Military Risk Assessment: From Conventional Warfare to Counter Insurgency Operations

C.W. Johnson

8/1/2012

University of Glasgow Press, Glasgow, Scotland.

ISBN 978-0-85261-933-9

Contents

List of Figures

1 An Introduction to Military Risk Management

All human activities carry an element of risk. However, military operations typically involve greater risks than most other occupations. These risks arise from by the use of innovative technologies by young people in uncertain and changing environments against enemies who constantly adapt their tactics to exploit perceived vulnerabilities. Many military organizations advocate risk management to support strategic, tactical and operational decision making. The UK Ministry of Defence has established a joint risk management policy between the Chief of Defence Materiel and the UK's Chief Scientific Advisor. This is intended to ensure that risk management techniques are used across all phases of military procurements from conception through to decommissioning. For instance, the British Army's Joint Doctrine Publication (JDP) stresses that commanders must develop Operational Estimates that take into account the hazards to their own forces but that also consider opposition vulnerabilities. Risks should be taken when they are justified by the potential rewards or when a unit can 'stand' the potential consequences. The US Army's Composite Risk Management initiative is intended to help soldiers avoid hazards both while they are on duty and during their free time.

This complexity of military risk management can be illustrated by a river crossing scenario. In the civil version of this scenario, you are faced with a choice between crossing a fast flowing stream using a rotten tree trunk or by walking two kilometers down to a ford. One way of approaching this is to consider the state of the tree trunk; to assess the probability that it will break and that you might fall into the water. As part of this assessment process you must also consider the consequences, including your ability to safely swim to the shore if you did fall in. The military version of this scenario now assumes that you have to cross the river under enemy observation. You must assess the risk of crossing quickly by using the tree trunk, which would expose you to a short period of intense fire. Or you could decide to move down to the ford, with an increased amount of time exposed to lower intensity fire. The introduction of opposition forces into any risk assessment opens up new dimensions of complexity – in this case we have to consider risk exposure from the short exposed crossing or the longer walk to the fording point, which did not arise in the civil case.

The hazards associated with military operations are compounded by the problems of fatigue and of distributed communication across multiple units from different nations. The increasing need to satisfy military objectives and at the same time protect both the civil population and the local environment within recognized rules of engagement add pressures that seldom arise in other contexts. In consequence, an increasing number of military organizations have proposed risk management techniques as a means of helping personnel to anticipate the hazards that they face. This book uses operational experience gained from post action reviews and boards of inquiry to explain how these hazards arise. It also argues that we often place undue confidence in civilian risk management techniques as means of predicting and mitigating the hazards of military operations.

Risk can be described in terms of deviations from an expected outcome. These deviations can be beneficial or they can involve some form of cost. This book is concerned with adverse outcomes in military operations. These deviations include the failure to achieve some expected objective. They also include situations that result in unanticipated harm. For example, there is a risk that a military operation will fail to achieve its intended outcome. There is also a risk that an accident will compound this failure by injuring the soldiers involved in an operation.

The field of risk management has extended this high-level view to define risk in terms of the likelihood of a hazard leading to an adverse outcome multiplied by the consequence of that outcome. For instance, coalition forces are exposed to high levels of risk associated with the threat from Improvised Explosive Devices (IEDs) in many areas of Afghanistan. This risk is high because they are likely to encounter these devices and the increasing sophistication of their design has also increased the damage caused by IEDs. There is, however, a decreasing risk from these devices to UK forces in Northern Ireland – not because they would be any less lethal but because recent years have seen a reduction in the likelihood that these devices will be used.

		Outcome				
Likelihood		Severe	Major	Moderate	Minor	Negligible
	Almost certain	U	U	H	H	M
	Likely	U	H	H	M	M
	Possible	H	H	M	M	M
	Unlikely	H	M	M	M	A
	Rare	M	M	M	A	A

Fig. 1. Risk Assessment Matrix

(U = Unacceptable, H = High, M = Moderate, A = Acceptable)

Figure 1 formalizes some of these ideas. As can be seen, high likelihood events associated with relatively severe outcomes can be classified as unacceptable (U). In such circumstances, action must be taken to mitigate the risks. This can be done by reducing the likelihood of a hazard, for example be denying opportunities to plant IEDs, or by reducing the impact of the hazard, for instance by protecting troops in hardened vehicles. Other hazards in Figure 1 that are less likely to occur or that have more benign outcomes can be classed as acceptable (A). In these cases, mitigations are not needed. Other threats that fall between these two extremes must be ranked in terms of their consequences and likelihood so that resources are allocated in proportion to the perceived High (H) or Moderate (M) risks.

Risk assessments have guided management decision making and engineering practices across healthcare, aviation, electricity distribution, nuclear power generation etc. Within each of these industries, risk management has been integrated into international standards. For example, IEC 61508 governs the application of programmable systems in safety-critical process applications. DO-178B advocates the use of risk management during the development of avionics. ISO 27001 advocates risk assessment as a key stage in information security management. ISO 14971 describes the application of risk management to clinical devices. Common to all of these standards is the assumption that we can increase safety by mitigating the consequences or reducing the likelihood of hazards.

Many of the military mishaps described in subsequent chapters were caused by wider management processes rather than from problems in quantifying likelihood or consequence using assessment matrices. Risk management, therefore, refers to the processes involved in first identifying hazards then assessing their relative significance before implementing controls and monitoring outcomes. It is possible to identify a number of common processes that characterize risk management techniques:

1. Identify potential the hazards or threats that can lead to operational risks;

2. Conduct risk assessments to identify the consequences and likelihood of each threat, taking into account potential vulnerabilities;

3. Prioritize risks using techniques such as a risk assessment matrix, illustrated in Figure 1.

4. Identify controls to either mitigate the consequences or reduce the likelihood of any unacceptable risks;

5. Implement and monitor the effectiveness of those controls;

6. Go back to step 1.

Risk management is an iterative process. The implementation of mitigation measures can lead to new hazards. For example, Improvised Explosive Devices (IEDs) are a growing threat or hazard to many conventional land vehicles. In consequence, many armed forces rely on rotary-winged aircraft to reduce the vulnerability of land convoys. However, the increased tempo of helicopter operations has, in turn, led to an increase in the number of mishaps involving these aircraft, for instance during dusty, brown-out conditions. It is for this reason that risk management techniques encourage iterative monitoring as hazards change over time.

1.1 Risk Management Processes in MIL STD 882D

US Military Standard (MIL-STD) 882D provides guidance on an eight stage process for risk management [1]. Firstly, it is necessary to document the approach. Without appropriate documentation, it can be difficult to determine whether hazards have been adequately addressed. MIL-STD 882D also recommends the use of hazard tracking software to ensure that potential risks are not overlooked during the subsequent stages of development.

The second stage identifies potential system hazards. This process must be informed by operational expertise and by mishap reports. It must extend across the system lifecycle to consider hazards associated with installation and decommissioning. It is important to consider hardware and software failures. Analysts must also account for interactions between a military system, its users and their environment.

The third stage advocated within MIL-STD 882D assesses the risks associated with each hazard in terms of severity and probability, using a risk matrix similar to that presented in Figure 1. The latest draft standard goes into considerable detail over the need to assess the contribution that software makes to overall system risk. This poses particular problems because software is often application specific; previous failure rates cannot easily be used to assess the probability of future failures. In contrast, 882D focuses on software criticality, partly determined by the degree of control that code has upon system hardware. This criticality assessment is used as a measure of importance rather than more conventional risk metrics. Later sections of this chapter will return to the issue of software risk management in greater detail.

The fourth stage of MIL-STD 882D's systems approach to risk management identifies mitigation measures; "Risk mitigation is an iterative process for eliminating or reducing risk to the lowest acceptable level within the constraints of operational effectiveness and suitability, time, and cost" [1]. The guidance advocates a number of different techniques. These include:

- *Hazard elimination* through the development of appropriate designs. For example, the use of non-flammable materials can potentially eliminate the risk of fire;

- *Hazard reduction* is also supported by design changes that reduce the probability of a risk even though it cannot be eliminated. It is seldom possible to remove all flammable material from military systems;

- *Additional protection devices* may also be used to reduce the impact of a hazard once it has occurred. For instance, sprinkler systems can be introduced to suppress a potential fire;

- *Warning devices* provide a further means of mitigation. However, it is important to consider the impact of false alarms and of missed warnings;

- *Procedures and training* can be used to prepare staff for adverse events. For example, drills can be used to simulate the response to potential hazards.

The fifth stage involves the selection and implementation of appropriate mitigation techniques. This stage of the risk management process must account for the costs associated with different interventions. It must be supported by appropriate technical reviews that span each of the different development strands in large scale, military procurement projects. This is important because the introduction of risk mitigation techniques can inadvertently increase the hazards associated with other sub-systems. For instance, the costs associated with new protection devices in one area might reduce the budget available for the development of other application components.

The sixth stage of the MIL-STD 882D process is intended to verify that the intended risk reduction has been achieved. As we shall see, this raises a host of problems. For hazards with a relatively low likelihood it may not be possible to observe direct evidence that mitigation measures will continue to provide the anticipated level of protection. In such cases, monitoring and inspection must be supported by analysis and testing techniques.

The penultimate stage ensures that all stakeholders in a project understand and accept the residual risks after any mitigation has been implemented. In particular, the certifying or approving authority must recognise any remaining hazards. It is impossible to guarantee that any complex system is absolutely safe. 882D requires that an ‘end user’ representative is involved in this process. The acceptance of risk must, typically, be documented in a formal manner that recognises any associated constraints on the operating configuration or on the environment in which a system is used.

The eighth and final stage of the MIL-STD 882D system safety process focuses on the management of risk after deployment; “the life-cycle effort shall consider any changes to the interfaces, users, hardware and software, mishap data, mission(s) or profile, system health data, and similar concerns”. Each of these factors can trigger additional risk assessments following the same eight stage process, described above [1].

1.2 Risk Management Techniques

This book examines the challenges that arise when using risk management techniques, such as those described in MIL-STD 882D, to support military operations. The focus is less on particular methodologies. In contrast, the intention is to identify the limitations of existing approaches and provide case studies that might be used to guide the development of future techniques. In order to do this, it is first necessary to introduce civilian risk management tools, including Hazard and Operability Studies (HAZOPs) or Failure Modes, Effects and Criticality Analysis (FMECA). The intention is not to provide an exhaustive account. Brevity prevents a detailed introduction to the use of Fault Trees, Cause Consequence Diagrams etc. Instead, the aim is to illustrate common concepts and then move on to consider the operational, tactical and strategic issues that complicate military risk assessment.

1.2.1 HAZOPS

Hazard and Operability Studies (HAZOPs) helps analysts to identify potential hazards and then assess their potential impact. Although it was initially developed in the process industries [2, 3], it has subsequently been used by a number of armed forces. For example, the former UK Ministry of Defense (MoD) Defence Standard 00-58 supported the application of HAZOPS in software control systems [4]. The US Army has used HAZOPS to manage the risk associated with commissioning, operating and decommissioning a wide range of systems [5]. Although there are a number of more recent alternatives, HAZOPS continues to be used in many high-risk areas. For instance, it has been integrated into the US Army's risk management processes "to assess and manage hazards and ensure scientists, engineers, maintenance personnel, and Environmental Safety and Occupational Health personnel contribute to the safe operation of specific processes" associated with nanotechnology programmes [6].

HAZOPS provides tools that help to minimize the subjective biases that undermine risk assessments. These biases arise when individuals cannot agree on the hazards that might undermine the safety of a complex system. They are reduced in HAZOPS by ensuring that the analysis is conducted by multi-disciplinary teams that work together in order to validate risk assessments.

HAZOPS can only begin after developers have identified the major functions in a potential system. Functional block diagrams can them be used to map out the interactions between these components. Each function is associated with a set of requirements or intentions. For instance, a personal radio system should provide point to point communications over 500 square meters for 256 channels with up to 20 hours continual use. The HAZOPS team then goes on to consider what would happen if the component deviated from its intended operating parameters. They must also identify the causes for each possible deviation. The HAZOPS team must determine whether the consequences of any failure threaten safety. If this is the case then additional mitigations must be introduced.

The previous summary makes it clear that the value of any HAZOPS study is determined by the selection of appropriate requirements for each functional component. If a key requirement is omitted from the analysis then HAZOPS teams are unlikely to identify what would happen if the function failed. The success of the approach also depends upon the use of guidewords. These prompt teams to consider a range of hazards that might undermine particular requirements. If a guideword is overlooked then teams may not consider a particular pattern of failure for a particular component. The following list provides examples of common guidewords:

- *Before:* consider what would happen if a function occurred before the intended point in a sequence;
- *After:* consider what would happen if a function occurred after the intended point in a sequence;
- *Early:* consider what would happen if a function occurred before the intended time even though it might occur at the correct point in a sequence;
- *Late:* consider what would happen if a function occurred after the intended time even though it might occur at the correct point in a sequence;
- *No or Not:* consider what would happen if a function did not occur or completely failed to satisfy the intended requirements;
- *More:* consider what would happen if a function went beyond the intended limits of a particular requirement;
- *Less:* consider what would happen if a function fell below the intended range of a particular requirement;
- *As Well As:* consider what would happen if a function introduced a number of unintended side-effects;
- *Part Of:* consider what would happen if a function met some of the intended requirements but not all;
- *Reverse:* consider what would happen if a function satisfied the opposite of an intended requirement;
- *Other Than:* consider what would happen if another function was substitutes for the intended function.

Guidewords can be combined during a HAZOPS analysis. For instance, ‘No or Not’ can be combined with ‘Early’ to consider failure scenarios in which the battery supply of a personal radio system failed to meet the intended endurance. Similarly, ‘No or Not’ can be combined with ‘Late’ to consider the hazards that could arise when a personal radio offered extended life beyond that anticipated in the requirements. It might be argued that such a scenario would not jeopardize operational safety. It might then be dismissed from further consideration. The key point is that the guidewords help to direct hazard analysis within the wider risk management processes.

The next phase of the HAZOPS approach identifies the causes that could result in a particular failure. This is important because there may be additional constraints, for example from the underlying physics, which prevent a problem from arising. Considerable care must be taken whenever HAZOPS teams dismiss a scenario; many accidents have occurred in ways that were never anticipated by system developers. It is also important to stress that there may be more than one way in which a failure can occur. In such situations, teams may reject one cause of a potential risk without adequately considering all of the other ways in which a hazard might occur.

After the causes of a hazard have been determined, HAZOPS teams must identify potential consequences. This creates considerable problems for military applications. The impact of a potential hazard often depends upon the context of use. For example, the consequences of losing a personal radio system will depend upon a host of mission-specific factors. These problems can be addressed by ensuring that HAZOPS teams draw upon a range of stakeholders with operational experience. It is also important to study any existing mishap information to determine the consequences of previous failures. After the causes and consequences of a hazard have been determined, teams must identify potential mitigations. Any potential changes must be subjected to a further round of HAZOPS analysis before they are implemented. The guidewords must be applied to the revised functional design to ensure that the mitigating actions have not introduced any new hazards.

A number of caveats can be raised about the support that HAZOPS offers to the wider processes of risk management. The use of team based techniques does not entirely eliminate potential bias; for example, if individuals ignore the opinions of their colleagues. The subjective nature of the analysis makes it difficult to provide objective evidence that risks have been reduced to an acceptable level. It can also be difficult to use HAZOPS to convince external agencies that all potential hazards have been mitigated, considering that each particular failure might be the result of numerous different causes. A final problem is that this approach provides relatively limited support for the analysis of knock-on failures. These arise when a particular hazard might, in turn, trigger a more serious failure in another area of a complex system. It is for these reasons that a number of military organizations have used variants of FMECA.

1.2.2 FMECA

Failure Modes, Effects and Criticality Analysis (FMECA) has been embedded within military standards [7], handbooks [8] and technical manuals [9]. It has also been adapted to analyze software related hazards [10], meeting some of the risk management requirements that were identified in MIL STD-882D. FMECA begins by defining the system under consideration. It is important to identify any implicit assumptions that are not described in the initial requirements documents because these might create hazards that would not be identifiable in subsequent stages of analysis. In particular, it is important to identify a range of mission profiles. It is also necessary to consider maintenance and decommissioning requirements that extend beyond the initial development.

The second stage develops functional block diagrams. These provide high level overviews of system processes, similar to the opening stages of HAZOPS. Alternatively FMECA can be used during later stages of development. Rather than focusing on high-level functions, it is also possible to analyze individual component and sub-components. This more detailed approach can exploit objective data for sub-system failure rates. This information often cannot be obtained for high-level functional descriptions because it is not possible to determine the precise components that will be used in the early stages of development. In either case, the analysis proceeds in a similar fashion but at different levels of abstraction.

The next stage of FMECA is similar to the application of guidewords within HAZOPS. The analysis progresses by considering different failure modes for each function or component:

- *Untimely operation:* consider what would happen if a function occurred either earlier or later than the intended time or point in a sequence;

- *Failure to operate when required:* consider what would happen if a function did not occur at all;

- *Loss of output:* consider what would happen if a function or component did not yield the anticipated output to other functions or components within a system;

- *Intermittent output:* consider what would happen if a function or component did not continue to provide the anticipated output throughout its lifetime;

- *Erroneous output (given the current condition):* consider what would happen if a function or component provided output that failed to meet functional requirements for a particular set of input conditions;

- *Invalid output (for any condition):* consider what would happen if a function or component provided output that did not meet the functional requirements for all input conditions.

This is not an exhaustive list but it does indicate the range of potential failure modes that must be considered within FMECA. The results of this analysis are, typically, recorded in a matrix where each row is used to denote a function or component. Each column represents a particular failure mode. This ensures that the same set of failure modes are considered for each part of an application. The analysis proceeds by examining each row in the matrix. The cells are annotated to indicate the consequences of each failure mode on the component referred to in the corresponding row. These are the 'effects' mentioned within the name of the approach. They include the total failure of a system as well as degraded modes of operation. Cells in the matrix may also be annotated to show that there is no effect.

The next stage in FMECA helps to prioritize design changes. This is done by calculating the risk associated with the failure mode for each cell in the matrix. The level of risk is derived from the product of severity and likelihood. This helps to distinguish low severity, improbable failures from higher consequence, more probable hazards. MIL-STD 882 provides a classification scheme for the severity of military hazards [1].

- *Category I: Catastrophic.* Could result in one of the following death, permanent total disability, irreversible significant environmental impact or loss exceeding $10M.

- *Category II: Critical.* Could result in one or more of the following permanent partial disability, injuries or occupational illness that may result in hospitalization of at least three personnel, reversible significant environmental impact or loss exceeding $1M but less than $10M.

- *Category III: Marginal.* Could result in one or more of the following injury or occupational illness resulting in 10 or more lost workdays, reversible moderate environmental impact, or loss exceeding $100k but less than $1M.

- *Category IV: Negligible.* Could result in one or more of the following injury or illness not resulting in less than 10 lost work days, minimal environmental impact, or loss less than $10k.

MIL STD 882D [1] also provides a similar classification scheme for the likelihood of particular failure modes. The assessment of both severity and likelihood can be derived from expert judgments. They can also be based on quantitative data about previous failures:

- *Category A: Frequent.* For individual items it is likely to occur in the lifetime of that item; with a probability of occurrence greater than 10^{-1} in that life. For a fleet or inventory item the failure would be continuously experienced.

- *Category B: Probable.* For individual items it will occur several times in the life of that item; with a probability of occurrence less than 10^{-1} but greater than 10^{-2} in that life. For a fleet or inventory item the failure will occur frequently.

- *Category C: Occasional.* For individual items it will occur sometime in the life of that item; with a probability of occurrence less than 10^{-2} but greater than 10^{-3} in that life. For a fleet or inventory item the failure will occur several times.

- *Category D: Remote.* For individual items it is unlikely but possible in the life of that item; with a probability of occurrence less than 10^{-3} but greater than 10^{-6} in that life. For a fleet or inventory item the failure is unlikely but can reasonably be expected to occur.

- *Category E: Improbable.* For individual items it is so unlikely it can be assumed that the occurrence may not be experienced in the life of that item; with a probability of occurrence less than 10^{-6} in that life. For a fleet or inventory item the failure is unlikely to occur but possible.

- *Category F: Eliminated.* Incapable of occurrence in the life of an item. This category is used whenever potential hazards are identified and later eliminated. For a fleet or inventory item the failure is incapable of occurrence within the life of an item.

Analysts can then inspect the risks calculated for each row of a FMECA to derive a ranked list of functions or components for future intervention. Any failure modes that continue to pose an unacceptable risk should become the target for mitigation. Variants of the initial FMECA process have proposed Risk Priorty Numbers (RPN) as a more elaborate means of calculating risk [8]. These RPNs are derived from the product of detectability, likelihood and severity; where each is measured by a scale from 1 to 10. From this it follows that the highest RPN is 10x10x10 = 1000; the failure is impossible to detect, very severe and the occurrence is almost sure. However, the ranking process is the same irrespective of whether RPNs are used or risk is calculated using the product of consequence and likelihood.

The final stages of FMECA identify recommendations for the high-priority failure modes derived from the previous risk computations. For example, design changes can be introduced to help end users identify a potential failure. Alternatively, analysts may recommend the redesign of critical functions or the introduction of components with higher known levels of reliability. It is important to document the output from each stage to support external, independent audit.

As with HAZOPS, there are a number of limitations with FMECA. The scope of the associated documentation can rapidly become overwhelming as analysts are forced to consider a large number of trivial or irrelevant failure modes. It is also hard to represent combined failure modes that simultaneously affect several different components/functions. FMECA can be resource intensive, requiring considerable training. Both HAZOPS and FMECA support large-scale military procurements. However, they provide little help for the tactical and operational risk assessments that are increasingly needed to inform counter insurgency and peacekeeping operations. There is no time to complete an FMECA matrix when you are under fire.

1.2.3 ALARP, ALARA and MEM

Techniques such as HAZOPS and FMECA, support risk reduction rather than risk elimination [11]. The only means of entirely protecting an aircraft from the risk of crashing would be to confine it to the ground. In consequence, the UK Ministry of Defence (MOD) has embodied the 'As Low As Reasonably Practicable' (ALARP) principle within much of its guidance:

> "A risk is ALARP when it has been demonstrated that the cost of any further Risk Reduction, where the cost includes the loss of defence capability as well as financial or other resource costs, is grossly disproportionate to the benefit obtained from that Risk Reduction." (Def Stan 00-56 Issue 4)

In contrast, the term "As Low As Reasonably Achievable" (ALARA) has been integrated into US military doctrine. For example, there is a requirement to demonstrate that the risks of radiation exposure have been reduced As Low As Reasonably Achievable within Army Regulation 700–48 on the Logistics Management of Equipment Contaminated with Depleted Uranium or Radioactive Commodities.

Other countries, such as Germany, have adopted Minimum Endogenous Mortality (MEM) as a means of guiding risk-based decision making. Organizations must ensure that they do not introduce hazards, which will 'significantly increase' the death rate beyond that expected from disease, congenital mortality etc. As we shall see, these requirements open a host of further problems. It is difficult to derive objective measures that can be used to demonstrate military risks are 'As Low As Reasonably Practicable'. Further problems arise when comparisons have to be made with civil operations that carry a lower level of intrinsic risk than many military missions.

1.3 Dimensions of Risk

Military risk management is further complicated by confusion over different dimensions of risk. Although we have characterized hazards in terms of likelihood and consequence, risk can also be viewed as a measure of uncertainty. Further confusion arises over subjective approaches and objective measurements of risk.

1.3.1 Risk and Uncertainty

A risk can exist even though we are uncertain about the likelihood or consequences of an associated hazard. This can be illustrated by UK case law, which established that risk exposes members of the public to the possibility of danger (R. vs Board Of Trustees Of The Science Museum 1993). There is no requirement for anyone to have been harmed or for a hazard actually to have been present. It is sufficient for there to be the possibility of danger for a risk to exist. Defendants must demonstrate that they have taken all practical steps to minimize the risk.

These legal provisions have parallels in the engineering and operation of safety-critical systems where we might need to assess the risk associated with a proposed design even though it has not yet been built. In such situations we cannot be certain about the frequency or consequence of an adverse event because many of the detailed parameters will not have been determined. Similarly, there may be many factors in the operating environment that cannot accurately be assessed until after a system has been commissioned. In such circumstance, we can still talk about the risk of a hazard occurring even though we cannot accurately determine where it might appear on a risk matrix, such as that illustrated in Figure 1.

There are parallel in other domains. Many government agencies are trying to calculate the risk posed by IEDs. We cannot be certain how many people would be killed or injured if we do not know the composition of the device or the location where it might be placed. We are, therefore, often required to make informed judgments or ‘guesses’ about the potential composition of a charge. Furthermore if we have information about previous patterns of attack we can talk about the likelihood of particular outcomes given a range of further assumptions. In other words, we can use experience of past events to calculate the future probability of attacks. The statistical analysis of previous incidents, therefore, supports the assumptions that reduce uncertainty in military risk assessment.

1.3.2 Subjective and Objective Risks

Subjective risk assessments deal with individual and group perceptions of a hazard. These can be influenced by previous experiences, by political pressure, by the media etc. Objective risk assessments use a range of scientific and engineering techniques to identify the likelihood and consequences of particular hazards. These include laboratory studies, epidemiological investigations, probabilistic risk analyses and surveys of accident or incident data

Mishap reporting systems have been used by several military organizations to derive object assessments for strategic, tactical and operational risks [12]. By recording the occurrence of previous failures, it is possible to compile statistics about the probability of failure on demand and about the outcome from adverse events. The resulting risk assessments have been compiled into field manuals and standards that encourage military commanders to reach the same decisions in similar situations.

The complex nature of many military operations undermines the utility of objective risk assessments. For example, the situation facing a military commander can be very different from those in which reliability statistics were gathered. This helps to explain the increased failure rates that were observed when a range of systems were first deployed to the deserts of Iraq or to the mountains of Afghanistan. In such situations, local ‘subjective’ expertise was often a more accurate arbiter of risk than the existing military doctrine.

Fischoff et al question any assumption that objective risk assessments are more reliable than subjective studies. They argue that experts’ opinions should not always prevail over public perceptions of risk [13]. There are many cases in which limited laboratory studies do not provide valid insights into the consequences of ‘real world’ hazards. In other words, there are often elements of subjectivity that influence expert opinions on risk even when they are backed up by empirical studies.

Even if military commanders can agree on a particular level of risk, there is no guarantee that they will reach the same decision in similar situations. Different individuals will react to external threats in different ways. Individual differences are exacerbated when individuals have limited information and significant time pressures. A range of organizational pressures can also make one unit more inclined to accept certain risks than another.

Cognitive biases also complicate subjective risk assessments. For instance, the ‘availability heuristic’ describes how individuals predict the frequency of an event based on how easily an example can be brought to mind. This leads to systematic biases in military organizations that disseminate information about major mishaps that are relatively rare. For instance, personnel may become more concerned about a relatively small number of well publicized friendly fire incidents, compared to a larger number of less publicized road traffic accidents. In such circumstances, subjective risk assessments are likely to overestimate the probability of rare events and under estimate the likelihood of more frequent mishaps.

1.3.3 Units of Measurement

It can be difficult to determine appropriate metrics to use in objective risk assessments. Consequences can be expressed in terms of injuries, fatalities, loss of life expectancy, loss of days of work, the monetary costs of an adverse event etc. The US Army uses a hybrid classification system [14]. This is based on monetary costs and the severity of any injuries. Class A mishaps cost $1,000,000 or more and/or destruction of an Army aircraft, missile or spacecraft and/or fatality or permanent total disability. Unmanned Aircraft System (UAS) accidents are classified based on the cost to repair or replace the UAS. Class B incidents involve damage costs of $200,000 or more, but less than $1,000,000 and/or permanent partial disability and/or three or more people are hospitalized as inpatients. The categorization of Class C incidents changed in 1992. Prior to that date they were defined to incur damage. Since then, they refer to incidents that result in damage between $20,000 and $200,000 or that incur nonfatal injury or occupational illness that causes one or more days away from work or training beyond the day or shift on which it occurred or causing disability at any time.

In addition to this three level classification, US Army Aviators have introduced three other severity levels. Class D accidents involve property damage that is $2,000 or more but less than $20,000. They can also lead to nonfatal injury or illness, which might prevent an individual’s activities or lead to medical treatment that is greater than first aid. Such incidents would include needle stick injuries or cuts from contaminated sharps. Aviation class E accidents involve damage that costs less than $2,000. Class E incidents are associated with operational or maintenance missions that are interrupted or not completed due to a failure or malfunction of a component. Finally, class F incidents refer to Foreign Object Damage (FOD).

The US Army accident severity categorizations cannot easily be mapped onto the severity classification for risk management within MIL-STD 882D. Previous sections have described how this ranges from level I (Catastrophic) through IV (Negligible). In 882D, a category I hazard results in death, permanent total disability, irreversible significant environmental impact or loss exceeding $10M. This compares with a US Army level A accident being valued at $1 Million or more. In terms of the MIL-STD, this would be classified as a level II ‘Critical Hazard’. The lack of any clear relationship between the severity measurement of previous accidents and the risk assessments for future hazards is confusing.

The metrics used to assess likelihoods are just as complex. Risks can be expressed in terms of the number of failures at different levels of severity per time of operation. They may also be defined as the number of failures on demand or when a system is first activated. Risks can also be expressed in terms of the number of different severity incidents per hour of exposure to a hazard, including enemy action. For instance, in the year up to October 2010 the US Army experienced 0.246 Category A ground accidents per 1,000 soldiers. In 2009 this accident rate was 0.274 and in 2008 it was 0.309. For category B accidents the rates were 0.051 (2010), 0.135 (2009) and 0.133 (2008).

Great care must be taken when interpreting risk metrics. For example, a relatively high likelihood of failure on demand can be hidden by a statistic that is stated in terms of failure per hour of operation. Similarly, by averaging out the total number of fatalities either over some previous period of time or in a forward looking risk assessment then there is a danger that analysts might ignore the impact of individual high consequence events. For example, there might appear to be a relatively low risk when considering the safety record across all UK in-flight refueling operations. However, this would overlook the profound impact of the loss of all 14 crew members on RAF Nimrod XV230 in Afghanistan 2006. The high number of fatalities during this accident helped to trigger a review that questioned the foundations of risk management across the UK Ministry of Defence [13].

