about whether an organisation is or is not an HRO.

The definition of HROs in terms of organisational characteristics suggests an obvious research strategy. The investigator can compare an organisation of interest to the ideal type and assess how well it measures up. As noted earlier, this was the strategy adopted by the Columbia Space Shuttle Accident Investigation Board. The research design
involved comparing the NASA culture revealed in the inquiry process with other theoretical models. In the Board’s words,

“To develop a thorough understanding of accident causes and risk, and to better understand the chain of events that led to the Columbia accident, the Board turned to the contemporary social science literature on accidents and risk and sought insights from experts in High Reliability, Normal Accident, and Organisational Theory… Insight from each figured prominently in the Board’s deliberations… The Board selected certain well–known traits from these models to use as a yardstick to assess the Space Shuttle Program, and found them particularly useful in shaping its views on whether NASA’s current organisation … is appropriate” (CAIB, 2003:180).

In fact the insights on which the Board ultimately relied came almost exclusively from High Reliability Theory and its conclusion was that NASA fell a long way short of the ideal.

More generally, Weick has developed check lists of questions that can be used to assess where organisations stand in relation to the various HRO characteristics (2001: 90, 95, 96, 100, 102, 104, 106, 108, 110). These are of value to any organisation seeking to operate more mindfully. Weick’s work is a practical application of HRO theory, a good example of theory in use. It offers a way forward to organisations that feel they have stagnated in their efforts to enhance safety.

There has been little new research on HROs since the initial investigations of the Berkley group in the 1980s. No doubt this is in part because of the definitional difficulties discussed above and the consequent difficulty of selecting new HROs for study. It is probably because of this definitional problem that the concept has been subtly transformed into a model, a yardstick, an ideal, against which real organisations can be measured.

References


Hopkins A (2005), Safety, Culture and Risk, CCH Sydney.


Rees, J (1994), Hostages of Each Other: The Transformation of Nuclear Safety Since Three Mile Island (Chicago, UP)


1 The researchers note that during a trip aboard the Enterprise in 1987 it took its 250,000th arrested landing without the kind of error in arrestor wire settings that could have led to the loss of an aircraft overboard
Such evidence of absolute error-free operation makes the claim to high reliability more plausible. On the other hand, the researchers also note that none of the six US nuclear aircraft carriers has ever been destroyed (Roberts and Gargano, 1990:147). The interpretation of this information depends crucially on the standard of comparison. It is rare for large ships to be destroyed in peace time and from this point of view, the fact that none of the nuclear aircraft carriers has been destroyed seems hardly remarkable.

Railways are another context in which reliability and safety can be in direct conflict. Insisting that trains run on time can sometimes be at the expense of safety (Hopkins, 2005, part 2).

Showing how organisations fall short of the HRO ideal is not only a research strategy; it is also a way of developing just what it means to be an HRO. Weick’s influential book is structured in just this way. He shows how the Union Pacific railway company in the US and the Moura coal mine in Queensland fell far short of the HRO ideal (Weick and Sutcliffe, 2001).