The ways in which risks are calculated yields important insights into the values of the organization or individual responsible for risk management. For example, in many military operations adverse events may lead to fatalities amongst coalition forces and the civilian population. Stating risks in terms of the probability of fatality for military personnel ignores the impact of other deaths and injuries on the overall objectives for counterinsurgency and peacekeeping operations.

1.4 The Psychology of Risk

The previous section has identified issues that complicate objective measurements of risk. Subjective assessments raise further problems. In particular, it is difficult to avoid the psychological influences that bias our individual perception of the likelihood and consequence of future threats.

1.4.1 Risk Aversion and Regret

Risk aversion can be thought of as a preference to remain in an existing situation rather than take an action that might lead to a worse situation. Military doctrine continually warns against such aversion. Inaction seldom guarantees a favorable outcome. Enemy actions can quickly transform any short term advantage into a longer term failure.

Regret is the subsequent preference for an action that was not taken. Risk aversion serves to delay intervention while regret can motivate decision makers to take potential hasty decisions when faced with a similar situation in the future. Unfortunately, it is typically only possible to assess the impact of these factors in the aftermath of any military operation. Risk aversion might persuade a commander not to take a decision, if things go well then this will be interpreted as a measured response to a threat that was not realized. If things go badly then the same lack of action can be condemned as a contributory factor in any mishap. Similarly, outcomes will determine the assessment of any precipitate actions inspired by regret for lost opportunities in the past.

1.4.2 Fear

The possibility of an adverse event can exact a psychological cost even if the adverse event does not occur. Fear influences all military operations, especially when forces are deployed many miles from home over long periods with relatively short opportunities to reassure friends and family. This form of dread has significant consequences for risk management. Fear can be seen both as a positive and a negative stressor. Some individuals thrive in high-risk nvironments, viewing them as a challenge. The possibility of an adverse outcome encourages them to think of new and innovative ways to reduce the likelihood or mitigate the consequences of any potential problem. Other individuals develop coping strategies, for example by dismissing the possibility of an adverse outcome. This can have negative consequences when it leads to recklessness. For the commander, it can be difficult to determine an appropriate level of concern that is proportionate to the risk created by any particular military operation. This is compounded by uncertainty – for example, there may be insufficient information about enemy dispositions to determine an ‘appropriate’ level of concern.

1.4.3 Risk Equity

Risk equity requires that those who suffer the potential consequences of a hazard should also receive a proportionate share in any potential benefits. In many civilian domains, regulators and decision makers tend to favor situations that encourage risk equity. In contrast, military operations create risk inequality. Opposition forces should receive all of the risk and none of the benefits associated with particular interventions [15].

Problems arise when conflicts also increase the hazards for civilian populations. This can be illustrated by the use of high-density depleted uranium (DU) in armor piercing projectiles. More than 2,000 tones were used by coalition forces during the 2003 conflict in Iraq. There is contradictory evidence on the health impacts of long term exposure to this material. DU is mildly radioactive and has a half-life of up to 5 billion year. DU aerosols are distributed over a wide area following any impact. Landmines and cluster munitions have also raised questions about risk equity in military operations. These technologies provide highly effective support for many military operations. However, their use is increasingly restricted by international agreements. It is difficult to control the level of risk that is created for non-combatants by landmines and cluster munitions both during and after a conflict.

Many other psychological concepts can be used to analyse and support military risk assessments. For example, Wickens [16] has investigated the ability of experts to identify the risks that influence their scope for action when making critical decisions.

Humphrey et al [17] show that the priorities, which individuals associate with safety requirements, change closer to the delivery of complex systems. In other words, project completion can become an overriding priority in the later stages of procurement. Rather than provide a more exhaustive review of the individual psychological factors that contribute to these observations, the following paragraphs focus on the ways in which risk management influences military decision making.

1.5 Risk and Decision Making

Risk is only one of several attributes that must be considered when making complex decisions. Political influences, interpersonal relationships, individual preferences all shape real-world interventions. Risk management processes can, however, encourage a systematic examination of the potential outcomes from complex decisions [12].

1.5.1 Framing

In many situations, military personnel are forced to make complex decisions under extreme time pressure. They cannot 'think of everything'. Instead, it is necessary to consider the risks associated with a small subset of all the possible options that could be selected. This subset refers to the 'frame' or scope of a decision. Framing plays an important role in the practical application of risk assessment to inform military decision making. We must impose limits or bounds on the options that we consider.

Framing suffers from a number of biases. There is often a tendency to ignore familiar hazards. For instance, military people will drive long distances from remote bases in order to see their friends and family. In many cases, they will decide to make extremely long journeys when they may already be tired after periods on duty. In most cases, individuals will decide to make these journeys without considering the associated risks. In other words, they exclude the hazards of privately operated vehicles (POVs) from the frame or scope of their decision.

Framing can also lead individuals to exclude disturbing events that have a non-zero probability. These might include kidnap attempts in counter-insurgency operations. The dread of such events prevents an objective analysis of the associated risks even though the likelihood of kidnapping might be extremely low for any individual. In consequence, by not thinking about the possibility of such an incident military personnel may 'decide' not to take precautions that might protect them from being kidnapped.

Framing can also be shaped by wishful thinking. For instance, individuals are extremely bad at predicting the impact of fatigue on decision making. We become less and less accurate at predicting when we are likely to fall asleep as we become more and more tired. This leads many military personnel to underestimate the hazard. This is exacerbated by the hope that fatigue does not affect us in the same way that it affects everyone else.

1.5.2 Recognition Primed Decision Making

Previous paragraphs have explained how risk management techniques guide decision making. The potential consequences and likelihood of a range of different hazards are first identified. These assessments are then used to select the mitigating actions that support safe and successful military operations. Unfortunately, real-world decision making tends to be more complex. As we have seen, time pressures and uncertain information can prevent individuals from enumerating all potential threats and mitigations. Recognition Primed Decision Making (RPDM) builds on the notion of framing to model differences between theoretical risk management techniques and real world observations [18]. RPDM has also been used to support operational and tactical decision making in military organizations [19]. RPDM models the way in which individuals rapidly assess potential risks. In order to do this, decision makers develop plans that are sufficient to meet high-level goals. RPDM assumes that individuals do not exhaustively search through every possible plan in order to find an optimal solution.

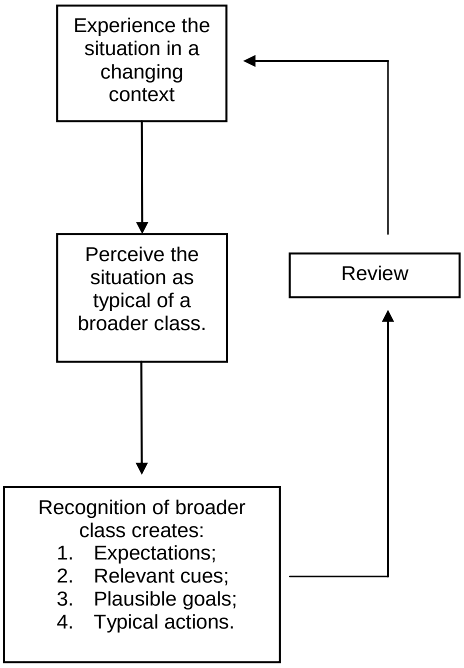

Experience the situation in a changing context
Perceive the situation as typical of a broader class.
Review
Recognition of broader class creates:
1. Expectations;
2. Relevant cues;
3. Plausible goals;
4. Typical actions.

Fig. 2. High-Level Overview of RPDM
(Adapted from [18])

Figure 2 provides a high-level overview of RPDM. The first step is for the decision maker to identify changes in their environment that can act as a trigger for subsequent action. Individuals use these changes to identify their present context as typical of a wider class of similar situations. This is the ‘recognition’ process within RPDM.

Once an individual has identified previous similar situations, they can simplify decision making by drawing on previous knowledge to derive expectations about their existing context. In other words, they can draw on the previous expertise to improve their understanding of the present situation. For example, they might remember to look for further information or cues that can guide their intervention. They can also use previous expertise to identify appropriate goals or actions that have been successful in similar situations. This recognition phase is at the heart of the RPDM process, illustrated in Figure 2. It is characterized by uncertainty; individuals lack perfect knowledge in complex and dynamic environments. In teams this can lead to time-limited negotiation to improve mutual situational awareness prior to collective decision making.

Once an action has been selected and executed, the decision maker must monitor and ‘review’ further changes in their environment. These changes help to confirm the effectiveness of their intervention. Alternatively, they may trigger further decision making if their initial actions did not have the intended effect. RPDM is based around a number of assumptions:

1. Experienced decision makers are usually only concerned to identify a ‘workable’ solution; they do not need to exhaustively consider every possible alternative;

2. Experienced decision makers generate and evaluate options one at a time, instead of comparing the advantages and disadvantages of all options;

3. Experienced decision makers evaluate an option by imagining the outcome, and by finding ways to avoid problems that may arise from its implementation;

4. Experienced decision makers focus on assessing the situation and looking for familiar cues, these cues then help to identify a potential solution;

5. Experienced decision makers are concerned to act quickly, they try to avoid prolonged analysis that may otherwise delay necessary intervention.

Figure 3 shows how the more theoretical perspective of Figure 2 has been applied to characterize interaction with US Army battlefield simulations [19]. As in the previous diagram, information about the present situation is gathered by decision makers. The next stage is to ‘conceptualize the course of action’ by identifying a potential mission. This needs detailed planning in order to ‘operationalize’ high-level objectives. War games can then be developed to determine whether the detailed mission plans are robust against potential threats.

The links between this RPDM model for simulation and the original theoretical models of decision making are reinforced by the argument that any course of action to survive the war gaming will become 'the plan'. There is no need to compare further options once a viable course of action has been developed. Orders are then disseminated and executed to implement the plan. Additional improvisation may then be necessary as changes in the actual situation force revisions to the initial course of action.

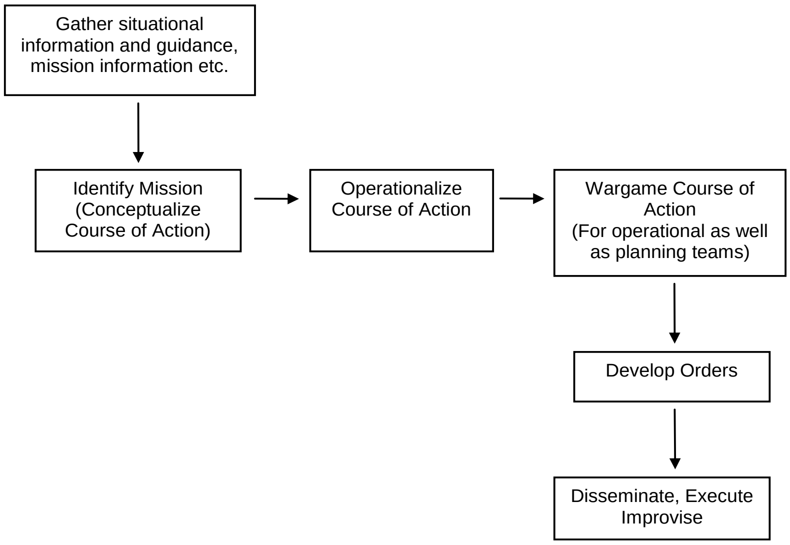

Gather situational information and guidance, mission information etc.
Identify Mission (Conceptualize Course of Action)
Operationalize Course of Action
Wargame Course of Action (For operational as well as planning teams)
Develop Orders
Disseminate, Execute Improvise

Fig. 3. Key Stages in RPDM with Simulated Scenarios (Adapted from [19])

Previous paragraphs have identified a tension between many convention risk management techniques and the decision making processes embedded with the RPDM model. In particular, RPDM suggests that decision makers do not exhaustively consider the likelihood and consequences of all potential hazards. It is possible, however, to reconcile some of these differences. For example, the war gaming techniques in Figure 3 use risk-based simulations to help validate the plans that are developed using RPDM techniques. These simulations model high probability threats to ensure that a potential plan is robust against common hazards.

Similarly, HAZOPS and FMECA can guide the recognition phase in RPDM by helping decision makers to identify common hazards across a class of similar situations. The HAZOPS guidewords focus on common hazards and previous solutions without necessarily considering every possible adverse event.

1.5.3 Situation Awareness and Confirmation Bias

Recognition Primed Decision Making (RPDM) assumes that decision makers cannot exhaustively consider every hazard that might arise in complex situations. Instead, individuals recognize common features in past experiences to help guide future decisions. Endsley's model of situation awareness takes this one step further [20]. Figure 4 shows how this framework begins with a process of filtering that affects our ability to perceive changes in the environment. We cannot monitor the many thousands of simultaneous changes in the hundreds of complex systems that characterize the modern battlefield. By focusing on certain changes, we can minimize the noise that would otherwise overwhelm our ability to respond to potential threats. In other words, Endsley's model begins with a process that strongly resembles framing and recognition in RPDM.

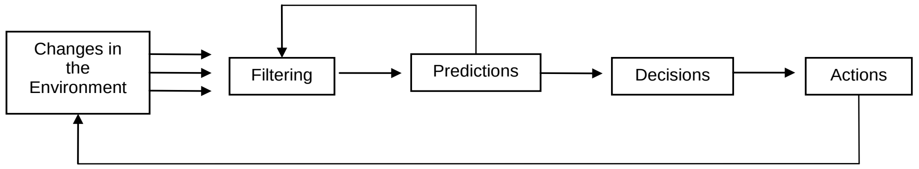


Fig. 4. A Simplified Model of Situation Awareness

The initial filtering in the Endsley model assumes that decision makers will not exhaustively consider the hazards associated with every change in their environment. However, this monitoring can be directed by informal risk assessments. We actively look for signs that might provide information about those hazards that pose the greatest threat to a mission. This leads to problems of confirmation bias. Decision makers actively look for evidence that confirms a particular hypothesis while excluding counter indications that support other hypotheses. Individuals favor information that confirms their preconceptions or hypotheses, irrespective of whether they are true or not.

Within the model illustrated by Figure 4, it is possible to identify three different levels of situation awareness. At the lowest level, decisions have to be informed by the perception of information that is relevant to a mission. However, it is not enough simply to perceive a threat. Decision makers cannot intervene in an appropriate manner if they do not understand the nature of the threat that they face. Finally, higher levels of situation awareness depend upon an individual's ability to make accurate predictions about the future – both in terms of the consequences of a hazard but also in terms of the effectiveness of any mitigating actions:

- *Perception.*

The first level of situation awareness depends upon individuals perceiving key attributes of the entities in their environment. Effective decision making relies upon monitoring a host of objects, events, people, systems, and environmental factors as they change over time. Decision makers must identify critical changes across this broad range of entities if they are to monitor a the many different hazards that undermine military operations;

- *Comprehension.*

The second level of situation awareness relies upon processes of interpretation and synthesis to integrate the information derived from the first level. Individuals have to understand how any changes in their environment will affect their tasks and objectives. In particular, decision makers must be able to interpret the threats that a potential hazard might pose to their operational and tactical goals;

- *Projection.*

The third level of situation awareness focuses on a decision makers' ability to project their knowledge from levels one and two to make predictions about the future. Risk assessment plays an important role in projection because it helps direct our finite analytical resources towards the probable changes that carry the greatest potential consequences for future operations.

Endsley's work is important because risk assessment provides a formal system of projection, by encouraging teams to consider the likelihood and consequences of future outcomes. As we have seen, however, both projection and risk assessment rely on the level 1 (Perception) and level 2 (Comprehension) processes of situation awareness. The following chapters will present many examples where the apparent failure of military personnel to recognize potential risks stemmed from problems in these precursors to projection.

1.5.4 Target Levels of Risk

RPDM has been applied as a means of representing and reasoning about operational military decision making [19]. Situation awareness has also been used to reason about the tactical and strategic ability of senior commanders [21]. However, other theories about risk management remain more controversial. For example, Wilde uses the term 'risk homeostasis' to refer to an individual's preference to meet a target level of risk [22]. The introduction of a new safety feature reduces the actual level of risk. In consequence, individuals are tempted to trade-off the safety benefits by pushing for increased levels of performance.

Much of the work on target levels of risk has focused on automotive safety. The aim has been to explain why the development of advanced braking systems or the compulsory use of seatbelts has not resulted in the anticipated reductions in accident rates. One explanation is that drivers have maintained a target level of risk by driving faster; in the knowledge that their new safety systems will provide a degree of protection from the consequences of their actions.

There are military applications of risk homeostasis. For example, the introduction of infrared and image enhancement technology can reduce the risk of accidents during night operations. However, subsequent chapters in this book will argue that the safety benefits derived from these systems may be offset by an increased pressure to accept missions that would not otherwise have been considered feasible.

Risk homeostasis is controversial because it suggests that individuals have access to feedback mechanisms about the changing levels of risk in their environment. The application of this theory to driver behavior has been attacked because few individuals have the technical knowledge to accurately assess the safety improvements derived from novel braking systems. Hence there seems little prospect that they could accurately predict the degree of protection that any safety device might offer if they were to be involved in an accident. Similarly, it is difficult for military commanders to accurately assess the safety benefits of night vision technology that might then be offset against increasing levels of operational risk.

1.6 Sociology of Risk

Previous paragraphs have focused on risk assessment in terms of individual decision making. However, even in military operations, critical decisions are the product of consultation and negotiation especially where operations rely on the tight integration of multiple units from different nations. One person may be responsible for the decision; however, many people help to inform that decision.

1.6.1 Risky Shift, Group Caution and Risk Polarization

Individuals adapt in different ways to fear, regret and uncertainty. From this it follows that they will perceive risks in different ways. This has important implications for any team that cooperates to make complex decisions. Group interactions will be shaped by individual perceptions of risk. Laboratory studies have helped to identify particular phenomena that characterize team-based decision making [23]. For example, risky shift occurs when groups gradually move towards the positions adopted by individuals with the greatest tolerance for risk. Their involvement in a group situation makes an individual more willing to accept a risk than they would have been if they had acted alone.

Risky shift can be explained in terms of ‘social comparison’ by which individuals do not want to appear risk averse in comparison to their colleagues. This may be a significant factor behind many military decisions that accept high risk strategies in the face of uncertain information.

It is also possible to identify an opposite phenomena to risky shift. Group caution occurs when individual members of a team urge their colleagues to consider risks that might otherwise have been ignored. This can prevent or delay groups from taking necessary actions and can be disastrous for many military operations.

Risky shift and group caution are both captured by the term 'group risk polarization'. This described the way in which groups tend to reinforce or amplify an accepted view. The wisdom of a particular decision is supported not through the validity of any associated argument but through the repetition of a particular point of view and a resistance to any dissenting opinions.

1.6.2 Risk Transfer

Risk transfer refers to the way in which some individuals and groups seek to shift a risk onto others. In everyday life, this can be done by taking out an insurance policy. The risks associated with a particular hazard are transferred to the underwriters who must consider both the likelihood and consequences in determining an appropriate premium. Such contingencies are less easy to identify within military operations that are excluded from most forms of insurance cover. However, risk aversion may lead individuals and units to off-load a potential risk by asking others to perform high-risk operations. One example of this has been the increasing use of private enterprise and commercial sub-contractors to support conventional military forces in counter insurgency operations.

Shaw has argued that risk transfer lies at the heart of the 'new Western' approach to war [15]. He uses recent operations in Iraq to argue that political and social pressures are forcing Western governments to minimize the risks facing their armed forces. However, this can lead to the problems of risk equity described in previous sections. Risk equity assumes that those groups which benefit most from an activity should also face the associated risks. Moral dilemmas and operational problems arise when the transfer of risk away from a military unit increases risk for the civilian population. For example, relaxing the rules of engagement can reduce the risk to a patrol by enabling action to be taken before a vehicle reaches a checkpoint. However, the local population faces an increased risk of being fired upon before the patrol has fully determined whether or not they pose a threat to them. Shaw argues that this risk inequity creates social and political pressures to conceal the impact of risk transfer on local populations.

1.6.3 Decision Transfer

Decision transfer describes how some individuals or groups to shift responsibility for making a risk-based decision. An example would be when an individual or unit passes a decision up the Chain of Command. It is important to emphasize that shifting the responsibility for a decision does not necessarily imply a shift in the risk itself. More senior levels of command may still ask the original decision maker to perform the operation that exposes them to a particular hazard. However, decision transfer usually implies a transfer of responsibility should anything go wrong.

Decision transfer is a critical issue in the sociology of military risk taking. Existing doctrine is intended to increase autonomy so that units can rapidly respond to dynamic, complex and uncertain environments. In contrast, decision transfer enables local units to refer operational matters higher up the chain of command. This can introduce delays and potentially undermines the flexibility provided by limited local autonomy. There is also a danger that higher levels of command will be overwhelmed by requests to assist in operational and tactical matters.

Later chapters in this book will argue that the introduction of formalized risk management techniques into military operations have encouraged risk aversion and decision transfer. By explicitly asking commanders to enumerate potential hazards prior to an operation, they are less willing to take responsibility for those risks. Instead, officers seek additional reassurance from higher levels in the command chain. This removes responsibility for decision making from those individuals and groups who are best informed about the nature of local threats and opportunities.

1.6.4 Risk Society and the Precautionary Principle

Giddens [24] and Beck [25] have challenged many underlying assumptions about the wider role of risk as a tool for scientific and engineering discourse. In the past, risks stemmed from environmental factors that were largely outside of our control, such as floods or earthquakes. In contrast, many of the risks for modern society are created by society itself; they are ‘manufactured risks’. Giddens and Beck have observed that successive accidents and incidents have undermined public faith in industry, government and experts. Events from Chernobyl to the Challenger disaster raise doubts about our ability to maintain public safety. This leads to a form of reflection where public concerns over the risks associated with particular activities are used to determine whether or not those activities are permitted. One consequence of this has been the precautionary principle; if there is no scientific consensus that an activity will not create a risk then it should not be permitted.

Beck’s Risk Society [24] goes on to argue that knowledge has supplanted wealth as a means of mitigating risks. In previous centuries, wealthy individuals often contributed to the threats that were faced by society. For instance, industrialists benefited from the products that created the pollution which affected a wider section of society. They could, however, use the profits from such activities to mitigate their personal risks –by buying a house away from the source of pollution. In contrast, the risks posed by many modern processes are not easily identified by the lay person, for instance when the longer term effects of toxins are unknown. Engineers and scientists gain significant political power and social advantage through their ability to identify and explain these risks.

The work of Beck and Giddens has important implications when the public observe the ‘manufactured risks’ of military operations. Media reports of high and sustained numbers of casualties raises questions about the judgments of government, experts and senior commanders. However, it is clear that civil standards of risk, such as those embodied within the precautionary principle, cannot easily be applied to military operations. Wexler identifies a host of concerns from the application of ‘risk society’ concepts to modern warfare. Using Agent Orange and depleted uranium projectiles as case studies he argues that the military application of the precautionary principle will focus public attention on those weapons with the most easily imagined and feared risks; “older weapons, to which the public may have grown inured, often pose equal or greater risks, but they are less likely to raise alarm… a military precautionary principle often fails to prioritize those actions that best promote the public health and the environment” [26].

1.7 Overview of the Remaining Chapters

The intention in this section has not been to provide an exhaustive overview of the many different aspects of risk and risk assessment that are relevant to modern military operations. Instead, the aim has been to develop a limited vocabulary of common concepts in risk and decision making that will support the more detailed analysis of operational risk assessment in subsequent chapters. The remainder of the book can be summarized as follows:

- Chapter Two identifies the paradoxes that arise from military risk assessment. For instance, many military organizations focus on the investigation of adverse incidents or accidents. This creates unnecessary conservatism when future risk assessments are informed by a detailed record of previous failures. The intention in this section of the book is to draw out some of the particular issues that affect military as opposed to civil approaches to safety-related operations. For example, in order to prepare participants for the hazards of modern warfare, military training must simulate situations that inevitably place individuals at some level of risk. It is important to identify these paradoxes if we are to understand some of the problems that arise when civil risk assessment techniques are applied to complex military operations.

- Chapter Three builds on the first of the two paradoxes mentioned above. In particular, it looks at the manner in which military organizations collect and collate data about previous adverse events. This is important if we are to derive objective assessments of the likelihood of future incidents. However, chapter three identified numerous problems that prevent accurate estimates for the probability of military incidents. These include the difficulty of exchanging information between units that are distributed around the globe. They also include the difficulty of conducting incident investiogations and causal analysis whilst under enemy fire.

- Chapter Four looks beyond the high-level statistics to consider a particular example of the problems that complicate military risk assessment. This chapter deals with an accident on-board a UK submarine involving Self-Contained Oxygen Generators. These devices were considered within a detailed safety case that was intended to demonstrate that they posed an acceptable level of risk. However, the chapter goes on to explain how these arguments underestimate some of the risks that arose from 'degraded modes' of operation. These arise when military personnel find 'work arounds' and 'quick fixes' that are intended to meet mission objectives even when thee may be considerable problems with the equipment that they must operate.

- Chapter Five uses data derived from incidents such as that described in Chapter Four to identify the systemic problems of fatigue for military operations. The insidious effects of this problem operate at multiple levels. Not only does fatigue increase the likelihood of many hazards it also undermines our ability to assess the risks of those hazards. Fatigue undermines effective decision making, it also undermines our ability to identify that we are making poor decisions.

- Chapter Six extends the analysis of the previous chapter and identifies more sustained problems for military risk assessment. Night vision devices (NVDs) have been proposed as one means of addressing the causes of fatigue. However, the introduction of these devices can create new risks. Sustained scanning using existing technologies increases rather than reduces levels of fatigue. These devices have also created new forms of perceptual error when drivers have to rapidly react to the terrain displayed in NVD displays. There are deeper systemic effects that arise when commanders accept risks on missions that would never otherwise have been permitted without NVD support.

- Chapter Seven again goes beyond the high-level statistical data to look in detail at the operational risks that can arise from the introduction of new technologies into modern warfare. In this case, the focus is on the loss of a rotary wing aircraft whose crew were using NVDs during brown-out conditions. The chapter begins by an analysis of the risks associated with these operations and at the technological or doctrinal mitigations that have been proposed. The second half of this chapter describes how operational demands systematically undermined each of these safeguards, exposing the crew to unnecessary levels of risk.

- Chapter Eight takes some of the lessons learned from the loss of helicopters and land-based vehicles to look at the use of innovative and disruptive technologies to mitigate military risk. In particular, this chapter looks at the rapid introduction of Unmanned Airborne vehicles (UAVs) to meet political commitments within NATO. The irony is that these systems were intended to reduce the risks to conventional forces but often, instead, exposed them to increased levels of risk as they were forced to retrieve them from vulnerable crash sites.

- Chapter Nine provides an example of the risks that are identified in Chapter Eight by examining the operational hazards that led to a fatality during the retrieval of a UAV. These risks were entirely overlooked from the risk assessments conducted by the manufacturers of these devises even though they were deliberately marketed as essential infrastructure in counter insurgency operations.

- Chapter Ten expands the scope of our work to consider the application of civil risk assessment techniques to consider the systems level threats posed by Improvised Explosive Devices (IEDs). In particular, we argue that it is difficult to anticipate the consequences that stem from the rapid dissemination of IED technology and tactics around the globe.

The closing sections of this book review the myriad of operational insights that have been summarized in previous chapters. The aim is to reiterate our underlying motivation to expose the difficult of using civilian risk assessment techniques to anticipate and mitigate the hazards of modern warfare. The closing chapter goes on to identify intial solutions that might be used to address these limitations.

1.8 References for Chapter One

[1] US Department of Defence, Standard Practice For System Safety: Environment, Safety, and Occupational Health, Risk Management Methodology for Systems Engineering, MIL-STD-882D w/CHANGE 1, Washington, USA, 29 March 2010

[2] T. Kletz, HAZOP and HAZAN, Taylor & Francis, London, UK, 2006.

[3] D. MacDonald, Practical HAZOPS, Trips and Alarms, Oxford, UK, 2004.

[4] UK Ministry of Defence, HAZOPS in Software Control Systems, UK Def STAN 00-58, 2000 (now superceded).

[5] Woodward-Clyde Consultants, Hazard and Operability Study (HAZOP) Rocky Mountain Arsenal, Basin F Liquid Incineration, Available via US Defence Technical Information Centre, July 1992.

[6] US Army Environmental Policy Institute, Managing the Life Cycle Risks of Nanomaterials, Arlington, Virginia, July 2009.

[7] US Department of Defense, Procedures for Performing a Failure Mode, Effects and Criticaility Analysis. MIL–HDBK–1629A. 1980.

[8] US Department of Defense, Electronic Reliability Design Handbook: Section 7.8 Failure Mode and Effects Analysis (FMEA), MIL–HDBK–338B, 1998.

[9] US Army, Failure Modes, Effects and Criticality Analysis (FMECA) For Command, Control, Communications, Computer, Intelligence, Surveillance, And Reconnaissance (C4ISR), Facilities, Technical Manual 5-698-4, September 2006.

[10] J.H. Graham, FMECA Control for Software Development. In IEEE 29th Annual International Computer Software and Applications Conference, COMPSAC 2005, 93-96, 2005.

[11] N. Trewina, U. Ojiakoa and J. Johnsona, Risk management and its practical application: lessons from the British Army, Journal of Risk Research, (13)5, 2010.

[12] C. Haddon-Cave, The Nimrod review: An independent review into the broader issues surrounding the loss of the RAF Nimrod MR2 Aircraft XV230 in Afghanistan in 2006, HC 1025 London: The Stationery Office, October 2009.

[13] B. Fischhoff, S.R. Watson, C.Hope, Defining Risk, Policy Science, 17, 123-139, 1984.

[14] US Department of Defence, Accident Investigation, Reporting, and Record Keeping, Department of Defence Instruction (DODI) 6055.7, Washington, USA, 2000.

[15] M. Shaw, The New Western Way of War: Risk-Transfer War and Its Crisis in Iraq, Polity 2005.

[16] C.W. Wickens, Engineering Psychology And Human Performance, Harper Collins, London. 1992

[17] S.E. Humphrey, H. Moon, D.E. Conlon and D.A. Hofmann, Decision-Making And Behavior Fluidity: How Focus On Completion And Emphasis On Safety Changes Over The Course Of Projects. Organizational Behavior and Human Decision Processes, Volume 93, Issue 1, January 2004, Pages 14-27.

[18] G. A. Klein, J. Orasanu, R. Calderwood and C.E. Zsambok, Decision Making In Action: Models And Methods, Ablex Publishing, Norwood, New Jersey, USA, 1993.

[19] K.G. Ross, G.A. Klein, P. Thunholm, J.F. Schmitt & H.C. Baxter, Recognition-Primed Decision Model, US Army Military Review, l:6-10, 2004.

[20] M.R. Endsley, Measurement of Situation Awareness in Dynamic Systems. Human Factors, 37, 65-84, 1995.

[21] J. Eid, B.H. Johnsen, W. Brun, J. Chr. Laberg, J.K. Nyhus, G. Larsson, Situation Awareness and Transformational Leadership in Senior Military Leaders: An Exploratory Study, Military Psychology, (16)3:203-209, August 2004.

[22] G.J.S. Wilde, Target Risk 2: A New Psychology of Safety and Health, PDE Publications, Toronto, Canada, 2001.

[23] C.L. Sia, B.C.Y. Tan and K.K. Wei, Group Polarization and Computer-Mediated Communication: Effects of Communication Cues, Social Presence, and Anonymity. Information Systems Research, 13, 1, 70-90, 2002.

[24] A. Giddens, Runaway World: How Globalization is Reshaping Our Lives. Profile Publishing, London, 1999.

[25] U. Beck, Risk Society, Sage Publications, London 2000.

[26] L. Wexler, Limiting the Precautionary Principle: Weapons Regulation in the Face of Scientific Uncertainty, University of California at Davis, Law Review, 39:459-529, 2006.

2 The Challenge of Military Risk Assessment

The language of hazards and threats, of consequences and likelihood has begun to influence strategic, tactical and operational military planning. This chapter identifies a number of issues that complicate the use of civil methods to support military risk assessment. The intention is not to make a series of political points or to criticize particular risk assessment techniques. Instead, the intention is to enumerate some of the underlying problems that are often ignored by the proponents of these approaches. The opening sections identify problems that affect the use of risk assessments at a strategic level, for example in the use of threat assessments to justify military expenditure. Subsequent sections look at tactical and operational issues.

2.1 Threat Inflation

US military policy is strongly influenced by the risk assessments that the Chairman of the Joint Chiefs of Staff present to Congress. The US Code United States Code (10 USC Sec. 153), which documents the laws of the United States, contains a requirement that:

> (b) Risks Under National Military Strategy. - (1) Not later than January 1 of each odd-numbered year, the Chairman shall submit to the Secretary of Defense a report providing the Chairman's assessment of the nature and magnitude of the strategic and military risks associated with executing the missions called for under the current National Military Strategy.

For example, in 2007 General Peter Pace reported that the strains imposed on the US military forces had increased the risks it faces in defending the nation from ‘moderate’ to ‘significant’ as a result of the ongoing wars in Iraq and Afghanistan. The important of these high-level risk assessments does not stem from any underlying quantitative calculation but from their impact on public opinion and on political decision making. Following Pace’s upgrade in the level of military risk, Defense Secretary Robert Gates was required to provide Congress with a mitigation assessment explaining how the Pentagon would respond to potential contingencies in the light of the CJCS statement. Subsequent Senate debates over the Emergency Supplemental Spending Bill reveal the complexity of these statements. **The link between perceived levels of risk and increased expenditure has left the military vulnerable to claims that threats have been over estimated.**

There is a danger that public and political opinion will perceive a form of ‘threat inflation’ as senior command seeks to justify additional levels of spending through exaggerated claims about the likelihood or consequence of particular hazards. This stems from some of the fundamental issues identified in the opening chapter. It can be hard to demonstrate that a threat would have been relaised if the money had not be spent. In other words, we have to justify military expenditure by showing that something did not occur.

2.2 Bias Towards Short Term Threats

Public and politicians have become increasingly sensitive to the use of military risk assessments as a means of justifying increased expenditure. The economic downturn 2007-2009, increased pressures on government budgets. It also led many to question whether the remaining threats justified the level of military expenditure. This led to significant pressures within the Democratic Party as president Obama sought to justify his decision to double the US military presence in Afghanistan. One element of this review has been a realization that by tying military expenditure to specific and immediate threats, military organizations may become fixated on short term requirements. There is a danger that longer term risks will be ignored.

A second concern can be seen in a joint briefing about the FY2011 military budget held by US Secretary of Defense Robert Gates and Admiral Mike Mullen, the present chairman of the Joint Chiefs of Staff[1]. This was intended to help justify a 1.8% growth in expenditure at a time of rising fiscal deficits. Gates opened the meeting with a declaration that his objective was now to "rebalance our programs in order to institutionalize and enhance our ability to fight the wars we are in today, while at the same time providing a hedge against current and future risks and contingencies...The budget and the reviews are also shaped by a bracing dose of realism – realism with regard to risk, realism with regard to resources. We have, in a sober and clear-eyed way, assessed risks, set priorities, made tradeoffs and identified requirements based on plausible, real-world threats scenarios and potential adversaries". The sensitivity and care needed to address the accusation of threat inflation can be seen in Gate's closing remarks that "we must remember that every defense dollar spent on a program excess to real-world military needs is a dollar not available to take care of our people, reset the force, win the wars we are in and improve capabilities in areas where we are underinvested and potentially vulnerable... So I would say in terms of risk – a year-and-a-half ago, or 2 years ago, our level of highest risk was actually in the current fight, not in terms of our future capabilities. I believe that we have now, by taking a little risk on the high-end capabilities, have significantly reduced the risk in the current fight". This speech encapsulates the realisation that by focussing on immediate threats as the justification for increased military resources, there is an urgent need to rebalance longer term expenditure as a 'hedge' against future risks. **The rhetorical connection between short term operational risks and military budget provision has obscured longer term threats.**

2.3 Development-Risk and Innovation

Not only has the language of risk assessment come to dominate the higher levels of political debate over military expenditure. It has also begun to guide procurement and acquisition. The General Accounting Office (GAO) have argued that the Department of Defense (DoD) must exploit risk-based approaches to strategic investment given

[1] http://www.jcs.mil/speech.aspx?ID=1320

increasing financial pressures in an uncertain security environment [1]. The DoD's Business Transformation Agency has responded by developing the Enterprise Risk Assessment Model (ERAM) to mitigate risks during acquisitions [2]. A 'risk assessment team' spends two weeks reviewing existing project documentation. This analysis then informs a series of more focused interviews with program stakeholders that last from 2-3 days. A further two weeks are then spent reviewing material, formulating additional questions and devising a risk mitigation proposal. The program manager helps to review the initial findings before a final mitigation strategy is disseminated to program participants. ERAM outputs identify vulnerabilities, propose solutions, and provide an action plan to reduce program risks. The intention is to ensure that Department of Defense projects deliver *capabilities* rather than focusing on particular technologies.

One means of mitigating risk is to support multiple development strands. Programme managers can exploit diversity between different suppliers in case one team hits a technological or organizational barrier during design or procurement. However, a capability-based program that spreads development risk between alternate technologies can also lead to resource starvation and under-investment in key areas. Program management must find the money to fund several alternate development teams in case one or more of them fails. However, this may mean that none of the competing strands has sufficient funding to address all of the technical challenges. This approach can also dissuade companies from exploring innovative solutions. There is a concern that they will be compared to rival development teams before their approach is fully mature. A further practical concern is that there are no agreed ways for military programme managers to measure the degree of risk that they have accepted within a procurement initiative. This makes it difficult to accurately assess the resilience provided by diverse, competing suppliers. In consequence, the multy-party approach envisaged within many procurement initiatives only last as long as an initial contract negotiation, after which a reliance on single suppliers quickly undermines the mitigation strategies put forward within ERAM. **It is difficult to identify appropriate metrics for measuring the success of any risk-based approach to military planning and acquisitions, too stringent control might eliminate program novelty while too lax control may lead to program failure.** Risk mitigation through multiple development strands also implies a level of wasted effort for those initiatives that are not selected for subsequent deployment. This creates concerns because there are no accepted metrics for assessing the degree of risk across a programme. Hence it will be difficult for any risk based procurement to accurately anticipate the amount of investment that might be wasted by parallel development teams; we cannot identify the minimum, amout of redundancy that would be necessary to ensure project completion. At the same time, efforts to assess program risk can be undermined by ad hoc attempts to meet urgent operational requirements. In order to overcome some of the inflexibility inherent in stanardised procurement practices, the DoD can issue undefinitized contract actions (UCA). These authorize contractors to begin work before reaching a final contract agreement.

Recent GAO reports have identified the risks associated with such innovations; “contractors lack incentives to control costs during this period” [3].

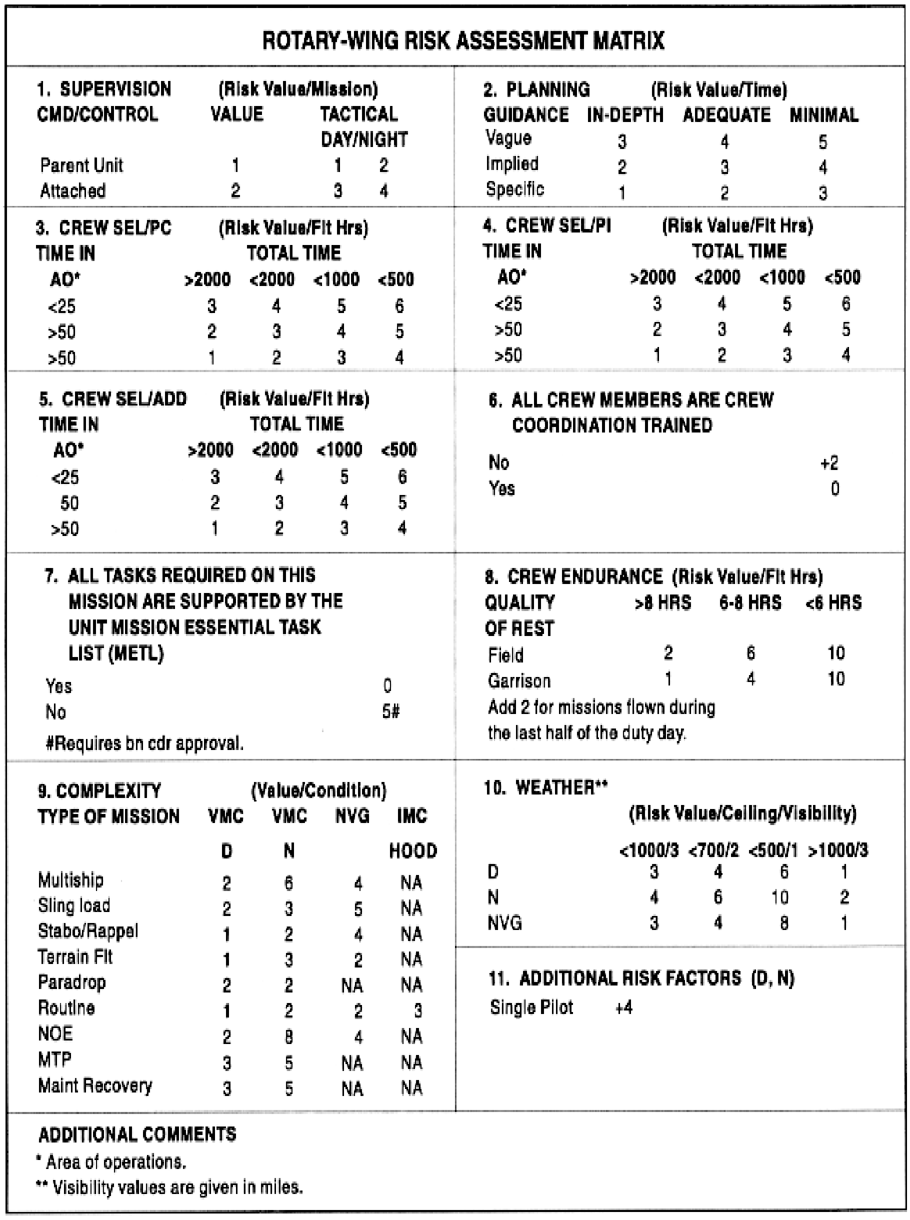
It remains to be seen whether or not ERAM can have the impact envisaged by the GAO. One reason for this is that ERAM was introduced in a piecemeal fashion. In April 2006, the Under-Secretary for Defense (Acquisition, Technology, and Logistics) approved a trial of ERAM focusing initially on the Defense Integrated Military Human Resources System, General Fund Enterprise Business System, and Integrated Data Environment/Global Transportation Network Convergence projects. These initiatives were chosen because they are typical of the business critical ICT applications that often pose particular problems for public agencies acquisition. However, the gradual introduction of the model crates problems because previous DOD initiatives have failed to achieve ‘critical mass’: “…there is no specific Defense-wide policy requiring vulnerability assessments or criteria for prioritizing who should be targeted first. This has led to uneven application of this valuable risk assessment mechanism.” [4].

ERAM is one part of a more general response to the principles encapsulated in Department of Defence Directive 5000.1 and Instruction 5000.2. These advocate the use of risk-based approaches across all procurement activities, including weapon systems and automated information systems. Instruction 5000.2 is intended to establish a management framework to translate ‘mission needs and technology opportunities’ into ‘stable, affordable and well managed’ acquisitions programs. Again, risk assessment is advocated as a key tool in achieving these objectives. The gradually development of ‘evolutionary’ prototypes or demonstrators will help end-users, testers and developers flush out any risks that were not identified during the inception stage. This was intended to address GAO concerns that pilot programs should be limited to low-cost, low-risk prototypes [5]. The evolutionary approach advocated in 5000.1 and 5000.2 helps to explain the piecemeal application of ERAM, described in previous paragraphs. It is unclear how the bureaucratic structures that support these initiatives will help with the higher-levels of strategic decision making. **In military acquisitions there is a tension between accepting sufficient risk to create innovative systems that exceed enemy capabilities and yet rejecting those projects that are so innovative that they are unlikely to yield operational benefits within a fixed timescale and to a specified budget.** The tension between the need to control program risks and at the same time enable innovation can be illustrated by the manner in which the DOD meets urgent operational requirements. In order to overcome some of the inflexibility inherent in standardized procurement practices, the DoD can issue undefinitized contract actions (UCA). These authorize contractors to begin work before reaching a final contract agreement. Recent GAO reports have identified the risks associated with such innovations; “contractors lack incentives to control costs during this period” [3].

2.4 Risk Assessments Fixated with Failure

Previous sections have focused on the problems that complicate risk assessment as a means of guiding military strategy and procurement. Risk assessment techniques are increasingly being used to guide the planning and execution of tactical military operations. For example, US Army Field Manual 3-04.513 deals with battlefield recovery and evacuation of aircraft. It places responsibilities on all soldiers who must: understand, accept, and implement risk reduction guidance and the concept of risk management and assessment; maintain a constant awareness of the changing risks associated with the operation; make leaders immediately aware of any unrealistic risk reduction procedure and report risks beyond their control or authority to their superiors for resolution. It states that “Risk management is a commonsense tool that leaders can use to make smart risk decisions in tactical and everyday operations. It is a method of getting the job done by identifying the areas that present the highest risk and taking action to eliminate, reduce, or control the risk. It is not complex, technical, or difficult” [6]. This view clearly contradicts some of the complexities that we have identified in the opening chapters of this book. It neglects the problems of risky shift or of decision transfer that influence the practical application of risk assessment techniques in military operations. Individual soldiers often have a strong, informal sense of the factors that increase the hazards of different operations [7, 8]. There is a danger that over-simplification will create dissonance between the official doctrine and the individual sense of risk. For example, many existing approaches ignore risk exposure; the likelihood and consequences of an adverse event are calculated without reference to the length of time a mission can take. This helps create straightforward techniques by simplifying the underlying mathematics that soldiers might need to perform in calculating risk metrics. However, this can undermine faith in any approach when individuals understand the basic difference between a hazard that lasts 10 minutes or an hour. If such factors are omitted in an attempt to simply the guidance then operational staff may ignore risk management programmes and their associated field manuals.

Figures 5 and 6 illustrate one of the US Army’s risk assessment tools. The box labeled ‘1. Supervision CMD/CONTROL’ provides a means of assessing the risks associated with operations involving personnel from the same unit or from an attached unit. Particular hazards stem from devolved lines of command hence a higher risk value is associated with operations involving crews from attached units than those for which all staff are drawn from the same command. This section of the form also associates a higher level of command and control risk with operations after dark. A mission involving attached units at night would be assigned an initial risk value of 4. In contrast, a mission that was conducted by an integrated unit in daylight would only score a risk value of 1. These scores are based on ‘human-error accelerator profiles’ that the US Army has derived from longitudinal studies of their accident data. An example of a high-risk mission profile would be an NOE (‘nap of the earth’) flight using night vision goggles with less than 23% and 30 degrees of illumination. The lack of illumination and limited visual field make crew scanning errors more likely to occur. Hence, these factors may be given a high risk-value weighting within the matrix.



ROTARY-WING RISK ASSESSMENT MATRIX

1. SUPERVISION (Risk Value/Mission)

CMD/CONTROL	VALUE	TACTICAL	
		DAY/NIGHT	
Parent Unit	1	1	2
Attached	2	3	4

2. PLANNING (Risk Value/Time)

GUIDANCE	IN-DEPTH	ADEQUATE	MINIMAL
Vague	3	4	5
Implied	2	3	4
Specific	1	2	3

3. CREW SEL/PC (Risk Value/Flt Hrs)

TIME IN	TOTAL TIME			
AO*	>2000	<2000	<1000	<500
<25	3	4	5	6
>50	2	3	4	5
>50	1	2	3	4

4. CREW SEL/PI (Risk Value/Flt Hrs)

TIME IN	TOTAL TIME			
AO*	>2000	<2000	<1000	<500
<25	3	4	5	6
>50	2	3	4	5
>50	1	2	3	4

5. CREW SEL/ADD (Risk Value/Flt Hrs)

TIME IN	TOTAL TIME			
AO*	>2000	<2000	<1000	<500
<25	3	4	5	6
50	2	3	4	5
>50	1	2	3	4

6. ALL CREW MEMBERS ARE CREW COORDINATION TRAINED

No	+2
Yes	0

7. ALL TASKS REQUIRED ON THIS MISSION ARE SUPPORTED BY THE UNIT MISSION ESSENTIAL TASK LIST (METL)

Yes	0
No	5#

#Requires bn cdr approval.

8. CREW ENDURANCE (Risk Value/Flt Hrs)

QUALITY OF REST	>8 HRS	6-8 HRS	<6 HRS
Field	2	6	10
Garrison	1	4	10

Add 2 for missions flown during the last half of the duty day.

9. COMPLEXITY (Value/Condition)

TYPE OF MISSION	VMC	VMC	NVG	IMC
	D	N		HOOD
Multiship	2	6	4	NA
Sling load	2	3	5	NA
Stabo/Rappel	1	2	4	NA
Terrain Flt	1	3	2	NA
Paradrop	2	2	NA	NA
Routine	1	2	2	3
NOE	2	8	4	NA
MTP	3	5	NA	NA
Maint Recovery	3	5	NA	NA

10. WEATHER** (Risk Value/Ceiling/Visibility)

	<1000/3	<700/2	<500/1	>1000/3
D	3	4	6	1
N	4	6	10	2
NVG	3	4	8	1

11. ADDITIONAL RISK FACTORS (D, N)

Single Pilot	+4

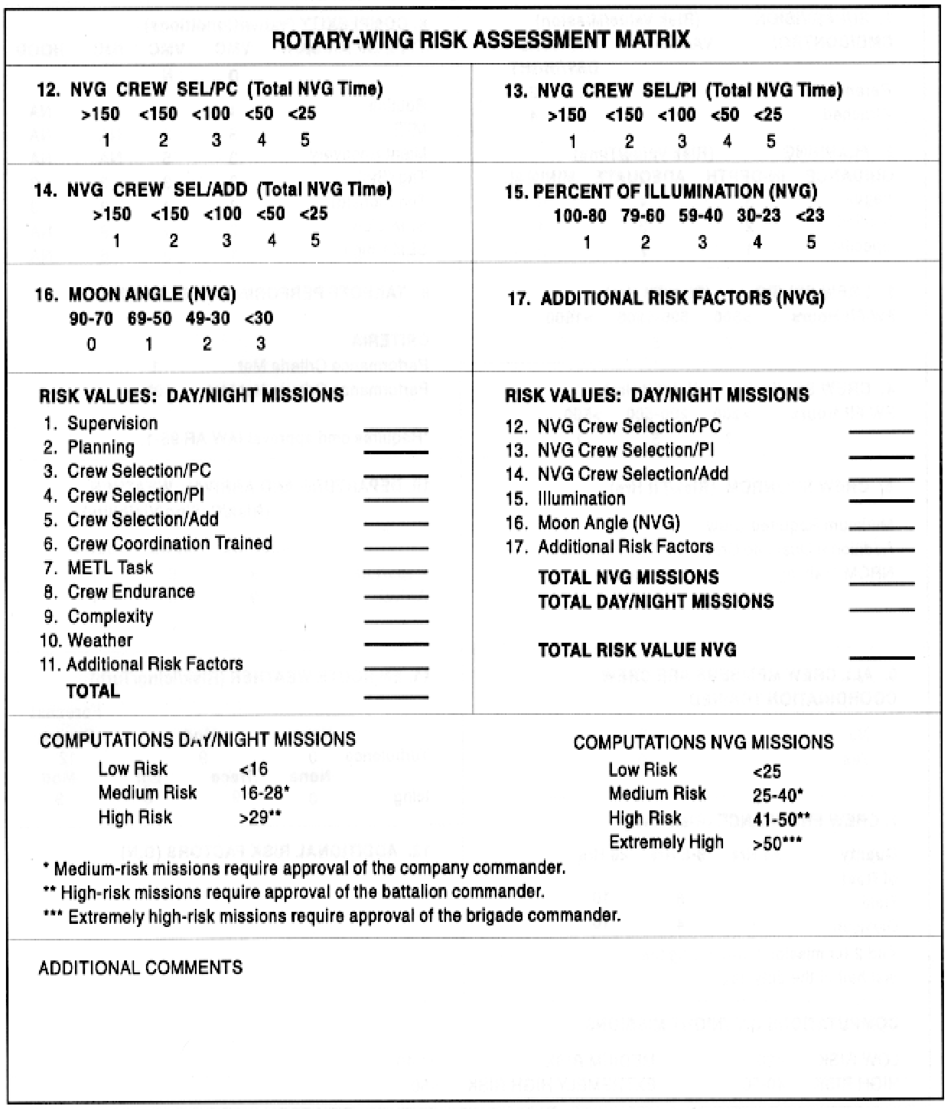
ADDITIONAL COMMENTS

* Area of operations.

** Visibility values are given in miles.

Fig. 5. Rotary-wing risk assessment matrix

(US Army TC 1-210 [9])



ROTARY-WING RISK ASSESSMENT MATRIX

12. NVG CREW SEL/PC (Total NVG Time)

>150	<150	<100	<50	<25
1	2	3	4	5

13. NVG CREW SEL/PI (Total NVG Time)

>150	<150	<100	<50	<25
1	2	3	4	5

14. NVG CREW SEL/ADD (Total NVG Time)

>150	<150	<100	<50	<25
1	2	3	4	5

15. PERCENT OF ILLUMINATION (NVG)

100-80	79-60	59-40	30-23	<23
1	2	3	4	5

16. MOON ANGLE (NVG)

90-70	69-50	49-30	<30
0	1	2	3

17. ADDITIONAL RISK FACTORS (NVG)

RISK VALUES: DAY/NIGHT MISSIONS

1. Supervision ______
2. Planning ______
3. Crew Selection/PC ______
4. Crew Selection/PI ______
5. Crew Selection/Add ______
6. Crew Coordination Trained ______
7. METL Task ______
8. Crew Endurance ______
9. Complexity ______
10. Weather ______
11. Additional Risk Factors ______

TOTAL ______

RISK VALUES: DAY/NIGHT MISSIONS

12. NVG Crew Selection/PC ______
13. NVG Crew Selection/PI ______
14. NVG Crew Selection/Add ______
15. Illumination ______
16. Moon Angle (NVG) ______
17. Additional Risk Factors ______

TOTAL NVG MISSIONS ______

TOTAL DAY/NIGHT MISSIONS ______

TOTAL RISK VALUE NVG ______

COMPUTATIONS DAY/NIGHT MISSIONS

Low Risk	<16
Medium Risk	16-28*
High Risk	>29**

COMPUTATIONS NVG MISSIONS

Low Risk	<25
Medium Risk	25-40*
High Risk	41-50**
Extremely High	>50***

* Medium-risk missions require approval of the company commander.

** High-risk missions require approval of the battalion commander.

*** Extremely high-risk missions require approval of the brigade commander.

ADDITIONAL COMMENTS

Fig. 6. Rotary-wing risk assessment matrix (cont.)

(US Army TC 1-210 [9])

The approach illustrated in Figures 5 and 6 has numerous potential benefits. It addresses some of the problems identified for subjective risk assessments by encouraging decision makers to consider a common range of threats. It may help to deter risk aversion by enumerating the grounds for delaying or abandoning a mission. However, it is essential to validate the risk values that are embedded within risk matrices. Considerable problems may arise if the consequent risk assessments are either too pessimistic or too optimistic. If they over-estimate the level of risk then commanders may be persuaded to consider hazards that have little impact on a potential operation. They may call upon resources that would have been better deployed on other operations. On the other hand, if the forms are too optimistic then commanders might accept hazards that threaten mission success and jeopardize the resources that are deployed to perform a particular operation.

The US Army has, therefore, used accident and incident data to help calibrate the risk assessments embedded within tools, such as that shown in Figures 5 and 6. The Army Safety Risk Management Information System and its various successors have helped to document previous adverse events. However, this raises further concerns because no record is, typically, kept of many successful operations. This undermines the sensitivity of the risk calculations. For instance, it might be observed that all previous recorded accidents involving the retrieval of rotary winged aircraft involved operations that included mixed deployments from a number of different units. It might, therefore, be concluded that this was a causal factor and that mixed deployments should be allocated a high risk factor for future missions. However, subsequent analysis could show that mixed deployments were present in the vast majority of *successful* missions as well. By focusing only on failure, there is a danger that risk matrices will undermine successful practices as well as those that characterize previous incidents and accidents. **Military risk assessments are usually validated by reference to the hazards that were realized in previous missions, this makes them overly conservative given that few records are maintained of successful operations where hazards were avoided.**

2.5 Risk Assessments Encourage Decision Transfer

Complex missions can be broken down into a number of activities using Mission Essential Task Lists. By summing the risk values for the hazards associated with each mission component, it is possible to form a partial ordering of those tasks that contribute most to overall risk. It is these high-risk sub-tasks that become the focus for risk reduction and mitigation. This relatively simple approach provides considerable flexibility. For example, an otherwise low risk mission might have a significant increase in the overall risk value if, for instance, one of the crews had less than 25 hours in the area of operation. Leaders might then intervene by introducing highly experienced crews into the operation. The overall mission risk is obtained by summing the hazards for each stage of the mission. The total can then be assigned to a particular risk level. For example, Figures 5 and 6 associate ‘Low Risk’ with risk values less than 16. Medium risk operations range between 16 and 28. High risk operations are associated with scores of 29 and above. In each case, commanders must seek additional levels of authorization before embarking on a mission. Company level approval must be provided for medium risk operations, while battalion commanders must support high risk plans. In this example, extremely high-risk operations associated with the use of night vision equipment must be approved at brigade level.

The opening sections of this book have identified decision transfer as a complicating factor in military risk assessment. This occurs when individuals and teams seek to pass on responsibility for complex, safety-related decisions. There is a danger that the codification of risk assessment procedures within organizational forms of decision making can provide an established mechanism for decision transfer. This is illustrated by the previous examples from the US Army. As mentioned above, the higher the level of risk derived from the individual risk factors then the higher up a decision must be referred in the chain of command. If a commander wanted to avoid responsibility for a particular operation then they could actively look to identify the highest risk factors from the tables presented in Figure 5. This would ensure a score that was sufficiently high for them to be justified in referring the decision up. **Military risk assessment techniques often explicitly identify the level in command hierarchy that must assume responsibility for different levels of risk. These architectures can be manipulated by individuals and units who want to transfer responsibility away from themselves towards senior command who may not be in the best position to accurately identify the hazards associated with a proposed mission.**

2.6 Risk Framing and the Hazards of Mitigation

Once leaders have identified critical tasks and hazards, such as crew inexperience, it is important to develop controls that reduce overall mission risks. FM3-04.513 argues that commanders should be presented with a series of options for risk control. Before presenting such a list, it is necessary for staff to consider any negative side-effects. For example, allocating experienced personnel to reduce the risks associated with a hazardous operation can increase the risks associated with other missions that might otherwise benefit from their participation. Similarly, deploying experienced personnel may reduce the opportunities to increase the skill set of other staff while increasing the levels of stress and fatigue on the crews who are allocated to the mission. US Army TC 1-210 urges commanders to think through the consequences of each potential risk control and then ‘visualize what will happen once the option has been implemented’.

The implementation of risk controls can involve changes to operations orders, standing operating procedures (SOPs), and training exercises. As might be expected, considerable emphasis is placed on communicating information about the purpose of controls, ‘from the commander down to the individual soldier’ so that any attempts to mitigate a risk is not inadvertently undermined. Similarly, commanders must take steps to supervise the application of risk controls. As with previous stages in the risk management processes advocated by the US Army, the superficial simplicity of the approach hides numerous detailed problems. For example, too close a monitoring may alienate staff, if they feel that their actions are under close supervision. Time may be wasted in providing evidence of controls to the point where supervision begins to undermine core mission objectives.

The opening sections of this book have argued that decision makers cannot enumerate all of the potential hazards that undermine complex, military operations. Instead, they learn to identify common threats that have affected previous situations. These framing judgments characterize recognition primed decision making and many contemporary theories about situation awareness. We use them to maximize scarce cognitive and perceptual resources. These constraints also affect our ability to assess the risks associated with particular mitigations. Having identified the threats posed by a mission, and then identified appropriate counter-measures, it is then difficult to find the time, expertise and commitment to conduct a further round of hazard analysis to identify the knock-on risks that might arise from the mitigating actions. **Most risk assessment techniques focus on hazard identification and prioritization with relatively little support provided for the analysis and planning of risks associated with the implementation of particular controls.**

2.7 Training Risks Decrease Operational Risks?

In order to prepare staff to make tactical decisions and execute complex plans under a wide range of environmental pressure, military organizations rely upon training and simulation. These exercises carry their own degree of risk: "Tough, realistic training conducted to standard is the cornerstone of Army war-fighting skills. An intense training environment stresses both soldiers and equipment, creating a high potential for accidents. The potential for an accident increases as training realism increases, just as it does in combat. The end result is the same; the soldier or asset is lost" [6]. In other words, there is a need to simulate risk. This creates tensions because it can be difficult to justify the use of hazardous training exercises that result in military fatalities, for example from the accidental discharge of weapons, or from military vehicles turning over during night exercises.

TC 1-210 emphasizes the need to minimize the differences between simulated and operational challenges during US Army training programs [9]. These differences can, however, be due to safety constraints. For example, the hazards of exposing troops to Multiple Launch Rocket System fire during training exercises can outweigh eventual mission benefits. Differences between training and operations may also be due to other practical constraints. For example there was insufficient time for all of the troops that were issued with Night Vision Devices during Desert Shield to train with those devices before deployment [8]. Each safety or functional constraint that creates differences between training and operations should be challenged. If possible, they should be removed to increase the veracity of the training program; 'With proper controls in place, these restrictions can be reduced or eliminated' [9]. If the constraints cannot be removed then they should be subject to risk assessment. **Operational effectiveness is perceived to depend upon the simulation of risk in training exercises that often result in accidents, this involves techniques that would not be acceptable in other safety-critical industries and which are often rejected as unethical by the general public.**

2.8 Military Training Encourages Risky Shift

Military training often encourages behaviors that run counter to many aspects of safety management. This includes an element of improvisation that is essential to respond to dynamic threats in a flexible and creative manner. The tension between creativity and safety can be illustrated by accidents during training to prepare troops for the hazards of improvised explosive devices (IEDs). US Army units have constructed 'makeshift' devices in pre-deployment training. In particular, several variants have been developed using ad hoc extensions to the M21 (Hoffman) Artillery Flash Simulator. This device is responsible for more explosives accidents and personnel injuries than any other simulator. The improvised IEDs often include flour mixtures with military grade munitions that can have extremely unpredictable results. The US Army Combat Readiness Center [10] observes that 'although their intentions are good, the risks associated with using homemade IEDs might be worse than the potential training benefits'. These devices contravene both Federal laws and Army regulations (eg AR385-63, Range Safety, paragraph 2-2). The development of homebrew IEDs devices can be explained both a desire to provide more 'realistic' exercises and also by the flexibility/creativity that is a primary aspect of military operations.

In retrospect, after an accident has occurred, it is far easier for personnel to identify the risks associated with military training exercises. However, one of the points behind many scenarios is to encourage the independence and mutual support that is based upon an acceptance of shared risks. For example, heat stress continues to affect many soldiers. 13 US Army personnel died from heat related injuries during 2005; there were more than 500 cases of heat stroke and 2,200 of heat exhaustion. Some of these injuries occur when individuals are forced to push themselves to the limit. However, the majority seem to occur from the internal motivation that arises when small teams work to meet shared training objectives. This is a particular form of 'risky shift' that must be identified by unit leaders. Commanders can, and have been, relieved of duty within the US Army when soldiers suffer from avoidable heat-related injuries. Even so, it can be difficult for individuals to raise concerns when they believe that undue risks are being taken during training exercises. **Military training often fosters a form of mutual support and a focus on mission success that often dissuades individuals who are best place to recognize the risks of training activities from raising their concerns.**

2.9 Availability Heuristic Undermines Risk Assessments

Previous sections have identified some of the limitations with FM3-04.513 and TC 1-201. In addition, it can be argued that these documents created arbitrary distinctions between the methods used to identify hazards both on and off duty. Such concerns led to the introduction of a new 'holistic' approach, known as Composite Risk management, within Army Field Manual 5-19 [11]. The dissatisfaction with previous methods was also reflected in the change of name from the US Army's

Safety Center to the new US Army Combat Readiness Center. This field operating agency located at Ft. Rucker, Alabama is the main agency for promoting operational risk management throughout the US Army. Senior staff leading this transformation summarized the need for change to CRM: "…the Army was still operating under a 1970's paradigm for safety, relying on lagging indicators, consequence management, and a compliance orientation… Mishaps behind the wheel accounted for nearly ¾ of the deaths in the past 2 years, the same proportion as reflected in the Army Safety Center's 1984 reviewPP… Thus, the (new) offensive on loss prevention has elements for the close fight and the deep fight. The plans consider the main effort (CRM in Army operations) and the flanks (CRM in off duty activities)" [12].

FM5-19 extended the scope of previous guidance by arguing that all Army personnel should be trained in the principles of risk management. In consequence, Composite Risk Management doctrine has been institutionalized in the Risk Management Chain Teaching program created by the Chief of Staff of the Army [13]. Under FM5-19 a hazard is interpreted to be any "condition with the potential to cause injury, illness, or death of personnel; damage to or loss of equipment or property; or mission degradation". They include "a situation or event that can result in degradation of capabilities or mission failure" [11]. However, the holistic nature of Composite Risk Management introduces important differences that reflect the US military concern to look after personnel on and off-duty. Hazards are defined to exist in all environments, including but not limited to "combat operations, stability operations, base support operations, training, garrison activities, and off-duty activities". The revised field manual also advocated the METT-TC mnemonic (Mission, Enemy, Terrain and weather, Troops and support available, Time available, and Civil considerations). METT-TC can be used to identify hazards irrespective of whether units are on or off-duty, ranging from terrorist attacks through to drunk driving.

One of the aims behind FM5-19 was to increase military awareness about the potential for terrorist attacks away from base. In addition, here has been a realization that risk assessments must be used to address hazards created by new forces threatening military personnel. Those operational risk assessments that have been developed within military organizations tend to focus on conventional enemies and do not consider the risks associated with terrorist attacks on off-duty personnel.

The first chapter has described how the availability heuristic creates systematic biases within risk assessments for such rare events. This heuristic explains how individuals predict the frequency of an event based on how easily an example can be brought to mind. This leads to systematic biases in military organizations that disseminate information about major mishaps that are relatively rare, including terrorist attacks, compared to more frequent failures, such as road traffic collisions, that are less well publicized. In such circumstances, individuals are likely to overestimate the probability of rare events and under estimate the likelihood of more frequent mishaps.

Composite Risk Management supports both on and off-duty safety assessments, however, the focus on rare and on frequent events has not addressed the biases created by the availability heuristic and may have hindered progress against some of the more 'mundane' causes of military mishaps, especially road traffic accidents.

2.10 Unthinking Enthusiasm Weakens Risk Assessment

Composite Risk Management doctrine emphasizes that commanders must identify those enemy capabilities that pose hazards to an operation or mission. Key to this hazard analysis of enemy capacity is the Intelligence Preparation of the Battlefield (IPB). The IPB is intended to support 'threat based risk assessments by identifying opportunities and any constraints the battlefield environment offers to both enemy and friendly forces' and hence must explicitly capture 'enemy capabilities and vulnerabilities' [11]. However, FM5-19 also recognizes the temporal constraints that create considerable pressures for the field commanders who must make key tactical decisions. The more considered quantitative approaches to likelihood and consequence assessment are ill suited to the rapidly changing context that many commanders must address: "In these situations, they perform hasty risk assessments. A hasty risk assessment may be performed mentally. It may be transmitted verbally or in writing via a FRAGO (Fragmentary Order)…Only the essential information necessary to complement the FRAGO and forward the risk guidance received from the battalion commander are included. As in the example, an overlay may be included with the risk assessment to clearly portray the location of hazards. The hasty risk assessment (can be) a separate document. However, it may be included within the FRAGO issued by the company to the platoon" [11]. The time limited nature of these situations and the critical nature of their decisions makes it essential that FRAGOs are successfully communicated to their intended recipients. FM5-19, therefore, provides detailed guidance on how hazard assessments can be passed in annotated form within these orders. The integration of 'ad hoc' risk assessments in fragmentary orders again illustrates the holistic approach advocated in the new Field Manual. Even where time is strictly limited, commanders should explicitly take the opportunity to consider potential hazards as part of the Military Decision Making Process.

There is a concern that attribution bias will impair the critical and unbiased assessment of risk assessment initiatives across the US military [14]. Attribution bias refers to inferences that are made by observers often with the benefit of information or resources that were not available to the individuals involved in an incident. This can be illustrated by a recent accident report that describes how two M1A2 Abrams tanks were assigned to escort an explosive ordnance disposal (EOD) team to an enemy weapons cache site. Neither the tank crews nor the EOD team was familiar with the location of the weapons cache. Maps and imagery provided insufficient detail to plan the mission. A process of trial and error led them the cache and the EOD team completed their task after dark, around 18:45. The leaders decided to return using the route over a sandy, clay road that ran alongside a canal. The trail tank crossed a bridge over the canal and turned right over a berm. It's rear began to shake violently and the track commander (TC) told the driver to go left as the right edge of the road collapsed under the tank's weight. The crew heard the TC announce "rollover, rollover, rollover" as the tank overturned into the water-filled canal. The TC's death was attributed to blunt-force trauma suffered during the rollover and a lack of oxygen after the tank settled in the water. The subsequent investigation identified two primary causes: a failure to adequately plan the mission and a failure to execute proper rollover procedures because the TC did not immediately drop inside the turret. Attribution bias can be seen in the commentary that accompanies the account of this accident: "Had the tank crews used CRM when they were trying to identify alternate routes, they might've realized the hazards they faced on the unimproved roads they ultimately selected. This instance wasn't the first time a canal road collapsed under a tactical vehicle in theater; similar roads have caved in under vehicles weighing far less than an M1 tank, including HMMWVs. The bottom line is every Soldier must take into account all the hazards, both tactical and accidental, that can hurt or kill them or their buddies. We need each one of you, so use CRM to stay ready and Own the Edge!" [15]. The key term here is 'might' – without significant additional operational experience in the application of CRM, considerable questions must remain as to whether the ad hoc risk assessments recommended in FM5-19 could really have helped the leaders and their crews to identify the hazards at the end of a long day, filled with other earlier missions as they made their way back to base through a potentially hostile environment. **The enthusiasm for novel military risk assessment techniques may create a hostility and cynicism amongst those personnel who are faced with the application of simple risk assessment techniques under complex, time limited constraints with incomplete information.**

2.11 Tactical Vulnerabilities from Military Risk Assessments

Does the dominance of risk assessment create any concerns? The concerns identified in this chapter capture problems in using these techniques to inform strategic, tactical and operational decision making in military organizations. Many approaches, including the Enterprise Risk Assessment Model and Composite Risk Management, are relatively novel. Their impact on operational effectiveness has yet to be studied; many units are still being trained in the doctrines embodied in FM5-19. Significant questions remain to be answered. For example, it is unclear whether there will be systematic biases in the risk assessments conducted by units with different operational backgrounds, including both full time and National Guard units.

Previous sections have described how lightweight risk assessments are to be integrated into FRAGOs (Fragmentary Orders) when leaders must make complex decisions against hard deadlines. The Composite Risk Management proposals are also intended to ensure that these assessments are communicated to the units involved in particular mission components. The precise format for both the FRAGOs and the communication of risk based decisions must be tailored to the particular situations facing individual units. Only time will tell if this emphasis results in the development of appropriate tools and techniques that can be used in the field.

Many military staff are pre-selected and then trained for decision-making characteristics that are very different from those in the civilian population. There seems to be very little direct evidence that CRM techniques will be able to compensate for the risk preference biases that are often seen in military personnel. This concern can be illustrated by an account from the US Army's Countermeasure safety publication in which the author describes the risk seeking nature of many soldiers and then raises an, as yet, unsubstantiated hope that Composite Risk Management will address some of the consequences in military activities: "Have you ever deliberately put yourself in a situation you didn't think you'd get out of alive, only to survive and vow never to do the same thing again? … Playing football on a semi-thawed lake, passing traffic uphill in a no-passing zone, driving drunk and boating in a lightning storm—none of these are sound decisions, but I've done them all. When you're young, it's hard to distinguish risk from what we perceive as adventure… We can step back and make smart decisions, which is the beauty of Composite Risk Management. Even in combat, Soldiers of all ranks have the authority to stop unsafe acts and implement controls to ensure everyone makes it home from the fight. Please take advantage of this great tool and apply it to everything you do, especially if you see some idiot pulling charges out of a powder pit!" [16]. Such assertions arguably underestimate the problems of 'groupthink' and 'risky shift', introduced in Chapter One, that can affect team-based decision making in combat operations.

One of the most vibrant areas of research within risk management and decision theory has focused on the development of models that explain opponents' behavior in various forms of games. These models assume that competitors make complex decisions with uncertain outcomes in order to maximize their returns while minimizing the rewards for competitors. Considerable benefits are to be gained if one player understands the decision making processes employed by their opponent. These theoretical outcomes have direct applications in the military domain. For example, TC 1-210 and FM5-19 stress that all 'unnecessary risks must be avoided'. This enables opponents to make direct inferences about the risk adverse behavior of the US military. These insights are, arguably, being applied by the insurgents' use of IEDs and snipers. In this case, the opponents are reacting on the basis of direct observations of risk-based decision making in the field. In the future, however, opposing forced could make strategic decisions based directly on the risk averse statements in public documents such as FM5-19. **If risk assessment techniques, such as Composite Risk Management, offer the benefits that are claimed in terms of encouraging consistent decision making across the US military then there is a danger that they will provide opposing forces with a 'blue print' for US military operational decision making.**

2.12 References for Chapter Two

[1] US General Accounting Office (GAO), High-Risk Series: An Update, Technical Report GAO-05-207, Washington DC, USA, 2005.

[2] US Department of Defense, FAQ: Enterprise Risk Assessment Methodology (ERAM), Technical report, Defense Business Transformation Unit, Washington DC, USA, April 2010. http://www.dod.mil/dbt/faq_eram.html

[3] Defense Contracting: DOD Has Enhanced Insight into Undefinitized Contract Action Use, but Management at Local Commands Needs Improvement , GAO-10-299 January 28, 2010

[4] US Government Accountability Office (GAO), Information Security: Computer Attacks at Department of Defense Pose Increasing Risks, Technical Report GAO/AIMD-96-84, Washington DC, USA, 2006.

[5] US Government Accountability Office (GAO), Business Systems Modernization: DoD Continues to Improve Institutional Approach, but Further Steps Needed, Technical Report Washington DC, USA, 2006.

[6] US Department of the Army, FM 3-04.513: Battlefield Recovery and Evacuation of Aircraft, Headquarters, Washington, DC, 27 September 2000. http://www.army.mil/usapa/doctrine/Active_FM.html

[7] C. Dandeke, On the need to be different: military uniqueness and civil-military relations in modern society. Royal United Services Institute (RUSI) Journal, 146, 4-9, 2001.

[8] C.W. Johnson, The Role of Night Vision Equipment in Military Incidents and Accidents. In C.W. Johnson and P. Palanque (eds.) Human Error, Safety and Systems Development, Kluwer Academic Press, Boston, USA, 1-16, 2004.

[9] US Department of the Army, TC 1-210: Aircrew Training Program Commander's Guide to Individual And Crew Standardization, Headquarters, Department Of The Army, Washington, DC, 3 October 1995.

[10] US Army Combat Readiness Center 2006. Simulated IEDs: Real Problems, Countermeasure, 27:06/06: 10-11, 2006.

[11] US Department of the Army, FM 5-19: Composite Risk Management, Headquarters, Washington, DC, August 2006.

[12] J.A. Smith and B.R. Yaeger, Tansforming Army Safety: Enhance Combat Readiness through Composite Risk Management, Technical Report, US Army Combat Readiness Center, 2007.

[13] US Army Combat Readiness Center 2007, Composite Risk Management Chain Teaching Training Packet. US Army Combat Readiness Center, April 2007.

[14] C.W. Johnson, A Handbook of Accident and Incident Reporting, University of Glasgow Press, Glasgow, Scotland, 2003. http://www.dcs.gla.ac.uk/~johnson/book

[15] US Army Combat Readiness Center 2006. Loss Investigation: A Map but No Direction, Countemeasure, 27:09/06: 14-15, 2006.

[16] R. Andree, Personnel Injury: Great Flying Stoves! US Army Countermeasure, Centre for Combat Readiness, Volume 27:10/06:10-11, October 2006.

3 Military Incident and Accident Reporting

Chapters one and two have identified some of the differences between civil and military risk assessment. Many of the factors that complicate military risk assessment stem from the active intervention of enemy forces that do not, typically, form part of more conventional hazard assessments. Other problems arise because many aspects of military life actively encourage risk taking in order to achieve tactical advantages or transfer effects between training and operations. Finally, there is a danger that opposition forces may make inferences about any organization that adopts a risk-based approach to military planning. In contrast, this chapter looks at the difficulties that arise when military organizations try to validate the products of any risk assessment. In particular, the following pages identify a host of further problems that arise when mishap reporting systems are used to provide feedback on the hazards and threats of military operations.

3.1 The Benefits of Incident Reporting for Military Risk Assessment

The following sections explain why incident and accident reporting systems are often integrated into military risk assessment techniques. They provide important means of validating the hazards that might be identified in previous missions. Mishap reporting systems can also be used to inform military personnel of the consequences and likelihood of previous incidents as a means of informing operational decision making. Subsequent sections go on to question the extent of these benefits.

3.1.1 Incident Reporting, Risk Assessment and Safety Management

Risk assessment does not occur in isolation. It, typically, forms one component within a wider Safety Management System (SMS). SMS provide a framework that integrates and coordinates the implementation of safety policy across complex organizations. A number of different organizations have provided detailed guidance on the implementation of Safety Management Systems [1, 2, 3]. However, common processes can be identified across many of these schemes:

- developing safety policy, which sets out the organizations' general approach, goals and objectives towards safety issues;
- planning and resource allocation, the organizational process which is used to determine the methods by which specific objectives should be set out and how resources are allocated. This includes establishing the structures, responsibilities and relationships that shape the total working environment;
- implementation, which focuses on the practical management actions and the necessary employee behaviors that are required to achieve success;

- auditing and review, which incorporates the process of gathering the necessary information to assess progress towards safety goals.

Figure 7 provides a graphical overview of the different phases with a Safety Management System. As can be seen, high-level safety policy is drafted to justify the allocation of resources and the planning of safety activities. These are then implemented across an organization. However, audits and reviews must also be conducted to provide feedback on whether or not the objectives in a safety policy have been met. This evaluation phase can trigger subsequent iterations by encouraging revisions in the high level safety policy.

Figure 7 also illustrates the way in which risk assessment contributes to the allocation of resources within a safety management system. As we have seen, it is intended to help engineers and management identify those hazards that are likely to have the most significant consequences for safe and successful operation. However, previous chapters have identified a range of complicating factors that can affect the accuracy of risk assessments. These include psychological processes, such as fear and risk aversion , as well as group factors, including risky shift. For new systems or novel applications of existing technologies and processes, there may not be any objective or empirical data about the likelihood of potential hazards. In such cases, it is important that the audit and review phase of Figure 7 is used to provide feedback on the accuracy of the risk assessments that guide resource allocation.

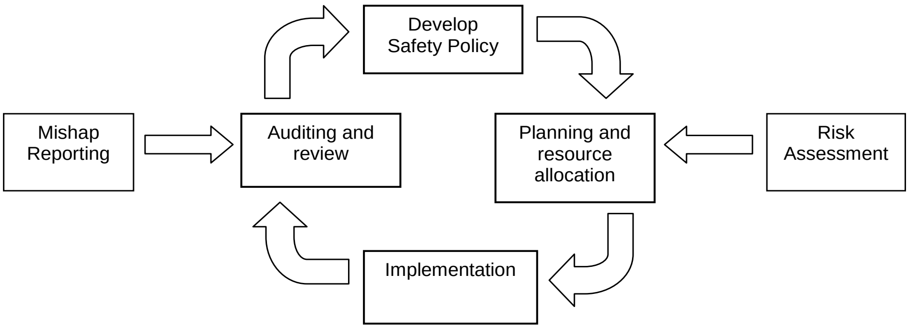


Fig. 7. Relationship between Incident Reporting and Risk Assessment

Incident reporting schemes offer a number of potential benefits within a safety management system. In particular, they can help to guide the allocation of finite resources to those areas of an application process that have proven to be most problematic in the past. In other words, incident reporting systems can focus risk assessment techniques using 'real world' reliability data that can be radically different from the results of manufacturer's bench tests.

Incident reporting systems have provided an important means of learning about ‘real world’ hazards in many safety-critical applications. For instance, the UK operates a

Confidential Human Factors Reporting System across the aviation and maritime industries[2]. The NASA Safety Reporting System provides means of reporting safety related concerns from across their diverse operations[3]. There are a range of similar resources across the healthcare industries. These include the Australian Patient Safety Foundation[4], the US National Patient Safety Foundation[5] and the UK National Patient Safety Agency[6]. However, many of these reporting systems face common problems. It can be difficult to persuade individuals to provide feedback on safety concerns if this might lead to disciplinary action. Many reporting systems gather information about well known problems, it is often hard to ensure that organizations allocate sufficient resources to ensure the effective implementation of subsequent recommendations/mitigations. Military incident reporting systems face a number of additional problems. For example, the command structure can make it difficult to sustain promises of confidential, blame-free reporting. This is a significant barrier because other systems have identified such assurances as a prerequisite for the development of trust in any reporting application [4].

There are further differences between military and civil reporting systems. For instance, military organizations recruit young people and place them in what are often high-risk situations. These individuals are expected to make highly complex decisions, often involving a range of different technologies. In such circumstances, it can be very difficult for soldiers to diagnose the detailed causes of any particular system failure. Military personnel are also expected to work in teams and to coordinate their activities with those of their colleagues who may be collocated or who may be hundreds of miles away. This creates problems when it may be unclear which individual or group should be responsible for submitting a safety concern. Military personnel face physiological and cognitive stresses that are seldom seen in any other domain. In such circumstance, it may be unrealistic to expect soldiers to find sufficient time away from their primary tasks to complete an incident reporting form. Equally, however, military personnel must often face hours of inactivity or of repetitious training that induces fatigue, boredom and inattention. This creates a concern that reporting systems will elicit spurious reports that have little relationship with operational threats and hazards.

3.1.2 Feedback of Lessons Learned to Military Risk Assessment

Military reporting systems contribute to the wider processes of risk management, just as civil systems provide feedback on hazards in other domains. The US Army's Accident Investigators' Handbook argues that the objective is to provide "information that the Director of Army Safety can use to immediately effect changes at Department of Army (DA) level, but will also provide information necessary to identify Army-

[2] http://www.chirp.co.uk/, all URLs referenced in this book were last accessed July 2010.

[3] www.hq.nasa.gov/nsrs/

[4] http://www.apsf.net.au

[5] http://www.npsf.org

[6] http://www.npsa.nhs.uk/

wide hazards and controls" [5]. This information is to guide the future application of the Composite Risk Management techniques introduced in Chapter Two.

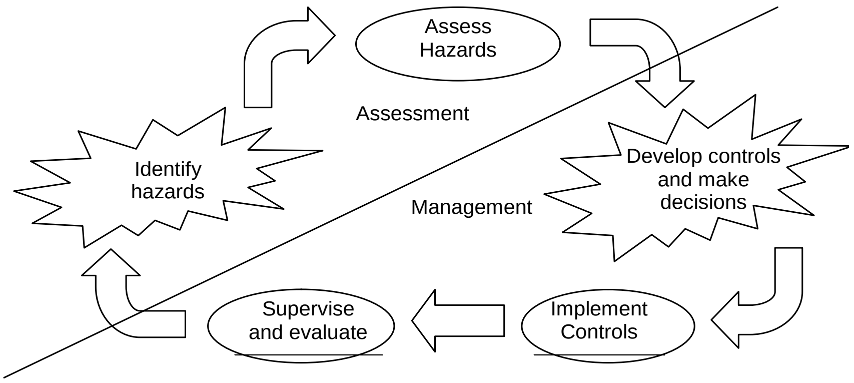


Fig. 8. US Army Composite Risk Management Processes

Figure 8 illustrates the main processes in CRM and can usefully be compared with the more generic processes of Safety Management Systems in Figure 7 [6]. As can be seen, hazard analysis contributes to the assessment of risks and the identification of controls. The effectiveness of these mitigations must be validated by a range of activities including debriefs, lessons learned and post action reviews. Accounts of previous adverse events or near miss incidents can directly inform the identification of potential hazards. For example, the US Army eleased a safety notification when its reporting system showed that drowning and water-related mishap were occurring at almost twice the anticipated rate. An analysis of the incidents revealed that most of the incidents stemmed from the incorrect handling of small boats. This triggered an analysis of the previous decade's incident statistics. Over the Summer period there were 141 reported incidents, some involving more than one fatality. Only 1 death occurred at a swimming pool with Army lifeguards. Most incidents occurred in open water or near the shoreline. Lake and river recreation produced 41% of the drowning incidents. Ocean swimming fatalities accounted for a further 16%. Military training operations accounted for 11% of the drownings. 9% drowned after vehicle accidents [7]. The significance of these statistics is that they can be used to promote awareness of the actual risk factors that contributed to previous adverse events. They also reinforce the central message behind CRM's insistence that military personnel learn to use the same techniques to identify the hazards both on and off duty.

As can be seen from Figure 8, the US Army relies upon Composite Risk Management techniques to identify the controls that are intended to mitigate particular risks. These controls fall into a number of broad categories:

- "Educational (awareness) Controls. These controls are based on the knowledge and skills of units, organizations, or individuals. It includes their awareness of the hazard and control. Effective educational control is implemented through individual and collective training that ensures performance to standard.

- Physical Controls. These take the form of barriers and guards or signs to warn individuals, units, or organizations that a hazard exists. Special controller or oversight personnel also fall into this category.
- Avoidance/Elimination Controls. These controls include positive action to prevent contact with an identified hazard or the total elimination of the hazard" [6].

Accounts of previous mishaps often describe ways in which controls failed to protect military personnel and their equipment. This enables safety officers to refine the measures that are intended to prevent the causes or mitigate the consequences of adverse events. For instance, a recent incident report described how a soldier fell while attempting to negotiate an 'inverted' rope descent. Previous training related incidents had led to the development of US Army standard FM 21-20. This requires that the obstacle should include a platform at the top of the tower for the instructor and the student. A safety net should also be provided. This standard also requires that the obstacle should be constructed to reflect the Corps of Engineers drawing 28-13-95. Unfortunately, this diagram did not include a safety net or a platform. The incident investigators, therefore, concluded that 'confusion exists concerning the proper design and construction of this obstacle' and the diagram was revised to remove any potential inconsistency. The incident not only helped to reinforce the need for the physical protection of the safety net, it also helped to clarify the procedures and guidance that were intended to ensure that a net had been provided in the first place [8].

The insights provided by accident and incident reports are of limited value unless they are communicated back to the units and commanders who must use them to inform future risk assessments. It is for this reason that many military organizations have implemented lessons learned applications. These include web sites as well as paper-based newsletters that alert military personnel to the consequences of previous incidents, as well as providing frequency statistics such as those presented in previous paragraphs for water-related fatalities.

The US Army's Picatinny Arsenal reporting system illustrates the human factors and other operational insights that can be gained by these applications [9]. The Arsenal maintained a lessons learned system for technicians working on the 155mm M109A6 self-propelled howitzer, known as the Paladin. A series of problems stemmed from situations in which the driver's hatch was opened with 'extreme' force. This caused the pin that secures the drivers hatch-stop to break. This defect could have rendered the vehicle inoperable according to the Paladin Operators Manual (TM 9-2350-314-10). Rather than follow this procedure or order new parts, many technicians applied ad hoc 'field solutions' using poorly fitted cotter pins, nails, and similar devices. These fixes resulted in more sustained damage to the expensive hatch stop assembly, which in extreme cases became totally inoperable. The Army's analysis of previous incidents, therefore, advocated that each unit hold a small number of spare Grooved Pins at a cost of $1.78 each. This simple measure was intended to reduce the need for ad hoc maintenance procedure and thereby avoid the safety risks associated with damaged hatch assemblies. This case study is instructive because it illustrates the type of ad hoc behaviors that characterize many military operations. There is a a strong desire to use creativity and lateral thinking to meet mission objectives, even if they may undermine safety. Incident reporting and lessons learned systems are critically important because they capture these ad hoc behaviors that would otherwise never appear as an explicit hazard within more formal risk assessments. Chapter Four will retrun to this issue in describing the problems of using military hazard analysis to identify the impact that 'degraded modes' of operation have upon many safety-critical systems.

Military reporting systems remind personnel of the hazards that can arise in a range of different contexts. For example, the Canadian Defense Force Digest is available in both printed form and on-line. It provides members of the military with information about common hazards and recent accidents. One article described an electrocution incident. This publication Co-workers exposed themselves to considerable risk in trying to rescue their injured colleague. The editors of the digest concluded by reminding staff that electrical canes can help to reduce the risks involved in isolating electrocution victims. This incident not only shows that role that reporting systems can play in reminding staff of hazards and counter-measures, it also illustrates the way in which they help to open dialogues across an organization. Engineering staff criticized the recommendations made by the editor of the Defense Force Digest. Canadian national rules require that trained personnel should be no closer than 0.9 meters from the victim during incidents involving 425V to 12,000V. This would prohibit the use of electrical canes. The correspondent argued that the focus should have been less on rescuing the injured person and more on the prevention of such incidents; "unless management creates a safety culture based on risk management and unless supervisors instill this workplace ethos in their workers… and then enforces this view consistently, we will never break the chain and accidents will continue to occur" [10].

3.1.3 Insights into Patterns of On-Duty and Off-Duty Hazards

Many of the incidents reported to military systems are similar to those that are elicited by civil applications. The previous incident could have been reported in the process industries or in power generation. There are, however, some important differences. In particular, military reporting systems reflect an extended 'duty of care' that is not present in other domains. For instance, many army systems analyze road traffic accidents involving their personnel even when they are off duty. In most commercial systems, these incidents would be left to the police. This additional level of involvement not only reflects the duty of care that military organizations extend to their personel; it also reflects an appreciation of the particular demands that are created by life in the forces. For instance, the geographical distribution and rotation policies employed by armies often leads individuals to travel enormous distances to meet family and friends. A US Air Force commander recently described how one of their staff attempted to drive 2,000 miles for a weekend reunion. The inclusion of off-duty incidents which military reporting systems is also consistent with the holistic approach to risk management, advocated within the US Army's CRM methodology.

The particular characteristics of military life not only affect the circumstances that surround more familiar adverse events, such as road traffic accidents. They also affect the recommendations that are made in the aftermath of an adverse event. For example, a report on road traffic accidents involving military personnel pointed out that the fatality rate for car passengers increased by 1.1% for every 100lb decrease in the weight of the car. Occupants of a lightweight car are, therefore, four times as likely to die in a collision with a Sport Utility Vehicle (SUV) than the occupants of the SUV. The recommendations made from this investigation caused one correspondent to question the underlying argument. He responded, "…by this logic, our family car should be a multi-wheeled armored fighting vehicle" [11].

The inclusion of on-duty and off-duty incidents within the same reporting system can help to identify some surprising parallels. For example, the US Army investigated an incident in which an Abrahams tank (M1A1) was overturned. The outcome was exacerbated because some of the occupants were not safely secured within the vehicle. The report has strong similarities with many recent investigations into the lack of seatbelts in civilian road traffic accidents; "once again, human error became a contributing factor in the loss of a soldier. Leaders must ensure that they and their crewmembers are positioned correctly in their vehicles and are taking advantage of all safety features. The nametag defilade position increases your ability to lower yourself safely inside the vehicle and prevents excessive exposure of body parts to the elements outside" [12]. Unfortunately, incidents involving incorrect crew positions within military vehicles are almost as common as incidents in which drivers did not wear their seatbelts. The following month, the US Army Safety Center received a report of another incident in which a soldier was crushed by an M551A1. This overturned after being accidentally driven into an excavated fighting position. The injured soldier was standing in the hatch above the nametag defilade position.

3.1.4 Feedback on Risk and Decision Making

Not only do incident reporting systems provide military personnel with feedback about the likelihood and consequences of potential hazards. They can also be used to illustrate some of the higher level factors that complicate military risk assessment. Many of these were introduced in the opening chapter of this book. For instance, it might be argued that the accidents, cited above, provide examples of 'target levels of risk' or risk homeostasis. The over-engineering of vehicles such as the M1A1 and the M551A1 provides soldiers with an impression of safety that they then trade against performance. They will perform more dangerous maneuvers and expose themselves to greater risk because they are confident that the vehicle will protect them against the adverse consequences of any hazard. The two previous incidents illustrate the dangers of such a view. Unfortunately, the human factors analysis of risk homeostasis provides few insights that can be used to avoid future incidents. Without these insights we are as likely to eliminate crush injuries from nametag defilade incidents, as we are to persuade everyone to wear seatbelts. This illustrates the mutual dependencies between incident reporting and risk assessment. In other words, risk assessment techniques are unlikely to provide accurate insights about potential hazards if they are not informed by information about previous adverse events. Conversely, incident and accident reporting systems are of limited value if they cannot be used to inform subsequent decision making.

In the previous examples, the soldiers were performing maneuvers that the vehicles had been designed to execute. They had been informed of the correct seating positions and yet still exposed themselves to the risk. Such incidents illustrate the tension at the heart of military risk assessment. To what degree should we attribute responsibility to individual operators? To what degree should we attribute responsibility to those that manage the personnel involved? If we focus on the individual operator then we may miss the wider systemic factors that influenced their behavior. Equally, the focus on management responsibility often makes unrealistic assumptions about military organizations' ability to ensure personnel follow safe operating procedures in the face of dynamic demands and complex problem solving situations.

3.1.5 Insights into Safety Culture and Military Justice

Incident reporting systems can be analyzed for common patterns between incidents that suggest wider problems within a military origination. Many mishaps stem from flawed 'safety culture'. Safety culture has been described as *the way safety is done around here*; emphasizing that it is concerned with the realities of safety, and not what people say *should* be done. The US Federal Aviation Administration defines safety culture to be the "product of individual and group values, attitudes, competencies, and patterns of behavior that determine commitment to, and the style and proficiency of, an organization's safety health and safety management" [13]. They identify four principle components:

1. A reporting culture encourages employees to divulge information about all safety hazards that they encounter.
2. A just culture holds employees accountable for deliberate violations of the rules but encourages and rewards them for providing essential safety-related information.
3. A flexible culture adapts effectively to changing demands and allows quicker, smoother reactions to off-nominal events.
4. A learning culture is willing to change based on safety indicators and hazards uncovered through assessments, data, and incidents.

Figure 9 provides a high-level overview of these components. As can be seen, the reporting culture can be associated with the implementation of the feedback mechanisms described in this chapter. The learning culture refers to techniques and processes that include risk assessment. A flexible culture captures aspects of the implementation mechanisms that must support the development and monitoring of recommendations, derived both from incident reporting and from hazard mitigation.

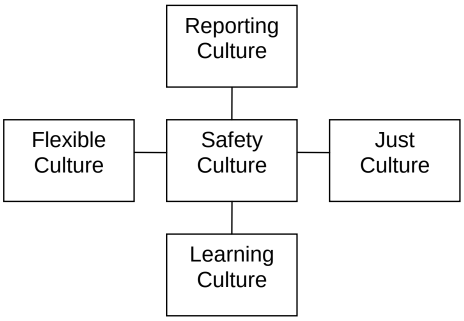


Fig. 9. Four Principle Components of Safety Culture

The US General Accounting Office monitored the implementation of recommendations following a series of Army Ranger training incidents [14]. They identified problems in the implementation of those recommendations but also in the way in which they were drafted in the first place. For example, one recommendation required that the Army develop 'safety cells' at each of the three Ranger training bases. These were to include individuals who had served long enough to develop experience in each geographic training area so that they understood the potential impact of weather and other local factors on training safety. However, the National Defense Authorization Act that embodied these provisions did not establish specific criteria on the makeup of a safety cell. The General Accounting Office concluded that the approach chosen by the Army 'represents little change from the safety oversight practice that was in place' at the time of the incidents. In other words, they lacked the 'flexible culture' identified in Figure 9. They also found problems in the 'learning culture' with deficiencies in the conduct of annual site safety inspections. Those inspections that took place often-focused on 'checklists of procedural matters', such as 'whether files of safety regulations and risk assessments are maintained' rather than on monitoring the effectiveness of recommendations after incidents have been analyzed.

A 'Just Culture' is the fourth element identified in Figure 9. The tension between identifying systemic causes and punishing individual violations can be seen in the relationship between reporting systems and the mechanisms of military justice. For instance, the US Army's Accident Investigators Handbook states that "First, historically, human error causes approximately 80% of all accidents. Second, identifying human error is the least objective of all the causal factors. Third, human error is often present in accidents caused by environmental factors and materiel failures. Finally, the complex nature of human behaviour and organizational culture mandates a systematic approach to investigations to ensure that all areas are thoroughly addressed" [5]. This more enlightened view can be contrasted with the legal provisions that govern military courts. For instance, Chief Justice Lamer of the Supreme Court of Canada explained in R. v. Généreux in 1992 that "Breaches of military discipline must be dealt with speedily and, frequently, punished more severely than would be the case if a civilian engaged in such conduct. As a result, the military has its own Code of Service Discipline to allow it to meet its particular disciplinary needs. In addition, special service tribunals, rather than the ordinary courts, have been given jurisdiction to punish breaches of the Code of Service Discipline. Recourse to the ordinary criminal courts would, as a general rule, be inadequate to serve the particular disciplinary needs of the military. There is thus a need for separate tribunals to enforce special disciplinary standards in the military".

3.2 The Limitations of Incident Reporting for Military Risk Assessment

The previous section has identified a number of benefits that can be obtained from the integration of military risk assessment techniques and of mishap reporting systems. These two different approaches combine prospective hazard analysis with direct feedback on the frequency and consequences of previous failures. They can also be used to provide information about the influence of risk taking behavior, including target levels of risk, on military operations. However, a number of limitations affect this integrated approach to safety management.

3.2.1 On the Need to Support Proactive Decision Making

It is important to stress that incident and accident reporting systems examples of a wider selection of techniques that can be used to gain feedback on military risk assessments. For example, the US Army's Composite Risk Management initiative also recommends 'spot-checks, inspections, situation reports (SITREPs), back briefs, buddy checks, and close oversight' [6]. After action reviews (AARs) are also intended to provide a forum in which different units can assess the accuracy and effectiveness of prior risk assessments; "Based on evaluation and assessment of the operation and the effectiveness of CRM, lessons learned should be developed and disseminated to others for incorporation into future plans, operations, and activities... lessons learned from the CRM process, to include CRM worksheets, are captured and retained for use during future operations". One of the reasons why the US Army have looked beyond incident and accident reporting is that previous systems were perceived to place too much emphasis on gathering data about past failures and not enough emphasis on predictive hazard analysis. This concern is reflected by the manner in which the Centre for Army Lessons Learned was renamed the US Army Centre for Combat readiness as part of the transition toward CRM.

The Canadian army's review of their participation in the NATO Implementation and Stabilization Force in Bosnia-Herzegovina provides a further example of the need to ensure that reports about previous adverse events can be used to guide future decision making. "Many units stated first aid training packages lack realism and should be oriented to injuries sustained in combat. IV and morphine training were essential... During 6 months in theatre, no soldier gave artificial respiration, treated a fracture or did a Heimlich maneuver. However, they did treat 17 bullet-wound cases, 3 shrapnel-wound cases and 7 minefield cases (foot or leg amputated)". At first sight, the lessons learned review might therefore suggest that future peace keeping rotations in conflict areas should be provided with training in the treatment of major trauma. However, further analysis revealed that "As the threat level dropped for latter rotations, unit comments on the need for IV and morphine training waned, there seems to be much debate on the usefulness and dangers of teaching this subject. All unit medical staff strongly recommended that it not be completed because of the inherent dangers that administering IVs or morphine entails…" [15]. These comments illustrate the complex problems that military operations pose for risk assessment. Incident reports often cannot be taken at face value. In this case, the introduction of trauma training for many troops might have increased the risk to the civilian population from iatrogenic injury. The troop's perceived operational need for IV and morphine training documented in the incident reports cannot easily be balanced against the potential dangers of inappropriate use during subsequent operations. This analysis reiterates comments made in Chapter Two that military risk assessment cannot stop with the identification of mitigations. It is often the case that the interventions identified to address previous incidents will give rise to new and unintended hazards.

3.2.2 What Should be Reported?

In many cases, it is obvious whether or not an accident should be reported; "The unit was engaged in a river crossing operation when the decision was made to float downstream. Even though current readings had not taken place, a safety boat was not on standby and an exercise participant was not wearing a flotation device, the squad decided to proceed with the mission. The rivers current was strong enough that it pulled all the team's elements under an anchored barge. Some of the team members survived, but two of them did not. Leaders must re-emphasize when encountering an unsafe situation, the mission must now become safety" [16]. In other situations it can be far less clear whether or not an incident or near miss ought to be reported. Some organizations provide guidance by listing the types of adverse event that should be submitted to their systems. This leads to missions when significant safety concerns are not included on the list. Other organizations identify a reporting threshold in terms of consequential loss. A report may be required if particular items of equipment are damaged or if the repair costs of any damage exceed a particular threshold value. The task of making such assessments is complicated if the costs of an incident are taken to include environmental impact. Similarly, it can be difficult to account for injuries and diseases or psychological adverse effects using such financial thresholds. Most organizations, therefore, define additional reporting criteria to explicitly list the types of physical harm that should trigger an incident report.

Outcome measures cannot be directly used to assess the criticality of near-miss incidents because nothing has been damaged. The fact that an adverse event was avoided forces investigators to make crude estimates of the 'worst plausible outcome'. Similarly, there may be certain types of adverse event that should be investigated even though they resulted in outcomes that would not normally be serious enough to warrant an investigation. For instance, public anxiety over previous injuries to recruits or conscripts during initial training has persuaded many armies to devote additional resources to the investigation and analysis of these mishaps. Some military organizations hold review boards to discuss the circumstances surrounding an adverse event or near miss before deciding whether or not it should be analyzed in greater detail. This approach creates problems when inconsistent decisions are made to investigate some mishaps but to ignore other similar events. Most military systems rely upon a compromise approach that publishes a list of 'typical incidents' but also allows considerable scope for individual discretion.

The scope of a reporting system can also be determined by the resources that are available to administer the scheme. In some systems, the same procedures are used to report everything from minor equipment malfunctions up to the loss of aircraft to enemy action [4]. A significant level of investment is required to filter the mass of incident information to derive appropriate priorities for further intervention. This creates problems because if we set the criteria too low we can be swamped by reported and may lack the resources to conduct a proper investigation into adverse events. Conversely, if the criteria are set too high then we may lose valuable opportunities to learn before a fatality occurs.

The task of identifying an adverse event is more complicated in military than in civil incident reporting systems. Many operations carry an intrinsically high level of risk. In many cases, it is impossible to entirely eliminate the potential hazard from occurring again without also sacrificing military objectives. For instance, the risk of drowning in river crossing exercises is well understood but most armies cannot avoid these operations. Steps can be taken to reduce the risks involved in such operations but they are unlikely to entirely eliminate the risks [4]. The focus of military reporting is, therefore, often directed towards the particular decision making processes that led to a risk being accepted rather than to the individual actions that immediately led to an adverse outcome. For example, if soldiers choose not to accept a risk then they must often seek alternative means of achieving their objectives. This creates complex situations in which it might be preferable to accept a short-term safety hazard than adopt an alternative strategy that might involve prolonged exposure to a series of lesser hazards. It might be 'safer' to cross a river and take the risk of drowning than lead a team on a longer route through mountainous terrain. Sadly, however, individual blame is often assigned if the hazard is realized irrespective of the decision that was made. Hindsight bias is a familiar aspect of military incident reporting systems.

3.2.3 Dealing with the Unexpected

Previous chapters have described the perceived need to simulate operational risks during military training exercises. Troops are deliberately exposed to hazards so that they can acquire necessary operational skills. Military training exercises are carefully designed so that any exposure occurs under controlled circumstances. Initial exercises using simulated munitions are mixed with 'live fire' exercises. These simulated operations are choreographed; the position of every participant and every system is, in theory, determined down to the last second. Unfortunately, training mishaps still occur even under carefully controlled conditions. This can be illustrated by an explosives incident that took place during a US Army nighttime training exercise. The intention was that two platoons would lead engineers across the line of departure. They would then be followed by a third maneuver platoon. The two lead platoons were to occupy 'support-by-fire positions'. The engineers and the third maneuver platoon were then to occupy hide positions before attempting to breach a triple-strand, concertina wire barricade. Such nighttime maneuvers require considerable preparation and the exercise was rehearsed several times. A daylight walkthrough was conducted without weapons, munitions or explosives. This was followed by a ‘dry fire' exercise with weapons but without munitions or explosives. The detailed breaching plan involved a team leader and two team members. The supporting members were to deploy 1.5-meter sections of M1A2 Bangalore torpedo into the concertina obstacle. The team leader would then pass elements of the initiation system to the team members who were to tie in the torpedoes to the detonating cords. The initiation system ‘consisted of a ring main (detonating cord about 1 meter formed into a loop) with two M14 firing systems (approximately 1 meter of time fuse with blasting cap affixed to one end) taped to the ring main' [17]. At the opposite end of the M14 firing systems was an M81 fuse igniter that had been attached before the start of the operation. The team leader was to give each team member one of the M81 fuse igniters. On his command, they were then to pull their M81 and initiate the charge. The breaching team was then to retreat to their original hiding place. The detonation was to act as a signal for a marking team to use chemical lights to help the following platoons locate the breach.

The incident began when the breaching team approached the concertina objective. The two-team members successfully placed their Bangalore torpedoes on either side of a potential breach site. The leader then handed the initiation system to them so that they could tie-in the Bangalore detonating cord lines. The team leader then handed one of the two M81 igniters to the team member on the left side of the breach. The team leader departed from the original plan when he placed the second M81 on the ground between the two-team members. Instead, he handed a bag containing approximately eight meters of detonating cord and an extra M14 initiation system to the team member on the right-hand side of the intended breach. The team leader then radioed the platoon leader to inform them of his intention to fire the charges. The left-side team member picked up the M81 fuse igniter that had been left on the ground. He also had the original M81 that had been given to him by the team leader. The right-hand team member held the two M81s from the bag. The team members pulled the M81 fuse igniters on the leader's order ‘three, two, one, PULL’. A Battalion S3 (operations, planning, and training officer) observed the burning fuses and the added charge in the bag, which had been placed to the right of the Bangalore torpedoes. He asked about the additional charge but did not receive any reply. The demolition team and the S3 then moved back approximately twenty-five meters to separate hiding locations. As intended, the detonation acted as a signal for the marking team and a security team to rush towards the intended site of the breach. A second, larger, detonation occurred some 3-5 seconds after the first. Both of the approaching teams were caught by the resulting blast. The initial detonation had been caused by the additional charge in the bag that had been handed to the team member on the left of the breach. The second explosion was caused by the Bangalore torpedoes [4].

This incident illustrates the challenges that are created by nighttime military exercises involving multiple teams coordinating the use of sophisticated and potentially hazardous munitions. The complexity of such exercises makes it difficult to both predict and control all of the potential hazards that arise. The additional charge moved the training exercise beyond the carefully choreographed scenarios and risk assessments that had been identfied before the event. Subsequent investigations argued that the individuals involved should not have had access to the extra detonating cord and M14 initiation system. The excess munitions should have been relinquished before this phase of the exercise. However, it can also be argued that the ability to deal safely with such unexpected conditions is an intrinsic part of military training. **It can, therefore, be argued that military incident reporting systems point to the failure of military risk assessment techniques because most adverse events stem from unpredictable circumstances that would have been difficult if not impossible to anticipate before the incident occurred.**

3.2.4 The Politics of Acceptable Risk

Military organizations must often focus finite investigation resources on those adverse events that the public and politicians perceive to be of greatest importance. Incidents involving recruits provide one example. Heat related injuries during acclimatization training are another. For instance, a series of incidents involving the Singaporean military led to the development of detailed heat exposure regulations [18]. For the first 2 days of exposure, personnel should only perform light duties. For the next 2-3 days, the intensity of exercise can gradually be 1-2 weeks if exercise is limited to 2-3 hours in the heat. If the expose is less than 90 minutes then a carbohydrate-electrolyte beverage should be offered with no more than 8%, or 2 tablespoons of sugar per liter. If the exposure is greater than 240 minutes then this should be supplement with 1teaspoon of salt per liter. As a result, the frequency of heat-related injuries in the Singaporean army has declined since 1987. However, these recommendations have not eliminated the problem. In 2000, the Singaporean army's reporting systems found that there were still approximately 3.5 cases of heat related injury per 1,000 soldiers in service. In spite of the changes that have been made, the reporting systems continue to reveal an uneven distribution of cases. The majority occurs in training schools that prepare National Service recruits for military life. Cases are still clustered around the periods of physical activity that are allowed in the morning and the afternoon when the heat is less severe.

The Singaporean army faces particular problems in address heat-related injury because of the climate in which it operates and the diverse pool of recruits that it receives from the National Service intake. Similar adverse events are mentioned in every military reporting system. For example, a US General Accounting Office report highlighted a case in which a Marine's death in training was classified as being due to 'natural causes' even though he had just completed 5 pull-ups, 80 sit-ups, and a 3-mile run [19]. This creates a variation on Perrow's Normal Accidents [20]. He coined the term by arguing our desire to introduce increasingly complex and tightly coupled technologies will inevitably lead to future accidents. In contrast, the stubborn nature of heat-related military injuries points to a less sophisticated form of 'normal accidents'. These do not stem from the economic pressures for technological innovation. They stem from the organizational pressures that prevent individuals from escaping the lessons of the past. Operational and training procedures continue to depend upon a physiological mismatch between the environments in which a recruit must operate and their own physical resources. If we remove this mismatch, for example, by introducing more stringent exposure controls then incidents can be reduced. This hardly requires the support of complex risk assessment methodologies. The point of the exercise may be lost, however, if individuals are left unprepared for their operational environment when training finishes. This again illustrates the tension between the need to avoid adverse events and yet provide individuals with experience in controlling or avoiding the hazards that arise in operational contexts.

3.2.5 Incident Reports and the Politics of Military Risk Assessment

As we have seen, military forces have allocated considerable resources to understand and combat the causes of heat related injuries. The importance of learning from these incidents has increased in recent years partly as a result of political and public concern, especially over incidents involving new recruits or conscript and national service units. It is difficult to underestimate the impact of such influences on military operations. In democratic countries, armies must account for any adverse events to the politicians and public who employ their services. It, therefore, follows that operational concerns or training priorities cannot solely be used to determine what is an acceptable military risk. This is at least in part a political decision. The US and UK forces have recently felt these influences following apparent clusters of suicides within particular units.

Given the diversity of military operations, however, it is impossible for politicians and the public to provide detailed guidance about every possible hazard that soldiers might be exposed to. There is an assumption that operations will be conducted in a 'safe' manner. In consequence, single incidents will often act as a trigger for political and public interest in hazards that have not previous attracted widespread attention. This can be illustrated by the findings of a Board of Enquiry into the drowning of an Australian cadet. The Chief of Staff of Headquarters Training Command found that: "I accept the Board of Inquiry finding that Cadet S drowned as a result of the amount of weed in the water, the depth of water, the wearing of GP boots and Disruptive Pattern Camouflage Uniform clothing whilst in the water and the absence of safety devices and inadequate safety precautions for the swimming activity. These factors contributed to Cadet S's drowning. A swimming activity undertaken by cadets as young as 13 years with unknown fitness levels and unknown medical conditions in the circumstances existing on 18 Nov 00 at the Bjelke Peterson Dam, was inherently dangerous…I do not accept the finding of the Board of Inquiry that Corporal X was not fully qualified as an instructor of cadets in the Army Cadet Corps in accordance with the Army Cadet Corps Policy Manual. Corporal X had completed the Instructor of Cadets Course and First Aid Course in compliance with the Army Cadet Corps Policy Manual and was qualified as an Instructor of Cadets" [21]. As we have seen, however, incidents involving new recruits and cadets often trigger greater concern than those involving more experienced soldiers. The political and public reaction to this incident went well beyond the immediate events surrounding the drowning. It motivated a more sustained review of Cadet activities that questioned the Chief of Staff's assessment of the competency of the individuals concerned. For instance, the Australian Minister for Veterans Affairs contradicted parts of the previous statement when he argued that “the swimming activity was not authorized by (the) Army and that there was inadequate supervision or monitoring of the Army Cadet Corps activity" [22]. In consequence, he suspended all cadet-swimming activities conducted in areas other than supervised swimming pools until there was a more systematic analysis of the risks involved in their training.

Similar tensions can be seen in the findings issued by UK Coroner’s Courts into recent fatalities in both Iraq and Afghanistan. These courts conduct inquests, or legal inquiries, into the causes and circumstances of a death. In 2007 the Oxfordshire Assistant Deputy Coroner found that a UK soldier was ‘unlawfully killed’ when his Scimitar tank was hit by "friendly fire" from a US aircraft in 2003. The death was "an entirely avoidable tragedy". The pilots' attack on the British convoy near Basra "amounted to an assault" and was bordering on criminal. Such findings have a public impact that goes well beyond the usual effect of military risk assessments or incident reports. In this case, it increased the political prominence of ‘friendly fire’ incidents beyond any consideration of frequency or costs. It helped to focus political attention on the concerns of the families and friends of the victim [23].

3.2.6 Problems of Scale

It is important to emphasize the considerable differences that exist between military incident reporting systems. Many of these differences stem from variations in the types and scale of operations that are conducted by the armies of different nations. Some perform relatively limited, ceremonial duties within their own borders. Others are simultaneously engaged in offensive military operations, policing and peacekeeping missions throughout the globe. The diversity of such operations and the geographical distribution of people and material makes for pathological problems. Any recommendations that are made following an investigation have to be communicated across huge distances to potentially very remote locations if they are to have any impact on subsequent risk assessments. They must also be acted on before any recurrence could happen. It is for this reason that many of the larger military organizations impose tight timing constraints on their investigation processes. The US Army gives the responsible Department of the Army-level organization 60 calendar days to provide an initial response to the US Army Safety Center describing any corrective actions. Interim and follow-up reports are required every 90 days after this initial response until the actions are closed. If the responsible command does not accept the recommendations then a report must be filed with the Commander of the US Army Safety Center, with a supporting rationale within 60 days of the initial notification.

Such procedures are necessary because of the extreme complexity of military systems and the organizational structures that support them. It is extremely easy for safety lessons to be lost amongst a mass of other operational updates. For example, the US Army issued at least eight revision requests for the M9 Armored Combat Earthmover manuals in a single month in 2000: TM5-2350-262-10, TM5-2350-262-10HR, LO5-2350-262-12, TM5-2350-262-20-1 & 2, TM5-2350-262-20-3, TM5-2350-262-34, TM5-2350-262-24P, TM5-2815-240-34 & P [24]. In addition to these sources, Armored Combat Earthmover operators also had to monitor two additional web sites (http://ncc.navfac.navy.mil and http://www.tacom.army.mil/dsa/) that contained further information about modifications and revised operating procedures for their vehicles. In consequence, US Army engineers did not always receive necessary technical information.

The US General Accounting Office provides further examples [25]. Division personnel did not receive revisions to the manual describing the fuel subsystem on Apache attack helicopters. The aircraft were then grounded and the maintenance teams wasted many hours troubleshooting because the old manual did not provide necessary information about how to fit new fuel transfer valves. The lack of an adequate monitoring system created further logistical problems. It was difficult for engineers to coordinate the implementation of multiple modifications to individual pieces of equipment. In consequence, the same item might be repeatedly removed from service while multiple modification orders were completed. Some items of equipment did not always work together after modifications. This loss of integration further delayed other maintenance procedures and reduced operational capability. For instance, modified parts were removed from Huey utility helicopters. Non-modified parts were then reinstalled because there were no modified parts in stock when the new parts broke.

The US Army's Modification Work Order (MWO) program was intended to address many of the problems described above. This intention was to ensure that 'any identified operational and safety problems' were consistently implemented across the US Army [25]. A centralized database was developed to record the progress of different maintenance recommendations. Army headquarters officials and Army Materiel Command officers could issue queries to check whether individual units met the timescales and objectives that were recommended in safety notices. Unfortunately, the database was discontinued following a structural reorganization in 1990. Control over modification installation funding was transferred from the headquarters level to the individual program sponsors who are responsible for major weapon systems, such as the Abrams tank, or for product centers that support particular pieces of equipment, such as the Squad Automatic Weapon. The result of this decentralization was that 'Army headquarters and Army Materiel Command officials do not have an adequate overview of the status of equipment modifications across the force, funding requirements, logistical support requirements, and information needed for deployment decisions' [26].

3.3 'Normal Accidents' in Military Operations

There is a clear relationship between military risk assessment and the feedback that can be obtained from mishap reporting systems. We must learn from incidents in order to reduce the likelihood and mitigate the consequences of future accidents. Most previous work in this area has focused on civilian Safety Management Systems. In contrast, this Chapter has examined the ways in which military reporting systems can help us to understand the causes of technical failure, of managerial problems and of human 'error'. There are some similarities between military and civil systems. For instance, both encourage investigators to look beyond the immediate or catalytic causes of human failure. Similarly, they are both faced with a tension between attributing blame to individual violations and the need to identify the systemic causes of accidents and incidents.

There are also important differences between military and civil ‘lessons learned’ systems. Military incident reporting is complicated by the need to determine whether or not individuals were justified in reaching particular decisions about the complex risks that they must face. Soldiers are often forced to make complex, real-time decisions with limited information. Their difficulty of their task is often exacerbated by the realization that military personnel are punished when their gambles fail. A form of hindsight bias often ‘informs’ punitive actions by military tribunerals. .

Many military reporting systems contain incidents that reflect the wider tension between ensuring that training is ‘safe’ whilst still providing personnel with the necessary skills that help to ensure operational efficiency. Soldiers are often exposed to hazards in order to prepare them for the pressure of operational tasks. As we have seen, considerable efforts are made to sure that these hazards are controlled during training operations. Dry fire rehearsals are conducted. The movements and actions of each participant and their supporting systems are choreographed. Even so, mishaps continue to occur.

Military accidents are ‘normal’. Not because they are acceptable but because they continue to occur in every army throughout the globe. They stem from the complexity and interconnection of military operations. Unlike Perrow’s ‘Normal Accident’ theory [20], where the focus is on the impact of technological innovation, military accidents often have more prosaic causes. They stem from ‘wicked’ combinations of contributory factors that cannot easily be predicted by existing risk assessment techniques. Therefore, rather than supporting risk assessment techniques, the mishaps that are documented in incident and accident reporting systems point to the failure of existing hazard assessment methodologies. For instance, it would have been difficult to alter the plan behind the Bangalore torpedo incident to explicitly include a plan of what to do should additional munitions be available during the exercise. Only in hindsight can we identify this as a potential hazard for the personnel involved in the operation. Given that the material was available, it is similarly difficult to argue that planning staff should have considered the possibility of a preliminary explosion triggering the arrival of the breeching units before the torpedo exploded. The difficulty of predicting all of the possible ways in which mishaps might occur is further complicated by the ‘risky’ nature of many military operations. The Canadian Lessons Learned in NATO peacekeeping operations illustrated this point. Injuries from mines led soldiers to explicitly request training in the use of IV lines and morphine. This creates potential risk if these techniques are used inappropriately. The Catch-22 of military safety management is that there is also a risk that lives can be risked if soldiers do not have these skills.

3.4 References for Chapter Three

[1] UK Health and Safety Executive, The Southall Rail Accident Inquiry Report: {HSC} Action Plan to implement recommendations, Health and Safety Commission, Her Majesty's Stationery Office, London, United Kingdom, 2000.

[2] EUROCONTROL, ESARR 3, Safety Management Systems in Air Traffic Management, Brussels, Belgium, 2000.

[3] International Maritime Organisation, International Safety Management (ISM) Code, London, United Kingdom, 2010.

[4] C.W. Johnson, A Handbook of Accident and Incident Reporting, Glasgow University press, Glasgow, Scotland, 2003.

[5] US Army, Accident Investigators Handbook, (Headquarters Department of the Army: Washington, DC, USA), November, 2007.

[6] US Department of the Army, FM 5-19: Composite Risk Management, Headquarters, Washington, DC, August 2006.

[7] US Army Safety Center, Safety Notice: Water Safety Trend (US Army: Fort Rucker, Alabama, USA), 2002.

[8] US Army Safety Center, Safety Notice: Confidence Course Obstacle [Note SAN 210800MAR00] (US Army: Fort Rucker, Alabama, USA) 2001.

[9] US Army, Product Manager, Paladin/FAASV (2000), Drivers' Hatch, *Paladin/FAASV Newsletters*, (Picatinny Arsenal: NJ, USA, 2nd Quarter, Financial Year 2000, 8.

[10] Canadian Department of National Defense, Letters to the Editor, *Safety Digest, (*Vice-Chief of the Defense Staff: Ottawa, Canada), Safety Digest, 7-99, 1999.

[11] Canadian Department of National Defense, Letters to the Editor from L. Cdr. D. Alder, *Safety Digest, (*Vice-Chief of the Defense Staff: Ottawa, Canada), Safety Digest, 11-00, 2000.

[12] US Army Safety Center, A Turn for the Worse, M1A1, *Countermeasure, 22, 1, 14,* (US Army: Fort Rucker, Alabama, USA), 2001.

[13] J. Devine and A. Smith, Safety Culture Enhancement Activities and Next Steps within the Federal Aviation Administration, In Proc. of International Systems Safety Society, Baltimore, MD, ISSC, Unionville VA, USA, ISBN 0-9721385-7-9, 2007.

[14] US General Accounting Office, M.E. Gebicke, Army Ranger Training: Safety Improvements Need to Be Institutionalized, [GAO/NSIAD-97-29] (United States' General Accounting Office: Washington, DC, USA), 1997.

[15] Canadian Army Lessons Learned Center, Common Observations and Issues: Operation Palladium Rotations Zero to Four [Analysis 9901] (Canadian Army Lessons Learned Center, Vice-Chief of the Defense Staff: Ottawa, Canada), 1999.

[16] US Army Safety Center, Accident Investigation: The Other Side of Risk Management, *Countermeasure, 22, 1, 1-16,* (US Army: Fort Rucker, Alabama, USA), 2001.

[17] US Army Technical Center for Explosives Safety, Last minute change is fatal: how could this happen? Explosives Safety Bulletin, 12, 1, 6-8, 2000.

[18] Singapore Army Safety Organization, Preventing Heat Injuries: The Commanders Guide (Safety Organization, General Staff Inspectorate: Singapore), 2001.

[19] US General Accounting Office, M.E. Gebicke, Military Training Deaths: Need to Ensure That Safety Lessons Are Learned and Implemented, [GAO/NSIAD-94-82] (United States' General Accounting Office: Washington, DC, USA), 1994.

[20] C. Perrow, Normal Accidents: Living with High-Risk Technologies, (Princeton University Press: Princeton, NJ, United States of America), 1999.

[21] Australian Army, P. B. Retter, Brigadier Chief of Staff, Board of Inquiry into the Death of Cadet K. P. Sperling of the South Burnett Regional Cadet Unit arising from an Incident at the Bjelke-Petersen Dam (QLD) on 18 Nov 2000: Findings, Action and Implementation Plan, (Headquarters Training Command: Canberra, Australia), 2001.

[22] B. Scott and B. Nelson, Release of Board of Inquiry Findings and Action Into the Death of an Army Cadet [Min 096/01] (Department of Defense: Canberra, Australia), 2001.

[23] A. Masys, Fratricide in Air Operations: Opening the Black Box and Revealing the Social, PhD thesis, University of Leicester, Dept of Criminology, UK, 2010.

[24] US Army Safety Center, New Requirements for the M9 ACE, Countermeasure, 22, 6, 15, (US Army: Fort Rucker, Alabama, USA), 2001.

[25] US General Accounting Office, Army Equipment: Management of Weapon System and Equipment Modification Program Needs Improvement, [GAO/NSIAD-98-14] (United States' General Accounting Office: Washington, DC, USA), 1997.

[26] US Army Safety Center (2001d) Safety Notice: HMMWV Seatbelt (US Army: Fort Rucker, Alabama, USA).

4 Degraded Modes and the Military ‘Culture of Coping’

Previous chapters have used case studies drawn from a number of military organizations to illustrate the complexity of risk assessment for complex, dynamic operations. Many previous accidents have been caused by failures in the ad hoc mitigations that were introduced to guard against existing hazards. In contrast, the following sections look more narrowly at a single case study. The intention is to provides a more detailed account of the difficulties that face military personnel in conducting and maintained assessments of the risks posed by legacy technologies.

Many safety-critical subsystems have a mean time to failure that is less than the intended operational life of the applications that they support. In such circumstances, designers and operators must work together to ensure that maintenance intervals are scheduled so that critical components are repaired or replaced before they fail. Risk assessment techniques can be used to ensure that those components with the greatest likelihood or with the most critical consequences of failure receive greatest attention during preventative maintenance. However, the shorter the maintenance intervals then the higher the costs will be. There is an incentive to delay intervention as late as possible without jeopardizing safety. Once a sub-system has failed, the same financial pressures can persuade managers to find ‘work arounds’ or ad hoc patches that enable operations to continue. In other words, operators learn to maintain system functionality under ‘degraded modes of operation’ without increasing the level of risk beyond acceptable limits [1, 2].

In military systems, these pressures are considerably more complex. For example, operational constraints can prevent personnel from replacing failed components that must be delivered along extended supply chains [3]. This is a particular problem in naval operations in remote regions where it may not be possible to source replacement parts for weeks or months at a time. The pressure to maintain functionality in the face of sub-system failures can also be exacerbated by operational requirements in the field. For example, the risks of working without any surface to air missile systems will often outweigh the risks associated with continuing to use such an application even when there are known bugs or failure modes [4]. Partly in consequence, many military organizations rely on training and doctrine to help personnel find ways of working around design flaws that were not adequately addressed during the procurement of complex systems [5].

4.1 The Accident On-board HMS Tireless

This Chapter focuses on a fatal incident involving Her Majesty’s Ship (HMS) Tireless while the submarine was on deployment [6]. Two members of the crew were killed and others were injured. The number of injured and the nature of their injuries was redacted from the published findings of the Board of Enquiry for reasons of confidentiality and security. The incident occurred while Tireless was taking part in an under-ice training and tactical evaluation exercise close to the US Applied Physics Laboratory Ice Station in the North of Alaska.

At the time of the accident, the submarine was following Standard Operating Procedures (SOPs) using Self Contained Oxygen Generators (SCOGs) to maintain the oxygen level in the vessel. SCOGs contain a mix of sodium chlorate and iron powder or of potassium and lithium chlorate. When ignited, the mixture smolders at about 600 °C to produce sodium chloride, iron oxide and oxygen. These devices introduce a number of hazards. For instance, the risks associated with high operating temperatures must be mitigated by insulating the SCOGs both to maintain the reaction and to protect surrounding equipment.

HMS Tireless was using these devices because the primary oxygen supply relied on low pressure electrolysers that were liable to trip if ice formed in the hydrogen discharge piping. The two crewmembers that were killed in the incident had been qualified to operate these SCOGs having conducted maintenance procedures on many previous occasions. They were responsible for training other members of the crew on their operation. At 19:56, approximately an hour before the watch was due to change, a loud bang was heard throughout the submarine [6]. The forward end of the vessel filled with smoke. This triggered a fire alarm. The flood alarm was also raised because the explosion was sufficient to depress the manual flood alarm button in the Forward Escape Compartment (FEC). From this point on it took 44 minutes for personnel to regain entry to the FEC where the two victims had been working. The hatch doors leading from the forward bunk space to the FEC were blown shut causing them to buckle and jam in place. Several small fires were started, probably from pools of sodium chlorate and iron fillings that were scattered when the SCOG exploded. A far more serious incident could have occurred if other members of the crew had not taken prompt action to extinguish these fires. The following sections analyze the role that risk assessment played in anticipating the causes of this accident. This analysis is then used to identify the ways in which a series of human errors, organizational failures and equipment problems combined to undermine the initial hazard analysis.

A number of theories were put forward to describe the causes of the accident on HMS Tireless. Each of these hypotheses focused on the hazards of operating Self Contained Oxygen Generators. One suggestion was that the explosive cartridge, which is used to initiate the SCOG reaction, could itself have ruptured the canister. However, testing showed that this hazard was unlikely to have had the consequences seen during the accident. The blast generated by the cartridge was insufficient to cause such a failure.

The risk of water contamination was also discounted. Water ingress might have caused a potential accident through over-pressurization from the formation of steam as the SCOG reaction continued. However, the subsequent Board of Inquiry concluded that this could not result in the explosive fracture that was seen in the Tireless accident [6]. Blocked vents might also have led to the rupture of the canister from a buildup of internal pressure. However, this also would not have had the force witnessed in the Forward Escape Compartment. Physical damage to the sodium chlorate block and manufacturing defects, including an increased concentration of iron filings, would not have increased the stored energy in the SCOGs to the point where the damage would be consistent with that witnessed on Tireless.

It was concluded that the only 'plausible' cause was contamination of the sodium chlorate block by an organic liquid. In particular, oil contamination might have occurred while it was stored in a submarine. Unfortunately, it was not possible to be more precise about the source of such a problem; "due to the patterns of logistic management of… SCOGs, the SCOG that exploded may have been embarked and disembarked from many different submarines before its use in Tireless" [6]. As we shall see, the nature of many military operations means that the (mis)handling of equipment during the supply chain often introduces hazards that are overlooked during the preliminary risk assessments that guide their acquisition. These hazards, typically, arise many months or years after an initial risk analysis was conducted. They stem from the everyday operational pressures that characterize military life and hence they typify the 'normal' accidents that were identified in the closing sections of Chapter Three.

The Board of Inquiry identified a number of contributory factors. These included potential cracks in the sodium chlorate block and 'constraints' imposed by the location of the SCOG holder. These physical issues were again the result of deficiencies in the 'acquisition, manufacture, transport, storage, stowage and logistics management of the SCOGS' [6]. Hence this incident provides many lessons that can be learned about the contexts in which risk assessments fail to identify the degraded modes of operation that gradually erode the safety of military operations.

4.2 Procurement, Degraded Modes and Risk Assessment

In the late 1980s the UK Ministry of Defense formed the Submarine Secondary Improvement programme. One aim was to increase the 3.5 days of emergency oxygen that could be provided by existing oxygen candles. It was recognized that some rescue scenarios might require up to 7 days of oxygen. In consequence, a revised design was submitted to the MOD for a Self Contained Oxygen Canister producing approximately 30% more oxygen than the candles. Evaluation trials were commissioned to provide empirical evidence about the risks associated with the new design. This led to two incidents in which one prototype SCOG caught fire and another ruptured from a build-up of internal pressure.

In the mid-1990s, a revised design was submitted. This successfully passed the associated test programme. These SCOGs were designed under the requirements of Joint Service Publication 430, a standard describing Ship Safety Management for the UK Ministry of Defense. This document was revised during the development of the devices, however the principle requirements were unchanged. JSP 430 recognized the MOD's obligation to "to manage the often greater safety risks associated with military operations whilst ensuring that arrangements are at least as effective as statutory requirements". In other words, there was a requirement "To demonstrate that so far as is reasonably practicable, the safety and environmental management of MOD shipping activities to be at least as safe and effective as that required of UK commercial shipping activities" [7]. This raises a host of difficulties. Previous sections have described the difficulties in deriving objective measures of risk that might be used as a basis of comparison between civil and military operations. The certification was intended to "provide assurance by subject matter experts to both Duty Holders and the Ship Safety Board that, in certain key hazard areas, suitable technical standards have been selected and implemented that reduce risk to ALARP and either broadly acceptable or tolerable and represents adequate risk mitigation for the key hazard areas". Again, Chapter One has identified the psychological and social biases that influence subjective, expert assessments.

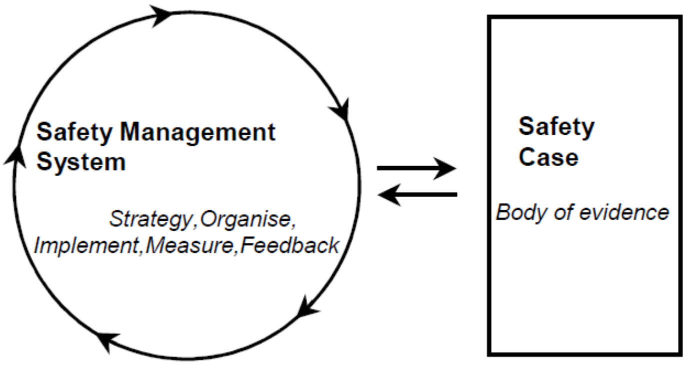


Fig. 10. Relationship between Safety Management Systems and a Safety Case

(UK Ministry Of Defence [7])

JSP 430 recommends the use of Safety Management Systems (SMS), introduced in Chapter Three. These provide the organizational structures that are necessary to meet the objectives in a safety policy. Figure 10 shows how the processes within an SMS are intended to support the generation and maintenance of Safety Cases. These documents provide a structured argument;

> "…that a ship or equipment is safe for a given application in a given operating environment. The scope should be proportional to the complexity of the system in question but will include for example; reports, risk management documentation and certification. The risk management process uses the principles of As Low As Reasonably Practicable (ALARP) and either broadly acceptable or tolerable, to verify that the level of whole ship risk is acceptable" [7].

A number of hazard review workshops were, therefore, held to develop a safety case for the deployment of SCOGs in Royal Navy vessels. The subsequent Board of Inquiry into the accident in HMS Tireless describes how the safety case assessed the contamination of a SCOG as "non credible due to design ('non credible' is described as 'extremely unlikely to occur during the operational life of the unit')" [6]. The risk assessment did not consider the possibility of contamination during transport, handling and storage.

After the Safety Case had been developed to meet JSP430, additional documentation was produced that did warn personnel about the potential hazards of contamination. These were not formally part of any risk assessment but instead included the operating document BR1326 ‘Submarine Air Purification Manual and in Planned Maintenance Documentation’.

Similarly, the safety case did not consider the potential consequences should a SCOG explode during operation. Consultations took place during procurement about whether the materials in these devices could constitute an ‘explosive store’ on board a submarine. However, the Defence Ordnance Safety Group rejected this by arguing that the Mk V oxygen candle, which it replaced, was not classified as an explosive. Self Contained Oxygen Generators were introduced on UK submarines in 2003. Similar devices have been supplied to the Australian, French and US naval forces.

4.3 Tolerance for Risk in Degraded Modes of Operation

The Tireless accident reiterates some of the problems identified in Chapter Three that limit the use of ‘lessons learned’ systems to inform subsequent risk assessments. There had been a long and well documented history of incidents involving devices similar to those in this accident. These included damage to the Mir Space Station that was caused by contamination of an oxygen generator; part of a latex glove was left inside the device during manufacture. There were also comparable incidents involving NATO vessels. In the 1980s, the US Navy suffered two ‘oxygen candle furnace fires’ accompanied by explosions that were ascribed to hydrocarbon contamination of the devices. The Naval Board of Enquiry found evidence of a number of previous mishaps involving the Self Contained Oxygen Generators in Tireless. These even included incidents that occurred during the deployment in which the accident occurred [6]. In one example, a continuous four inch flame burned from one of the outlet ports shortly before the device was successfully ignited. The crew tipped the SCOG into a bucket of water until it was extinguished. Seven of the Tireless’ canisters misfired and were subsequently reignited. Another device continued to ‘rattle’ in its holder after being ignited.

Defence Standard 00-44 (Part 1) described the collection and classification of reliability and maintainability data across the UK armed forces [8]. This standard details how Navy vessels should use forms S2022 and the submarine counterpart S2022(S) to report shortcomings “in any equipment, documentation, materiel or procedures”. The decision about whether or not to submit a report is left largely up to individuals. There are, however, a small number of mandatory events that must be reported. These change over time and, typically, involve equipment that is being accepted into service. The completed forms are then validated and transcribed for use within the Submarine Upkeep Data System (SUDS) software system. The subsequent examination of the SUDS data found several incidents in which SCOGs had failed to ignite. Recommendations focused on reducing the shelf life of the cartridges that were used to trigger the chemical reaction.

The S2022 reporting systems also provided information about more serious incidents. These included fires in HMS Superb (January 2006) and two incidents in HMS Trafalgar (October 2004). The first fire in Trafalgar was especially serious because the canister became so distorted that it could not be removed from its holder. This hampered attempts to extinguish the flames. In the second incident during October 2004, molten materials began to drip through holes on another of the SCOG canisters in Trafalgar. These incidents are significant because they show that crewmembers continue to entrust their lives to devices which have already failed. It also illustrates how multiple incidents can stem from problems that appear to be common across a batch of canisters. The manufacturer identified flaws in their production processes and recalled two hundred and ninety four SCOGs from four different production runs. All of these canisters were found on board the Trafalgar. eleven had already been used. The submarine was ordered to land all of these suspect SCOGS on return to the UK so that they could be withdrawn from service.

Following the incident in the Tireless, attempts were made to ensure that the SCOGs from these faulty batches had been destroyed. One hundred and three were found in storage at Devonport naval base, forty nine were still in the Trafalgar, one was in HMS Vanguard, two remained in HMS Tubulent, one was found in a dangerous waste store and forty eight remained unaccounted for. The investigation also heard that around one thousand canisters previously had been sent back to the Hazardous Waste Store as being unfit for purpose. However, in November 2006 the majority had then been returned to the supply chain. This had taken some six months before the incident in Tireless. The decision to mark these SCOGs as serviceable was made following a visual inspection. These canisters were supposed to have a shelf life of 10 years on board a submarine and most were no more than five years old; “these events serve to demonstrate that the logistics management of SCOGs has been poor. It is possible, but cannot be proved, that a number of the unserviceable SCOGs recalled following the Trafalgar incidents were amongst the 550 … received by Tireless on 5th February 2007 prior to sailing for the ICEX” [6].

The S2022 reports provide valuable insights into the pragmatics of safety management that lie behind the more theoretical models illustrated in Chapter Three. Previous sections have argued that incident reporting systems are intended to support risk assessments by providing an important source of operational feedback on the frequency and consequences of previous failures. However, the inquiry after the accident on HMS Tireless shows how difficult it can be to exploit operational information that is gathered in the months and years after initial deployment. It is, typically, when a catastrophic failure that attention is focused on common patterns between the adverse events that are documented in military reporting systems. Such post hoc revisions to risk assessments occur too late to protect the lives of many service personnel.

The Board of Inquiry into the Tireless accident also uncovered evidence that a number of previous SCOG incidents had gone unreported. This is significant because it suggests that any risk assessment that is derived from operational incident reports is likely to provide a considerable under-estimate of the likelihood of future failures. As we have seen, there is considerable discretion over the submission of these forms. One of the critical problems in combating 'degraded modes' of operation is that operators may not understand the significance of the risks posed by equipment failures. They may be so accustomed to using 'work arounds' and other forms of coping strategy that they do not file incident reports. In consequence, higher levels of command may have little idea of the problems that occur during operational service; "precisely why submarine personnel are under-reporting is a matter of conjecture but it is the opinion of the Board that the S2022 reporting system has its shortcomings" [6]. Disillusionment and under reporting are characteristic of adverse event monitoring systems when staff receive little feedback or suffer long delays before their concerns are addressed [9, 10].

4.4 Logistics and Management of Degraded Modes of Operation

Submarines typically carry SCOGs for two different purposes. They are either stored for escape or for 'ready use'. Escape devices are stored in purpose built sealed lockers in the vessels escape compartments. They were protected to reduce the probability of accidental damage. The escape SCOGs are not normally taken out except for an independent inspection conducted each year as part of the submarine's escape audit. These periodic inspections are designed to reduce the risks that oxygen generators will fail under emergency conditions. After the accident, an inspection of the escape SCOGs in HMS Tireless showed that they had not been damaged.

Attention began to focus on the management of the 'ready use' SCOGs. These were kept in the engineer's store and a forward naval store. Both locations had been fitted with appropriate fixed spray fire suppression systems. The 'ready use' SCOGs were intended for operational situations such as under ice exercises. They were not stored in purpose built containers. The cardboard packaging was often removed before they were stowed. Following the incident, it was determined that the Tireless carried more than seven hundred of these 'ready use' devices. A sample of two hundred and fifty eight were inspected, of which fifty nine (23%) had been burned while the remainder had not been used. One hundred and forty seven (57%) had suffered some form of physical damage, forty two (21%) were not fully sealed, seventy one (28%) showed signs of corrosion, twenty seven (11%) had suffered 'gross contamination' with either oil or grease [6]. In other words, accidental damage and contamination posed a considerable hazard to both escape and ready use oxygen generators. The protective measures that were used to mitigate the emergency devices were not provided for the 'ready use' SCOGs. The board concluded that these SCOGs posed a considerable risk to both HMS Tireless and the other submarines that carried them.

One of the reasons why crews did not appreciate the risks posed by 'ready use' SCOGs was that the safety warnings had worn off. Screen printing techniques had been used to mark the SCOGs. However, 'their dangerous goods classification appears to have little or no impact on how they are transported to and from submarines' [6]. The subsequent investigation found that the canisters for the Tireless had remained uncovered on a jetty from the 5th until 19th February before being embarked into the submarine. While those for the HMS Vanguard had remained uncovered at Devonport for just under two years before being investigated by Explosives and Health & Safety Officers. Many of the warnings had been rubbed off through contact with other surfaces or with contaminants including water. These problems were identified by the UK Marine Forces Marine Environment Survivability and Habitability group. Steps were taken to mitigate the risks by relabeling remaining canisters. However, significant numbers of unlabelled SCOGs remained in Navy vessels.

An important technique for reducing or mitigating the hazards created by degraded modes of operation is to provide staff with explicit opportunities to raise concerns about problems with the systems that they must operate. However, the only opportunity to reject SCOGs was when they were receipted after delivery to the submarine. No evidence could be found that this happened across the Navy. Personnel had not been trained to identify whether a canister was serviceable. The guidance that was available was ambiguous because it did not explicitly state the criteria for acceptance/rejection of the devices; "over the course of the investigation…, the Board has formed the view that complacency had set in since the introduction of the SCOG in 2003 and personnel were less cautious than they had been with the previous Mk V. oxygen candle. Despite the presence of warnings about the explosive risk presented by contamination with organic material there was no real experience or understanding within the MOD of just how violently a contaminated sodium chlorate candle could react" [6].

4.5 Trust and the Mitigation of Risk

Previous sections have described how Self Contained Oxygen Generators were only used as a secondary from of oxygen generation. The primary system was based around an electrolysis cell stack to separate water into hydrogen and oxygen. These systems carry their own risks. For example, it was understood before the mission that this primary system would not work in the waters off the North of Alaska at depths any shallower than 150 meters. The electrolyser hydrogen discharge would freeze causing the system to trip. This concern was documented in a Fleet Publication Notice (FPN 27) and partly explains why Tireless had loaded additional SCOGs before leaving for under-ice operations. Many in the Navy management structure understood the risk that the electrolysis stack would fail for long periods of the mission. Subsequent investigations by the Ministry of Defence therefore, questioned why modifications were not made to extend the operating range of the primary systems. This would have reduced the hazards created by carrying so many 'ready use' SCOGs in ad hoc storage areas [6].

The Board of Inquiry argued that sodium chlorate technology continues to provide adequate protection for the crew providing that an assessment is conducted so that 'proper risk mitigation to prevent liquid organic contamination is applied to any future system' [6]. They concluded that these technologies remain an acceptable secondary source of oxygen production for submarine operations; given that the associated hazards are no greater than those presented by other equipment in the submarine. However, this form of reasoning seems remarkably similar to the arguments that were used in favor of retaining the same explosive classification for SCOGs and Mk V Oxygen Candles. The SCOGs were no more dangerous than the candles. As we have seen, however, they exhibited different, more explosive failure modes during this accident.

Many submariners have become suspicious of SCOGs. They are now unwilling to light them. In the aftermath of this incident, steps were taken to remove all of the existing design and replace them with a revised device. However, it will be necessary to 'undertake a full internal public relations campaign, possibly including a change of name, before a reintroduction into service is considered' [6]. These are significant observations; they reveal the consequences that arise when crewmembers understand the potential hazards from degraded modes of operation. It is regrettable that these dangers are often not fully realized until accidents have jeopardized the lives of many individuals.

4.6 Degraded Modes and Dimensions of Coping

This chapter has argued that a culture of 'making do' exposes military personnel to levels of risk that would never have been considered during preliminary hazard analyses prior to deployment. Teams are encouraged to use their initiative to find 'coping strategies' that respond in flexible ways to unanticipated systems failures. It follows that individuals must be encouraged to raise concerns when these failures create risks that jeopardize successful operations. The previous sections have focused on problems in the handling and procurement of Self Contained Oxygen Generators. However, these issues cannot be viewed in isolation. This culture of coping extends across multiple platforms and systems. For example, the investigation into the explosion and subsequent fires in Tireless found several other examples of applications where the crew worked hard to overcome design flaws [6].

4.6.1 The Risks of Emergency Management

A number of coping strategies had to be deployed by the crew in the immediate aftermath of the incident. This is a significant observation because many of the risks associated with emergency management are never considered during the Safety Cases that justify the introduction of technologies, such as the SCOGs. This provides a further example of the more general argument introduced in Chapter One. Risk assessments often help to identify techniques that mitigate primary or secondary hazards. However, very few existing techniques take a systematic approach to the assessment of knock-on risks. In this accident, the hazards of electrolysis failure were mitigated by the introduction of 'ready use' SCOGs. As we shall see, however, insufficient steps were taken to mitigate the hazards created when those oxygen generators failed.

Twenty five personnel in HMS Tireless were forced to don breathing apparatus in the Forward Bunk Space as it began to fill with smoke. Visibility was reduced to less than half a meter. They could have chosen to use self-contained Emergency Escape Breathing Devices. These were available and were specifically designed to allow escape from a smoke filled compartment. However, the formal investigation noted that these devices were relatively complicated. Crewmembers had 'a lack of confidence in it and an immediate preference to seek the Emergency Breathing System (EBS)'. This was a tethered supply that imposed more restrictions on crew movements around the nozzles where they could access the masks. The Emergency Breathing System masks were stored in lockers and these had to be emptied out before they could be distributed. Not only were there insufficient masks for all of the crew in the bunk space. There were not enough PCL (Pneumatic Components Limited) couplings for them to attach their breathing apparatus to the Emergency Breathing System. When access to the Forward Escape Compartment is closed off there is only access to 18 PCL couplings for the crew who are located in the forward bunk-space beyond the 29 Bulkhead. In consequence, some of the crew were forced to use a 'buddy system' as they shared their masks. Others decided to crawl out of the compartment beneath the smoke. It was fortunate that the submarine officers decided not to close the 29 Bulkhead otherwise there might have been greater fatalities amongst those who were forced to cope with an inadequate number of marks and couplings.

This Tireless accident revealed further problems with the Emergency Breathing System couplings. These illustrate how risks may only be identified for some systems during extreme situations as crews struggle to devise coping strategies. There were two different designs for the PCL couplings. A more modern version required both hands to plug or unplug the user's mask. However, the older design only required one hand to make the connection. These differences were particularly exposed when personnel moved rapidly from one position to another in the vessel. In many situations, individuals who were carrying critical equipment had to put it down to disconnect the hose, move to their new location, put the equipment down and then reconnect before picking up the equipment again so that it could be operated. It is difficult to underestimate the additional workload that this created in cramped conditions with low visibility, especially when crewmembers had built up experience in the 'fleeting' one-handed operation of the older coupling devices. It is clear that the existing Safety Case for the SCOGs did not consider the interaction between these devices and piecemeal modifications to the PCL system. However, these interactions typify the system-level risks that expose military personnel to the greatest risks and they respond to major emergencies.

4.6.2 Training and an Expectation of Risk

It might be argued that the problems experienced in the aftermath of the SCOG blast should have been identified during training drills. However, the difficulty in connecting and reconnecting to the Emergency Breathing System was compounded by a lack of appropriate training. The detonation created a situation in which smoke rapidly accumulated in several areas of the submarine. The subsequent enquiry noted that “neither the pre-deployment training package nor general …safety training exercises a scenario whereby such a large volume of the submarine atmosphere is out of specification from the onset” [6]. This created a mismatch between the conditions that the crew experienced and their previous training. In other words the environment following the incident created a ‘degraded mode’ of operation, which forced a range of coping strategies that had not been considered in training. For example, standard emergency station actions focused on getting dressed and stowing loose gear. These actions were inappropriate, if not impossible, without first obtaining and then connecting their Emergency Breathing System masks. Inappropriate SOPs can increase the risks associated with emergency response to adverse events. Equally a failure to provide and rehearse SOPs can lead to an uncoordinated response in the aftermath of major accidents [9].

Other risks associated with improvised emergency response can be seen in the attempts that the crew made to open the hatch doors to the Forward Escape Compartment (FEC). These were buckled by the force of the initial explosion. It was critical to open these doors. The surviving crew members in the FEC struggled to put out fires that had been started by the material that had been propelled from the ruptured SCOG. However, members of the damage control team could not find appropriate equipment and had “to open the hatch doors to the FEC utilizing various pieces of equipment as improvised tools” [6]. There were insufficient crowbars and hacksaws. One reason for this is that the damage control scenarios used to train submarine crews did use these tools as much as emergency management drills did in the surface flotilla. Rescue teams were forced to remove a ladder so that they could access the buckled hatch doors. Crewmembers managed to stand on drums and peer into the FEC where they could finally assess the extent of the injuries to their colleagues. They were also able to pass back information about the damage created by the SCOG. Some forty four minutes after the initial explosion, the crew managed to force one of the hatches in the opposite direction to the way in which it normally would have opened. They then tore it from its hinges. The ladder was replaced and teams could enter the FEC.

By this stage, readers should have some understanding that military life creates an expectation of system failure and poor design. Personnel are used to finding ways of working with communication problems and inappropriate SOPs. They expect doctrine to be ambiguous and incomplete. For instance, the Fleet produced a Training Directive for Under Ice Training. This made extensive references to a capability training directive that had not yet been published. Hence, senior officers had to infer elements of the Pre-Deployment Training scenarios. Fleet also developed Command Guidance documentation for Under Ice operations. This arrived too late for the Flag Officer for Sea Training to use it in helping the crew of the Tireless prepare for their operation. The Submarine Support Integrated Project Team had reviewed existing materials used to prepare for ice missions. They had also assessed the previous guidance on environmental impact. However, there was no coordinated safety review or risk assessment prior to the Under Ice deployment. In consequence, it was difficult for senior officers to determine whether or not the crew of the Tireless had sufficient training for the range of hazards that they might be faced with.

It can be difficult to maintain under ice expertise in a crew. There may be significant intervals between these missions and the hazards will change with equipment modifications and alterations in SOPs. In this accident, further problems were created by a reduction in pre-deployment training from a maximum of 5 days down to 3. Ironically, this decision was taken for that Tireless could return to Her Majesty's Naval Base Clyde for repairs. Not only did system failures force changes in the operational procedures on-board the submarine. It also forced the Flag Officer for Sea Training to improvise and find 'work arounds' in order to complete the necessary Pre-Deployment training within a shorter period of time.

4.6.3 Communications Failures

Firefighting efforts were also hampered by equipment problems. A Self-Contained Firefighting Unit (SFU 90) was inadvertently deployed after the fires had been put out in the FEC. The crew struggled to divert the jet along 2-Deck and then directed it into the Junior Ratings bathroom. As the crew worked to stop the unit, the nozzle came off the hose allowing water to flow freely. Although this might seem like a minor issue, it is important to reiterate that the crew were working in cramped conditions in breathing apparatus without many of the usual communications systems with reduced visibility. The failure of the nozzle was also important to the subsequent investigation because it had already been modified following a similar failure during a fire in HMCS Chicotimi in 2004 [6]. This illustrates how even when previous accidents do trigger an investigation, there is no guarantee that a subsequent modification will mitigate the risk of future failure. The hose assembly failed in Chicotimi, was modified and then failed again. Fortunately, several of the crew in Tireless worked together to improvise a solution and put a kink into the hose that cut the flow. Their colleagues then isolated the supply on another deck. Once this was done, the unit was taken to the senior ratings' mess so that the missing nozzle could be repaired in case it was required again.

Communications systems are essential in coordinating any effective response to adverse events, such as those in the Tireless. The failure of these applications creates two different sorts of risk. Firstly, crew members must find improvised solutions to restore contact with their team mates. Secondly, the lack of efficient communications channels also delays or frustrates attempts to coordinate a flexible response to other systems failures. Both problems are apparent in the aftermath of the SCOG incident. Immediately after the explosion, one of the crewmembers who was trapped in the Forward Escape Compartment (FEC) heard the telephone next to the Forward Escape Tower ringing. The handset was missing. He then tried to use the handset in the canteen at the other end of the FEC. This was broken. In such circumstance, suffering from disorientation after the blast, injured and still trying to fight the fires, he could not find a suitable 'work around' to communicate with other areas of the vessel. Eventually, he realized that there was someone calling to him from the other side of the buckled hatch.

The damage caused by the blast also illustrates other hazards associated with degraded modes of communication. When the ward room filled with smoke, key officers had to move to the Switchboard Room. This created problems for other crew members who tried to relay information to them. Sometime later, an announcement was made using the main broadcast pipe that they could be contacted on telephone extension 234. This enabled a direct line to be established between the firefighting teams and their coordinators. In the meantime, crewmembers could not brief senior staff about the problems that they faced in opening the hatches to the FEC; the APL Cromwell VHF radio was not working. The Board of Inquiry later concluded that the incident demonstrated these devices were not fit for purpose. Several of the following chapters in this book will reiterate similar criticisms about many of the radio systems being used across NATO forces. For now it is sufficient to observe that the crew of the Tireless were able to identify a further 'work around'. Rather than using the Cromwell system, they used the SR Mess DC Net telephone system while tethered to the end of a 7.5m breathing hose. This exacerbated the hazards associated with the couplings, which were mentioned in previous paragraphs.

The aftermath of the Tireless SCOG detonation reveals further limitations for military risk assessment. Very often the consequences of any component failure are considered only terms of the loss of functionality or service that would otherwise be provided by that component. In this case, the loss of the Cromwell VHF radio had a marginal impact on communications. The risk was mitigated by the DC Net. However, very few risk assessments consider the increased workload or reduced situation awareness that can be directly associated with the failure of the primary system. In this case, considerable time was wasted as crewmembers continued to try and use the Cromwell radio system even though it had failed. It can also be dangerous to rely on redundancy as mitigation for many risks [11]. Further problems arose because pipes were inaudible due to the noise created by attempts to open the hatch and by the damage forward of the 29th Bulkhead. Again, the crew was forced to improvise communications channels and messages were relayed between officers and the fire fighting teams by word of mouth.

These examples show that military risk assessment is complicated by the integration of many different systems. This leads to compound risks. The damage created by the SCOG exposed problems in the storage and provision of emergency breathing apparatus. It also revealed the lack of appropriate tools for opening hatches. The initial explosion damaged primary communications systems. In each of these examples, crewmembers found ways to 'work around' the hazards created by poor design, inappropriate SOPs or by the damage caused by the incident. However, it is clear that these 'ad hoc', flexible responses may also have exposed individuals to additional hazards. If the decision had been made to respond to the flood alarm by closing the 29th Bulkhead then it is likely that more lives would have been lost from the lack of masks and couplings to the Emergency Breathing System. If the fire had developed then the lives of damage control teams might have been in peril from the passage of information by 'word of mouth' in high-stress situations.

4.7 Degraded Modes and the Compound Risks of Military Operations

This chapter has used a detailed case study to illustrate the interactions between initial risk assessments and what are termed 'degraded modes' of operation. Degraded modes force military personnel to find coping strategies that help them deal with the many different failures that complicate military life. In most cases, these ad hoc adaptations do not threaten safety. However, personnel often struggle to cope with the compound risks that arise when one failure triggers multiple problems in the operation of other complex systems. One of the reasons for this is that the initial failure acts as a catalyst for a host of other underlying problems, including inadequate maintenance, poor design, ambiguous SOPs etc. If these underlying failures are not addressed then they can gradually erode the barriers and other defensive measures that prevent hazards from arising.

These arguments have been illustrated by a recent accident that led to the deaths of two members of the UK Royal Navy on-board a submarine. At the time of the incident they were participating in under-ice training. The mission formed part of wider tactical evaluations with the US military. Inadequate risk assessments during system integration and maintenance were compounded by a lack of pre-deployment hazard analysis. Problems in the handling and storage of Self Contained Oxygen Canister seem likely to have resulted in a 'significant liquid organic contamination of the SCOG sodium chlorate block due to inadvertent ingress of oil inside the SCOG canister' [6]. These handling practices emerged as a means of coping with the need to store large numbers of SCOG devices for 'ready use'. They created a compound risk because the 'ready use' SCOGs were intended as a means of mitigating the failure of low pressure electrolysers when ice formed in the hydrogen discharge piping.

Inadequate risk assessments not only contributed to the cause of this accident but also complicated the response to the emergency. Crewmembers were forced to find numerous ways of coping with design flaws in their breathing systems. The Emergency Escape Breathing Devices were so complicated and difficult to use that teams preferred to use the Emergency Breathing System even though there were insufficient masks. Individuals also had to cope with numerous equipment failures – including the loss of the Cromwell VHF. This was compounded by the difficulty of hearing the piping system over the background noise in critical areas of the vessel. Further problems stemmed from the failure of fire fighting and rescue equipment, most notably when the nozzle of the SFU-90 system broke in a similar way to previous failures. In all of these instances, crew members were forced to adopt ad hoc solutions that were not anticipated in any Safety Case. The limited supply of Emergency Breathing System masks persuaded some crew members to crawl under the smoke while others shared their devices. The failure of communications systems forced teams to rely on word of mouth being passed along chains from command centers to the FEC. The failure of the SFU-90 nozzle was only rectified when team members put a kink in the hose to stop the water from causing further damage.

These ‘coping strategies’ had to be improvised in cramped conditions, with high levels of noise, low levels of visibility and significant quantities of smoke. It can be argued that no risk assessment can ever prepare a crew for incidents of this nature. However, it is clear that significant steps can be taken to learn the lessons provided by the Tireless incident. Firstly, many of the limitations of the equipment mentioned above were well known. However, information was not always provided to those who needed it most – for instance about correct handling procedures for the SCOG devices. In spite of significant investments within the UK MOD to promote ‘lessons learned’ systems, previous SCOG related incidents did not trigger revisions tom the initial risk assessments or safety arguments. Secondly, there were significant weaknesses in pre-deployment training. These exercises had themselves been curtailed as a result of systems failures, which forced the submarine to return to HMNB Clyde. A key insight from this study is that military organizations should reduce their tolerance of routine operational failures. This applies to the problems introduced in the acquisition of military systems, just as it applies to design failings, handling errors and even the planning of pre-deployment training.

4.8 References for Chapter Four

[1] C.W. Johnson and C. Shea, The Contribution of Degraded Modes to Accidents in the US, UK and Australian Rail Industries. In A.G. Boyer and N.J. Gauthier (eds.), Proceedings of the 25th International Systems Safety Conference, Baltimore, USA, International Systems Safety Society, Unionville, VA, USA, 626-636, 0-9721385-7-9, 2007.

[2] C.W. Johnson and C. Shea, The Contribution of Degraded Modes of Operation as a Cause of Incidents and Accidents in Air Traffic Management. In A.G. Boyer and N.J. Gauthier (eds.), Proceedings of the 25th International Systems Safety Conference, Baltimore, USA, International Systems Safety Society, Unionville, VA, USA, 616-626, 0-9721385-7-9, 2007.

[3] A. Chappell and H. Peck, Risk Management in Military Supply Chains: Is there a role for Six Sigma? International Journal of Logistics Research and Applications, (9)3:253-267, September 2006.

[4] United States General Accounting Office, Report to the Chairman, Subcommittee on Investigations and Oversight, Committee on Science, Space, and Technology, House of Representatives, Patriot Missile Defense: Software Problem Led to System Failure at Dhahran, Saudi Arabia, Februart 1992. Available on http://archive.gao.gov/t2pbat6/145960.pdf, Last accessed 11th June 2009.

[5] C.W. Johnson, Act in Haste, Repent at Leisure: An Overview of Operational Incidents Involving UAVs in Afghanistan (2003-2005). In P. Casely and C.W. Johnson (eds), Third IET Systems Safety Conference, NEC, Birmingham, UK, 2008, IET Conference Publications, Savoy Place, London, 2008.

[6] Board of Inquiry, Report of the Board of Inquiry into the Circumstances Surrounding the Deaths of LOM(WSM) Paul David McCann and OM(WSM)2 Anthony Huntrod on HMS Tireless on 20/21 March 2007.
http://www.mod.uk/NR/rdonlyres/08A7D08E-E092-4159-94C3-9A26E182A01B/0/boi_hmstireless.pdf, Last accessed 11th June 2009.

[7] Joint Service Publication, MOD Ship Safety Management: Part 3: Naval Authority Regulations, Issue 4, UK Defence Council, Ministry of Defence, London, United Kingdom, JSP 430, September 2007.

[8] Ministry Of Defence, Reliability And Maintainability Data Collection And Classification, Part 1: Maintenance Data & Defect Reporting In The Royal Navy, The Army And The Royal Air Force Defence Standard 00-44 (Part 1)/Issue 2, 9 June 1995.

[9] C.W. Johnson, A Handbook of Accident and Incident Reporting, Glasgow University Press, Glasgow, Scotland, 2003.

[10] C.W. Johnson and C. Shea, A Comparison of the Role of Degraded Modes of Operation in the Causes of Accidents and Rail and Air Traffic Management. In 2nd IET Systems Safety Conference", The IET, Savoy Place, London, UK, ISBN 978-0-86341-863-1, 89-94, 2007.

[11] C.W. Johnson, The Dangers of Interaction with Modular and Self-Healing Avionics Applications: Redundancy Considered Harmful. In J.M. Livingston, R. Barnes, D. Swallom and W. Pottraz (eds.) Proceedings of the 27th International Conference on Systems Safety, Huntsville, Alabama, USA 2009, International Systems Safety Society, Unionville, VA, USA, 3044-3054, 2009.

5 Systemic Risks of Fatigue in Military Operations

Previous chapters have identified a host of factors that complicate the use of risk assessment techniques to support military decision making. Risk assessment during procurement is complicated because many hazards only emerge after systems have been deployed. Previous accidents and incidents have shown that risk assessments are not updated with the operational information that is often captured in lessons learned systems. Further problems stem from identifying the compound risks that arise from the degraded modes of operation and routine system failures that characterize military life. At an operational level, personnel often have to act under extreme time pressure and with limited information about the threats that they must face. This chapter extends the analysis to identify the impact of fatigue on decision making. Fatigue increases the risks associated with different forms of human error [1]. It can also undermine attempts to accurately assess the risks that arise during military operations [2].

5.1 Introduction to Human Error

There is an increasing recognition that part of the responsibility for human error lies with developers and management, not just with the end users of military systems [3]. This joint responsibility can be illustrated by a simple example. A recent accident involving an US Army M985 Heavy Expanded Mobility Tactical Truck illustrates this joint responsibility [4]. The two individuals involved in the accident had spent the day preparing for a night tactical road march involving a 73-vehicle convoy. These preparations provided few opportunities for them to rest before they left for the marshalling area around 21:30. With 6% moon illumination and 2-mile visibility, the crews used Night Vision Goggles (NVGs). They linked up with the rest of the vehicles in their section and at 24:00 the company commander conducted a convoy leaders briefing. By 02:30, the section came to a halt as vehicles waited to refuel. Several drivers fell asleep during this unscheduled stop. Partly as a result, gaps began to appear in the convoy when it eventually started to move again. Shortly afterwards, the truck drove off the left side of a tank trail and partially turned over in a 4-foot deep stream. One of the vehicle occupants was drowned. This accident could have been blamed on driver error. However, the resulting investigation concluded that the principle cause was inadequate risk assessment. Given the size of the convoy, the company commander delegated responsibility to squad leaders for ensuring that personnel had sufficient rest before the mission. However, the squad leaders did not continuously review rest patterns on the day before the accident. Lack of sleep was compounded by the demands of driving using NVGs. This joint responsibility for error extends beyond the individual and their commander to include the designer of military systems. Even if personnel end-users have opportunities to take breaks, errors will still happen if people are expected to work in noisy environments or cramped conditions against tight deadlines.

5.1.1 Resilience and the Myth of Error Free Performance

It is difficult to understand the risks that fatigue and human error create for military operations unless we first consider the strengths that individuals bring to military operations [5]. It is important to recognize that people are particularly good at coping with failure. We constantly adapt to problems in our working lives. The previous chapter has provided numerous examples of the ways in which personnel learn to operate poorly designed systems. These coping mechanisms help to explain why we find so many legacy applications that continue to be used even though they show no evidence of even the most basic human factors involvement in their design. Over time pesonnel learn through their mistakes [6]. They find new ways of doing frequent tasks so that they avoid known problems in poorly designed or faulty equipment. Hence, failure is a necessary component of learning. The flexibility of human behaviour helps to explain why so many of the hazards identified in a risk assessment are never realized.

A great deal of attention has recently been devoted to the topic of 'resilience engineering' [5]. This shifts the focus away from the causes of human error and more towards the promotion of recovery actions. The proponents of this approach argue that we cannot anticipate all of the hazards that might arise in complex military operations. In consequence, we should enable personnel to cope with a broad range of potential failures. For example, this would encourage the development of multiple alternate communications systems that could be used in the aftermath of an accident such as the explosion on HMS Tireless, discussed in Chapter Four. Resilience engineering starts from the assumption that humans are not simply the cause of error, they act as a key means of mitigating failure in complex systems [7]. This creates considerable problems in trying to reconcile preliminary risk assessments with the date derived from operational experience given that individuals and teams actively work to reduce the consequences of any failures that do occur. Resilience engineering is a relatively novel concept. There are few established tools and techniques and only limited applications of the approach to military systems. Later chapters will return to this issue. In contrast, the remaining sections in this Chapter focus on the interaction between human error, fatigue and the hazards of military operations.

5.1.2 Violations, Slips, Lapses and Mistakes

Figure 11 provides an overview of the relationships between violations, slips, lapses and mistakes. A violation occurs when users knowingly break a rule [6]. In military operations, these include generic 'Army regulations' through to the more detailed Standard Operating Procedures (SOPs) that govern particular operations down to individual orders. The distinction between errors and violations is not always as clear as it might seem. For instance, individuals can unwittingly violate rules if they are unaware of them or if the rule is not clearly expressed. In such circumstances, it can be argued that an error has occurred rather than a deliberate violation. In other words, it is difficult to distinguish between errors and violations because these two forms of failure stem from different intentions even though the observable actions can be identical.

.

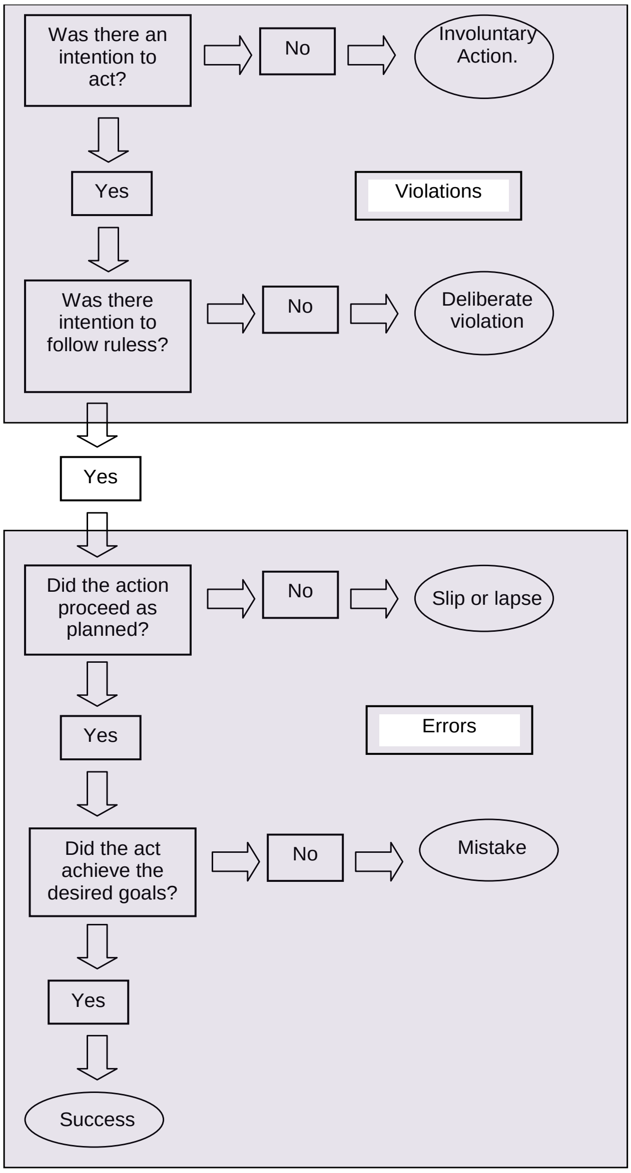

Was there an intention to act?
No
Involuntary Action.
Yes
Violations
Was there intention to follow ruless?
No
Deliberate violation
Yes
Did the action proceed as planned?
No
Slip or lapse
Yes
Errors
Did the act achieve the desired goals?
No
Mistake
Yes
Success

Fig. 11. Relationships between Violations, Slips, Lapses and Mistakes

Figure 11 distinguishes violations from the broader class of errors that include slips, lapses and mistakes. As can be seen, if there was an intention to act and to follow established rules but the action did not proceed as planned then there may have been a slip or a lapse. A lapse is an omission. An individual forgets to do something. For example, a pilot might forget to lower their landing gear during an approach. In contrast, a slip occurs when additional actions might inadvertently be introduced into a task or when key steps are executed in the wrong order. For instance, the pilot might attempt to lower the landing gear before a descent has begun. Slips and lapses occur even though the individual has a correct plan of action. However, a mistake occurs when a plan is correctly executed but it fails to deliver the intended outcome. Slips and lapses are often best addressed by introducing forms of checks. For instance, a co-pilot might monitor the pilot's actions. Training and the use of simulators as well as procedural cues, such as task cards, can help to reduce the likelihood of mistakes.

5.1.3 Task Based Approaches to Error

A number of different approaches use the distinctions illustrated in Figure 11 to introduce human error estimates into risk assessments [6, 8]. One of the most common techniques is to break down common tasks into their component stages. Any deviation from the ideal pattern of task completion can be regarded as an error. Human factors experts can then inspect the products of this task analysis to identify those activities that are most likely to be affected by a slip or a lapse. Similarly, an analysis of previous operations and training performance can be used to identify common mistakes. The consequences of these slips, lapses and mistakes can then be assessed just as they would be for component or system failures.

The task-based approaches to human reliability analysis work best in procedural tasks where it is possible to identify common patterns of working. For instance, they are widely used within the nuclear industry where regulators constrain the ways in which operators can interact with the underlying systems [8]. Unfortunately, this approach does not work so well in military operations. It can be difficult to exhaustively prescribe the stages of many complex missions. Very few units ever perform operations in exactly the same way that even their own Commanders might envisage. There are also tremendous variations in performance between different individuals. Hence, what might seem to be as optimal behavior for a new recruit might be seen as an 'error' for more experienced troops.

Instead of defining errors to be a departure from optimal performance, they might instead be defined as a departure from 'normal' operation. This definition acknowledges that we might expect tremendous variation between different individuals. It, therefore, addresses one of the problems that arose in comparing performance with a 'perfect approach'. What might be a 'normal' error to expect from a recruit might not be expected for more experienced personnel. This definition also acknowledges that we might be able to recognize unusual or erroneous performance that differs from expected behavior. However, a number of theoretical and practical problems complicate this approach. For instance, we tend to commit hundreds of small slips and lapses every day. This is a consequence of the key role that errors play in learning. As we have seen in Chapter Four, most people deploy 'coping strategies' so that they can recover from tiny failures within minimum disruption to their higher level tasks. Errors often occur in successful interactions in which individuals achieve their goal. In many cases, they may not even realize that an error has occurred.

5.1.4 Human Reliability Analysis

Human reliability analysis extends the task-based approach by providing numeric estimates of the likelihood and consequences of slips, lapses and mistakes. Quantified risk assessments works best for hardware systems for which it is possible to derive statistical estimates of the probability of random failures over time. For example, the US Department of Defense publishes a number of military handbooks that contain the probability of failure for many different hardware components [9]. Subcomponent failure rates can be combined to derive system level probabilities. In the same way, Swain and Guttman [10] have sought to publish handbooks of human reliability that provide high-level estimates for the likelihood of particular types of error. For example, the probability that a soldier incorrectly reads back a series of figures from a display might be 1 in 200 attempts. This approach has numerous advantages for the engineering of complex systems – the same risk-based approaches can be applied throughout all aspects of the development process.

5.1.5 Performance Shaping Factors

A number of concerns limit the practical application of human reliability analysis. Firstly, critics of the approach have argued that the probabilities can be difficult to validate [6]. It is difficult to derive operational information about error rates in different tasks. Laboratory based studies of military tasks raise further questions. If personnel know that their actions are being observed then they may be less likely to make an error; they will exploit a range of self-monitoring techniques to catch and rectify any mistakes that might jeopardize their tasks. Secondly, even if accurate data is available for previous errors in similar operations then high level estimates do not take into account a host of more complex cognitive and social factors. Hollnagel [11] voices this criticism when he describes Human Reliability Analysis as 'psychologically vacuous'. More recent methodologies have sought to address these criticisms by helping designers to first calculate the base probability for particular errors and then apply modifying terms to equations that account for the impact of what are called 'performance shaping factors'. These include fatigue, heat, alcohol and drug abuse, noise, stress etc. Performance Shaping Factors are defined to be any aspect of the individual characteristics, environment, organization, or task that improves or impairs human performance causing a consequent change in the likelihood of human error. For example, if an individual must carry out a mission under time pressure then the probability of incorrectly reading back some data is increased. If they have prior training in this task then a delta may be applied to reduce the probability of such errors. Although these developments help to address caveats about the application of Human Reliability Analysis, they also raise further concerns about the validation of both the base probabilities and also the 'fudge factors' that account for performance variations.

5.2 Fatigue and the Risk of Error

Performance shaping factors complicate the use of human reliability estimates for the risk of human error in military operations. Analysts must consider a host of individual and environmental factors that influence the likelihood of slips, lapses and mistakes. In contrast, the following sections in this Chapter go on to look at one performance shaping factor in additional detail. The decision to focus on the impact of fatigue on military risk assessment is justified because this is often cited as a causal factor in previous incidents [2]. Accidents can occur in relatively routine operations through a lack of continued monitoring of shift patterns. This is because fatigue influences different levels within military operations including front line units, logistics, as well as strategic and tactical command. It is, therefore, important not to view fatigue as a problem in isolation from other concerns. It is a contributory factor in a host of adverse events ranging from poor operational risk assessment through communications failures to poor situation awareness.

5.2.1 Problems of Self-Diagnosis

One of the reasons why it is difficult to assess the risk of uman error during military operations is because it is hard for individuals to accurately assess the level of fatigue affecting both themselves and the soldiers within their units. This is a recursive problem. Fatigue undermines an individual's ability to identify when their performance will be affected by fatigue. As personnel become increasingly tired, they are less able to identify the signs and symptoms of that tiredness [12].

In military operations, the difficulty in self-diagnosing levels of fatigue can be compounded by the same strong individual motivation to complete operational objectives that were identified under the degraded modes of operation in Chapter Four. For example, a US Army helicopter crew prepared for a training exercise by sleeping until mid-morning. By 16:00, they had finished a preflight inspection but then received immediate instructions to return to their home base. The pilot took off at 17:30 and arrived back by 22:30 only to be told that their unit was preparing for immediate deployment. He helped to complete preparations by 00:30 and was told to get some sleep but to be on the airfield by 05:00. The pilot reported that he only managed to sleep for less than an hour given the stresses involved in preparing for deployment. He also struggled to gain any sleep while a fixed wing aircraft transported them and their machines to their destination. On arrival, the unit took part in intelligence and threat briefing. This illustrates the difficulties of coordinating operational risk assessments during such operations. The crews then began detailed mission planning and map studies. By 24:00 on the second day, the unit was moved on transport aircraft to a forward staging base with further meetings being conducted during this flight. They landed at 02:30 and then helped to unload and prepare their aircraft. By 06:00 the pilot and his colleagues were waiting for clearance to takeoff. He concluded “In the preceding 46 hours, I could remember really sleeping for only 1 hour. As it turned out, I would not sleep again until approximately 22:00 that night, when I collapsed in exhaustion. My first day of combat had added another 16 hours of wakefulness, bringing the total to 62 hours. I couldn’t help wondering what I would have done if the mission extended any further before I could get some sleep” [13].

The limited time available for risk assessment in this operation was compounded by lack of sleep and rising levels of fatigue. Such examples illustrate the practical problems that undermine many of the principles expressed in the US Army’s FM22-51 Leaders’ Manual for Combat Stress [12]. This guidance reiterates the need for military personnel to continuously monitor the possible impact of fatigue on their teams. However, this may not be possible when competing demands force the rapid redeployment of scarce resources between training and operations. Such incidents illustrate the importance of viewing fatigue as a ‘systems problem’; one that arises from the interaction between different organizational demands rather than one which is the result of negligence or carelessness by a single individual.

5.2.2 Circadian Rhythms

A number of factors influence the risks associated with fatigue during military operations. For example, circadian rhythms affect a wide range of functions, including performance, alertness, behavior, and mood. The majority of people feel the greatest need for sleep between 03:00 and 05:00 and again between 15:00-17:00. In contrast, the two periods of greatest alertness are between 09:00-11:00 and 21:00-23:00. Many different factors affect the synchronization of these episodes of alertness and sleepiness. It can take more than three weeks for some individuals to adjust to a new time zone or to adapt to a new shift pattern. The period of adjustment is influenced by the degree of disruption and a range of external factors, including exposure to light.

These disruptions have contributed to a range of military incidents and accidents. For example, the Canadian Forces recently lost a Sea King helicopter during a deck-landing and C-6 gun training mission [14]. The aircraft rose to the high hover position and then suffered a loss of lift. The helicopter hit the flight deck and rolled over. Although the cause of the accident was traced to a compressor stall in one of the engines, the subsequent investigation identified a number of ways in which circadian de-synchronism contributed to the incident. At the time of the accident, the aircrew was on board a vessel in an operational ‘work up’ stationed in the Arabian Gulf. The evening before the flight, the ship advanced clocks from midnight to 01:00 and then immediately practiced an emergency fire drill. This required the entire ship’s company to be awake and active for approximately one hour. Because the emergency fire drill had disturbed the aircrew rest schedule, the Helicopter Air Detachment

Commander rescheduled the flights so that the first operation took place at 10:45 rather than 08:00. The official investigation argued that the effects of the early fire drill coupled with the time zone change may have prevented the crew for identifying signs that something was wrong with the aircraft before the accident. These included a delay in the engine starting and water spray pooling in front of the engine in-takes; "in light of these events, it is possible that 'circadian disruption' may have affected the aircrew during their assessment of whether or not to perform engine-related maintenance or to even continue with the launch" [14]. This incident again illustrates the 'systemic' nature of fatigue in military operations. The fatigue that prevented the crew from responding effectively to the incident as it developed also inhibited their ability to identify potential risks prior to take-off.

5.2.3 Other Performance Shaping Factors

One of the problems in assessing the impact of fatigue on the risk of human error is that it interacts with other performance shaping factors. In other words, levels of fatigue can be influenced by the noise, vibration and time-stress that often characterize military operations. Recent US military engagements have placed personnel in relatively hot environments. Sleep patterns do not, in general, improve over time where heat disrupts restorative rest intervals [15]. Perceived levels of fatigue also increase when the influence of heat is combined with a requirement to perform monotonous or repetitive tasks. For instance, many recent accidents have occurred during convoy driving while personnel have been wearing Kevlar helmets and body armor.

Other external factors, including noise, can both induce fatigue and prevent sleep. For example, a Canadian reservist was acting as a driver. She worked all day prior to helping with the move to an exercise area, attending military training classes before reaching bed around midnight. She was then assigned Fire Picket duty from 03:00-04:30. Her immediate superior requested that she be taken off the duty but he was overruled by his superior. She tried to sleep after her duty but was prevented by the noise coming from the Armory. She was then awoken shortly before 05:00 so that the unit could leave around 09:00. Within an hour of departure, she fell asleep at the wheel of a utility vehicle. It left the road, hit a culvert, flew through the air and rolled several times. She suffered multiple injuries and was unable to work or continue training for several months [16]. This accident illustrates how the demands on particular soldiers are exacerbated by the difficulty of sleeping in many operational environments. It also again illustrates the role that leaders play in tackling the potential consequences of fatigue given that her immediate superior was over-ruled in his attempt to provide additional rest for his troops.

Environmental factors, including extreme heat or cold, reduce the effectiveness of self monitoring as a means of detecting the risk of fatigue. A recent review from the US Army Combat Readiness Center on the interaction between heat and fatigue argued that "performing sentry or fire guard duty, surveillance activities, monitoring instruments, and operating a vehicle all demand vigilance. Temperatures higher than

85 degrees Fahrenheit with 63 percent relative humidity affect the vigilance of Soldiers, even those well acclimatized to the heat. It's important that commanders recognize this limitation and take steps to ensure their Soldiers get adequate breaks from extended duties" [15].

5.3 Consequences of Fatigue

The hazards associated with fatigue in military operations include: difficulty in thinking clearly; poor performance; greater tolerance for error; inattention to details; increased lapses of attention; increased irritability; decreased motivation, attempts to conserve effort; increased errors; slow and irregular reaction times; impairment in communicating and cooperating with other soldiers, headaches or stomachaches; poor morale [4].

5.3.1 Inability to Self-Monitor Fatigue

How et. al. [17] have shown that higher subjective assessments of 'sleepiness' are associated with poorer performance. However, it can be difficult to validate the insights that are obtained from self-reporting of fatigue. Itoi et al [18] asked sleep-deprived participants to predict whether they would fall asleep over the next 2-minute interval on a scale from 0% to 100% likelihood. On those occasions when a participant did fall asleep within the 2 minute interval, the average likelihood estimated by the participants was only 55%. In other words, they did not assign a very high likelihood to sleep immediately before they fell asleep. This has considerable practical implications; it can be difficult for military personnel to accurately assess their level of fatigue as they become increasingly tired.

Chapter Three has described how Composite Risk Management encourages every soldier to consider the negative effects of fatigue as part of the planning for every operation. This can be illustrated by a recent incident involving a KC-135 air refueling tanker in Iraq. The pilot inadvertently transferred too much fuel from one side of the aircraft causing a potential imbalance. The incident took place after the crew had completed twelve 6-7 hour missions with their associated briefings, operational risk assessments and debriefings. As in many military incidents, it is only the realization that an accident was averted that eventually forced the crew to accept how tired they had become; ''I don't think you really know the fatigue that sets into your body until you are finally able to rest. Even though I was getting enough rest at night ... my body and senses became very numb. I truly believe, because of the demand for the missions in Operation Iraqi Freedom, our crew had become so tired that we forgot the little things, which can add up to big things" [19].

5.3.2 Microsleeps and Encysting

Increases in subjective fatigue are associated with a rise in the frequency of both slips and lapses. These effects are compounded by the self-monitoring effects mentioned above. Errors of omission and of commission become harder to detect with increasing levels of fatigue. Further problems arise when personnel suffer from ‘microsleeps’. These lead to a loss of attention that lasts from 0.5 to 10 seconds. In continual monitoring tasks, this is sufficient time for military personnel to lose critical components of situation awareness.

Fatigue is also associated with the problem of encysting. This occurs when soldiers become preoccupied with one particular task to the detriment of overall situation awareness. This contributed to the loss of a transport aircraft at the Guantanamo Naval Base. The crew had flown overnight cargo schedules for two nights before the accident. Shortly before they were due to be released from duty, they were allocated the accident trip. During the crash, the pilot was so focused on finding a strobe light that he failed to respond to other crew members’ warnings that there was a risk of stalling [20]. By focusing on the individual behavior of the pilot, there is a danger that we will overlook the wider crew interactions that contributed to this incident. However, there have been very few studies into the team-based hazards of fatigue in military operations.

5.3.3 False Responding

As mentioned previously, fatigue increases the risks associated with errors of commission. These include the ‘false responding’ that occurs when individuals react to changes in their environment either in inappropriate ways or in response to signals that did not occur. These problems do not arise in isolation; hence some individuals in a team may suffer from false responding at the same time as their comrades suffer from lapses of attention. Many adverse events stem from errors of omission and commission that involve the same individual. This can be illustrated by accidents involving F-16s in the Gulf. The first occurred when an aircraft collided with a stationary F-16 during taxiing. The moving aircraft then suffered a hydraulic failure, which removed control of the nose wheel. The pilot noticed the hydraulic problems but did not set the parking brake and instead continued to taxi repeatedly activating the brakes as he went along the taxiway. This bled off pressure from the brake/jet fuel starter accumulators, eventually damaging the brakes. The safety investigation board found “mental fatigue caused by sleep issues, dehydration, and hunger slowed response time, decreased performance and led to distraction” [21]. The initial collision was arguably caused by a lapse, which was then compounded by a form of false responding as the pilot continued to taxi without parking the aircraft.

The second mishap involved a Canadian F-16. Trapped fuel in an external tank led the pilot to refuel. During this procedure they inadvertently disabled the aircraft’s low fuel warnings. Later in the mission, the pilot failed to conduct an in-flight operational “Dash One” check that would have caught this error. The pilot eventually identified the potential lapse and realized they were short of fuel. He attempted another refueling, but the engine flamed out due to fuel starvation. The pilot ejected and was recovered by rescue forces. The safety investigation board argued that the pilot failed to recognize the state of the aircraft because he was suffering from chronic mental fatigue. These were again due to systemic factors beyond the control of the individual pilot; “Inadequate quarters, the recent change in circadian rhythm, as well as the contingency operations tempo contributed to inadequate pilot rest” [21].

5.3.4 Poor Decision Making

An investigation of fatigue related accidents and incidents involving the US Air Force’s C-5 fleet reported that 55% were related to problems of attention, including lapses and false responding. A further 24% stemmed from ‘decision making problems’ including inadequate risk assessments [22]. The interactions between fatigue and military decision making have also been explored under ‘laboratory conditions’. For instance, one investigation looked at the reactions of young, severely stressed, sleep deprived military personnel when ordered to fire with live ammunition at ‘real people’ rather than the targets, which they had believed would be involved in the exercise. The majority of the participants fired their weapons. Only one student tried to warn his colleagues when he observed that there were people in the target area [23].

Other studies have looked at the impact of fatigue on decision making during particular operations. For example, the Evaluation of Risks (EVAR) technique has been used to assess the acceptance of risk by personnel involved in a maritime counter-terrorism exercise [24]. The EVAR approach relies upon a series of questions where participants are asked to draw the position on a 100mm line that best indicates their present feelings between two extremes such as . ‘I seek the thrill of danger’ or ‘I seek tranquillity’. These questions are used to assess 24 different items that are grouped to provide information on five factors that are interpreted to help shape risk based decision making: self control, danger seeking, energy, impulsiveness and invincibility. The results are scored by measuring the distance in millimeters from one of the poles to the participant’s mark. Sicard describes how ten pilots were submitted to strenuous night flights with limited sleep deprivation. Their results were compared to a control group that participated in the maritime counter terrorism exercise but who were otherwise well rested. Compared with this baseline data, the pilots reported an increase in impulsiveness. Correlations were also observed between mood and alertness and risk factors. The results indicated that risk taking, decision making, and stress factors were strongly associated during this exercise; “we observed change in risk proneness occurring within a short time under harsh conditions” [24].

The increased impulsiveness revealed from the EVAR study is illustrated by the following case studies where fatigue increased the risks associated with poor judgment and decision making. These accidents also illustrate the dangers that can arise when tired personnel eventually try to make up for a sleep deficit. A Canadian soldier had been involved in several days of high intensity training. He, therefore, took the opportunity to sleep in a military truck. He started the engine so that he could keep the heater on. He was subsequently told to move into position but he collapsed as soon as he got out of the cab. Medical personnel soon determined that the soldier was suffering from carbon monoxide poisoning and dehydration. He was extremely fortunate to be woken before he lost consciousness [25].

A section had been riding all day in an Armored Assault Vehicle Command (AAVC) [26]. By 21:00 the unit was ordered to rest. However, the location for their AAVC was constrained by the need to maintain good radio communications with the regimental Tactical Operations Centers. A Staff Sergeant set up a security plan for their position and took the first watch, the radioman then retrieved his gear and bedded down five meters behind the AAVC. At 22:00, the staff sergeant looked for the radioman to take a watch but ended up posting another soldier when he could not find him. Some three hours later, a fuel truck arrived at the group's position. Two members of the unit carefully searched the position with flashlights and then gave the AAVC driver permission to move 100 meters down the hill for the refueling operation to take place on the side of the road. Two soldiers guided the vehicle at walking pace between a number of boulders and other obstacles. The refueling was completed by 02:00 and the AAVC moved back to its original position driving over the body of the radioman on the way. Such incidents illustrate the practical difficulties of conducting risk assessments in the field when many individuals are simultaneously coping with rising levels of fatigue.

5.4 Countermeasures for Fatigue and Sleep Loss

Military personnel can employ a range of different measures to mitigate the risks of fatigue. These range from improved crew rotation to the use of drugs, both as stimulants or to promote sleep. All of these approaches have drawbacks. They tend to focus on the symptoms of the problem rather than the systems level causes. In consequence, it can be difficult to quantify the extent to which they can reduce the risks associated with many military operations.

5.4.1 Restorative Sleep

US Army FM22-51 [12] and FM 6-22.5 [27] advocate the maintenance of 'sleep discipline' in order to ensure that all soldiers maximize limited opportunities for rest under unpredictable circumstances. 6 to 8 hours of sleep should be obtained whenever possible, with a minimum of 4 hours uninterrupted or 5 hours interrupted sleep in every 24. These should be included within the standard 12 hours on, 12 hours off wartime shift that simplifies planning and staff rotation. Army doctrine acknowledges the need for flexibility in shift patterns where, for instance, 12 hours may be too long to maintain optimum performance in demanding tasks. Shorter shift patterns have to be balanced against the time that can be wasted as personnel move to rest quarters and dining facilities as well as the problems of ensuring the hand-over of sufficient information at the end of each shift pattern.

Many military operations provide limited opportunities for the continuous sleep that is required to mitigate the risk of fatigue. Similar problems have also been faced by NASA as they try to help astronauts maintain high levels of performance during long-duration space flights, where normal sleep patterns might become divorced from the usual circadian rhythms. The ‘NASA nap’ was, therefore, developed to combat the adverse effects of fatigue within safety-related operations. It lasts for 40 minutes but should not occur within four hours of a more sustained sleep cycle. However, it can be hard for personnel to ‘switch off’ enough to sleep. This is a particular problem if military personnel have recently used stimulants, such as caffeine, to reduce the impact of fatigue. Conversely, carbohydrates and sugary foods can be used to help induce sleep. Small meals that are rich in protein provide stop-gap measures to fight off fatigue. Exercise and hydration are also important in managing boredom over sustained periods of time. Restorative sleeps can, therefore, be seen as one component of wider ‘team management’ techniques that must also considers task prioritization, workload, nutrition, exercise etc.

Restorative rests cannot be relied upon under operational pressures when it is difficult for teams to diagnose that they are suffering from fatigue. This can be illustrated by a ‘near miss’ involving Canadian personnel flying out of Kabul. The crew was familiar with the area and this may have instilled a level of confidence that led them to omit necessary navigational planning. As a result they flew at low altitude into a box canyon. The crew performed a slow, tight turn that led the aircraft to stall. They managed to restart the engines at 250 feet above ground level. The investigation argued that the crews’ performance was adversely affected by ‘acute fatigue (‘jet-lag’) and chronic fatigue (‘sleep-debt’)’. In particular, the crew did not ‘exercise their option of calling a ‘‘time-out’’ – they perceived a definite pressure to get the job done. The operational imperative emphasized at the time may have created a mindset in the crew to push personal limits, thus unwittingly promoting skewed decision making processes. The crew did not advise their Chain of Command of their fatigued state in an effort to seek other risk mitigation strategies’ [28]. Fatigue combined with over-confidence and operation pressures to undermine their assessment of the hazards involved in navigating their way out of the box canyon.

5.4.2 Motivation and Task Rotation

Fatigue is often exacerbated by a small subset of the tasks that individuals are required to perform each day. It, therefore, follows that some of the adverse effects can be mitigated by rotating these tasks. This requires planning from senior personnel so that individuals have received sufficient cross-training to move between activities. One limitation with this approach is that shift handovers create potential risks when necessary information is not passed between personnel. Additional care must be taken when fatigue complicates communication at the end of one task and the beginning of another.

The effects of fatigue and boredom can be illustrated by an incident that was reported by a US Army pilot. He was supposed to be providing air support to a training exercise but the start was delayed and he was forced to maintain a hover; ‘…as we waited, boredom set in. I scanned with the target acquisition designation sight (TADS)—trying to find anything of interest—until my thumb was sore. The end of the mission was approaching quickly and the infantry we were supporting was finally situated and in need of our assistance locating and identifying the enemy. The infantry requested assistance from our company for an undetermined amount of time past the scheduled completion time, which happened to coincide with the official end of my flying duty day. The pilot in command (PC) was within his own duty day limitation because he had spent the night in the field and started duty well after I did. However, I didn't consider that my duty day was coming to an end because it always coincided with everyone else's duty day. A rookie mistake, I know, but such mistakes happen….I was tired, but I didn't know the extent of my fatigue until I caught myself doing the jello-neck head bob in the cockpit. I told the back-seater about falling asleep, and he said he knew because he had been watching my head tracker bob up and down as I fell in and out of consciousness. This should have been our first indication we should land or fly back to the assembly area. We didn't, nor did we discuss the need to. I tried to keep myself awake while the PC kept us at a hover, but I fell into a full sleep right before our company broke station to return to the assembly area. I think I was awakened by the radio call to break station…I'm no longer ashamed to say when I'm too tired to take an extension. *Fatigue in the cockpit is a risk that can't be mitigated with coffee or an instant energy drink. It can only be mitigated with the proper rest cycle*' [29].

This incident illustrates several of the systemic interactions between aspects of fatigue, human error and decision making that have been mentioned in previous sections of this Chapter. Inadequate planning and risk assessment led to a situation where the pilot extended the mission duration beyond safe limits. The pilot and his superiors underestimated the demands of repetitive and fatigue inducing tasks, especially sustaining a hover while waiting for the infantry. This analysis reiterates the observations in Chapter Four that many military risks stems from a culture of 'making do' and of 'getting the job done'. These same attitudes help to explain the large numbers of military incidents and accident that are associated with the adverse effects of operational fatigue.

5.4.3 Training and Monitoring

A number of strategies support cognitive functions when soldiers suffer from fatigue. These include read-back techniques or checklist reminders to confirm that instructions have been understood. US Army leadership manuals advocate 'over-learning' so that tasks become automatic. Key activities should be rehearsed until 'you can do it in your sleep' [12]. Many accidents and incidents are averted when colleagues notice that military personnel are beginning to show symptoms of fatigue. These include physiological changes, such as drooping eyelids or yawning, as well as psycho-social and cognitive effects, including irritability and forgetfulness. As we have seen, however, fatigue also impairs many of the cognitive functions that support the detection of performance problems in our colleagues. Initially, an individual may be able to identify potential 'errors'. This can lead to a sense of complacency as their level of fatigue increases. Less attention is then allocated to checking on performance deteriorates within a team. Attempts to increase self-motivation and effort can often be short-lived. These individual initiatives increase perceived levels of stress and can even exacerbate the underlying fatigue..

Crew selection techniques help to develop teams that are more resilient to the effects of sleep loss. However, it can be difficult to identify those attributes or markers that might be used to predict sustained performance under fatigue. The results from drills and exercises often provide poor indications of eventual ability under the stresses of military operations. Secondly, there are limitations in the individual differences that justify the use of crew selection criteria to combat the negative effects of fatigue. Although some people can resist the effects of fatigue for up to 24 hours, after 36 hours everyone suffers from measurable effects. The limitations of self-monitoring, of team and individual selection, of training have led many commercial and government agencies to seek alternate technological and pharmaceutical mitigations for the risks of fatigue.

5.4.4 Technological Countermeasures

Technology can be used to address the negative consequences of fatigue. For instance, Ground Proximity Warning Systems (GPWS) and Automated Ground Collision Avoidance Systems (AGCAS) have successfully reduced the number of controlled flight into terrain accidents in both military and civilian applications. However, these systems are less effective for ‘nap of the earth’ operations or night vision exercises. They tend to generate spurious warnings that can increase the risks of distraction during critical phases of flight. In the last two years, the US Army has lost five AH-64s in Controlled Flight Into Terrain accidents in Afghanistan and Iraq. Although fatigue was not a cause in all of these accidents, they all occurred in spite of the technological warnings provided by GPWS.

In highway operations, lane control systems have been developed to use video cameras as a means of sensing when a driver has inadvertently drifted out of position. However, this civilian technology has obvious limitations for military applications in areas with poor highways and few lane markings. Alternate ‘smart cruise control’ systems using short-range radar can be used to provide collision warnings. Although this technology has been deployed in a range of military vehicles, it suffers from further limitations. It is expensive to install and maintain. There are also calibration issues when units operate in convoys. The radar obstacle detection warnings may be continually going off even though drivers are alert to potential problems. This can create complacency if drivers learn to ignore alarms that warn about potential hazards. Spurious alarms also increase the temptation for military personnel to disable warning systems.

Chapter Six describes how night vision equipment can reduce some of the physiological, cognitive and perceptual hazards associated with prolonged monitoring during darkness. As with the technologies cited above, however, it can also contribute to adverse events especially when soldiers come to rely on these systems. For instance, the US Combat Readiness Center describes an incident in which an M1A1 had completed an attack and was moving to the assembly area. The driver and commander were both using NVGs to navigate but neither was using an approved scanning technique. They failed to see an unmarked fighting position and the tank slid into the hole. In another similar incident, a group of Bradley fighting vehicles were conducting reconnaissance in a desert environment with low contrast during a particularly dark night. They were expecting enemy fire and so were driving without any illumination using NVGs and GPS. As they approached their objective they came across what appeared to be two small ditches. They crossed the first then all three Bradleys went over a 15 foot cliff. Two soldiers were killed, and eight others were injured [30]. This incident again illustrates the problems of compound risk assessment in military operations. NVDs can reduce the risk of fatigue created by the need to mount nighttime operations. However, the risk mitigation provided by this technology must be offset against the novel range of perceptual and cognitive hazards that are created by these devices.

5.4.5 Drugs

Drugs can promote sleep. Ideally, they should have an immediate effect without any 'hangover' if the rest is interrupted for operational reasons. For example, studies have shown that 20mg, or twice the recommended dose, of zolpidem is required under simulated troop transport conditions to improve restorative sleep. At this level, the drug impairs complex mental operations 1.5 hours after administration. In further studies, twice the recommended dosage (0.5mg) of triazolam was shown not to improve the daytime sleeping of troops being deployed on a flight from the US to Europe during Operation Bright Star. Mental tasks were impaired for up to eight hours after the drugs had been administered. Another study focused on the use of triazolam by Ranger rifle platoons [31]. Doses of 0.5mg and 0.25mg were shown to improve sleep in the cold. Only the higher of the two doses was shown to impair performance 4 hours after it was taken. After 24 hours, the group given 0.25mg improved better in complex mental tests that those who received no drug to support sleeping. However, some soldiers fell asleep before they finished getting into their sleeping bags. This raises obvious concerns for the safe use of such drugs in harsh environments. In this case, the risk reduction from improved rest is outweighed by the compound risk of hypothermia, when drugs take effect before soldiers can seek shelter.

Rather than promoting sleep, stimulants provide short-term relief from the symptoms associated with fatigue. They are typically viewed as expedients that should only be used when it is impossible to create the conditions necessary for comfortable sleep [32]. The two principle approaches rely on caffeine or amphetamines. Present advice suggests that caffeine is best used for relatively short sustained periods of wakefulness up to 40 hours. Beyond 60 hours, it becomes less effective in reducing the symptoms of fatigue. Around 200mg of caffeine is required to maintain wakefulness, although the precise value depends on the individual's level of fatigue and background tolerance from daily use. Caffeine can be found in coffee (100 to 175 mg per cup), soft drinks (31 mg), tea (about 40 mg) and ‘over the counter’ stimulants (one tablet can contain around 65 mg of caffeine). An important benefit of caffeine is that in many armed forces, it is the only stimulant that can be used without prior approval.

Amphetamines have been widely used by the military as a means of combating fatigue. However, the social (mis)use of the drug has created a considerable need to monitor their consumption. The US Air Force first authorized amphetamines in 1961 and continues to permit the use of dextroamphetamine for some prolonged aviation operations. The US Army also authorizes the use of this drug to combat severe aviator fatigue. Dextroamphetamine, like many stimulants, improves self-ratings of energy level, vigor, and alertness. As we have seen from Sicard’s work [24], these factors are very likely to influence the individual risk assessments that guide military decision making. Amphetamines also interfere with recovery sleep and should be avoided within four hours of the soldier’s next sleep cycle [30]. There is, however, considerable controversy over the use of stimulants within the military. The US Army and Air Force Exchange Service (AAFES) banned the sale of products containing ephedra when a soldier at Fort Hood died from a heart attack and another suffered a heat-related injury during physical training [32]. The US Agency of Healthcare Research and Quality found insufficient evidence to suggest a causal link between such events and the use of ephedra [33, 34]. However, further deaths linked to the use of dietary supplements containing caffeine and ephedra have led to repeated warnings within the US military about their risks in combating extreme fatigue or enhancing physical performance.

5.5 Downwards Cycle of Fatigue

This Chapter has used a number of military incidents and accidents to identify the interactions between human error, fatigue and risk assessment. These interactions can be observed at two different levels. Fatigue increases the likelihood of operator error and hence must be factored into any operational risk assessments. Fatigue also affects our ability to conduct an accurate assessment of the hazards that might arise in a particular operation. There are also compound effects. For instance, not only does fatigue increase the likelihood of human error in military operations, it also reduces the likelihood that we will be able to diagnose the risk of fatigue related errors. Soldiers are often too tired to notice that they are too tired to complete an operation. There are also strong interactions with the ‘coping strategies’ that were identified in Chapter Four as military personnel drive themselves beyond safe limits in order to get the job done.

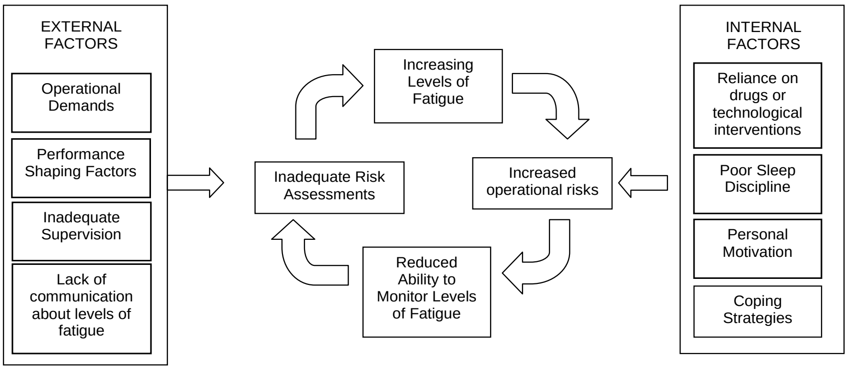

EXTERNAL FACTORS
Operational Demands
Performance Shaping Factors
Inadequate Supervision
Lack of communication about levels of fatigue
Increasing Levels of Fatigue
Inadequate Risk Assessments
Increased operational risks
Reduced Ability to Monitor Levels of Fatigue
INTERNAL FACTORS
Reliance on drugs or technological interventions
Poor Sleep Discipline
Personal Motivation
Coping Strategies

Fig. 12. Fatigue Monitoring Cycle

Figure 11 sketches some of the interactions between both external and internal factors that contribute to fatigue related accidents. Operational demands combine with performance shaping factors, such as heat, noise, time stress, as well as inadequate supervision and communications problems, for example about the risk of fatigue in particular units. Internal factors include an individual's reliance on technology or drugs to combat the hazards associated with fatigue. Poor sleep discipline and a strong motivation to 'get the job done' combine with the coping strategies identified in Chapter Four. All of these external and internal influences are likely both to increase operational risk and undermine our ability to recognize those risks.

Previous studies tend to have focused on the impact that fatigue has upon individual performance. In contrast, this Chapter has used a number of recent accidents and incidents to identify the systemic causes and effects of fatigue on military systems. It has been argued that the true nature of this problem can only be understood if the results from lab based studies are seen in the context of operational risks. In particular, we have shown at a strategic and tactical level that inadequate risk assessments and a lack of 'joined up' planning often leave soldiers in situations where they are likely to make the errors of commission and omission that are associated with extreme fatigue. At an operational level, we have argued that fatigue has an insidious effect on the interaction between and within teams of soldiers. Not only does it impair performance on shared tasks but it can also prevent teams from identifying the worst effects of fatigue in their colleagues. Unless greater attention is paid to these more complex, systemic aspects of fatigue then there seems little prospect that we will be able to mitigate the risks of future incidents and accidents in military operations.

5.6 References for Chapter Five

[1] P.A. Hancock and P.A. Desmond, Stress, Workload and Fatigue, Lawrence Erlbaum, New Jersey. USA, 2001.

[2] C.W. Johnson, The Role of Night Vision Equipment in Military Incidents and Accidents. In C.W. Johnson and P. Palanque (eds.), Human Error, Safety and Systems Development, Kluwer Academic Press, Boston, USA, 1-16, 2004.

[3] E. Hollnagel, The ETTO Principle: Efficiency-Thoroughness Trade-Off, Why Things That Go Right Sometimes Go Wrong. Ashgate, Aldershot, UK, 2009.

[4] US Army Safety Center, Fatigue, Countermeasure, (23)3:4-5, March 2002.

[5] E. Hollnagel, D. Woods and N. Leveson, Resilience Engineering: Concepts and Precepts. Aldershot, UK: Ashgate, 2006.

[6] J. Reason, Human Error. Cambridge: Cambridge University Press, 1990.

[7] J. Reason, The Human Contribution: Unsafe Acts, Accidents and Heroic Recoveries. Aldershot, UK: Ashgate. 2008.

[8] B. Kirwan, A Guide to Practical Human Reliability Assessment. London, UK: Taylor and Francis, 1994.

[9] US Department of Defense. (1995). MIL-HDBK-217: Reliability Prediction of Electronic Equipment. Washington DC, USA.: US Department of Defense.

[10] Swain, A. D., & Guttman, H. E. (1983). Handbook Of Human Reliability Analysis With Emphasis On Nuclear Power Plant Applications. Washington DC, USA.: NUREG/CR-1278.

[11] E. Hollnagel, CREAM (Cognitive Reliability and Error Analysis Method). North Holland: Elsevier, 1998.

[12] Headquarters, Department Of The Army, Leaders' Manual for Combat Stress Control, FM22-51, Washington DC, USA, September 1994.

[13] US Army Safety Center, Sustaining Performance in Combat, Flightfax, (31)5:9-11, May 2003.

[14] Canadian Defense Force, Canadian Forces Flight Safety Investigation Report 1010-12401 (Dfs 2-4-2), Category A Accident Involving CH124A Sea King, 27 Sep 2004.

[15] J. McKeon, ‘Protect Your Squash’, US Army Safety Center, Countermeasure, (26)5:3-5, March 2005.

[16] Director General Safety, Pushing the Limits Too Far, Canadian Defense Force, Ottawa, Canada, Safety Digest Volume 2, 1997.

[17] J.M. How, S.C. Foo, E. Low, T.M. Wong, A. Vijayan, M.G. Siew, R. Kanapathy. Effects of sleep deprivation on performance of Naval seamen: I. Total sleep deprivation on performance. Ann Acad Med Singapore. 23(5):669-75.

[18] A. Itoi, R. Cliveti, M. Voth, B. Danth, P. Hyde, A. Gupta, and W.C. Dement, Relationship between Awareness of Sleepiness and Ability to Predict Sleep Onset, American Automobile Association. Foundation for Traffic Safety, Washington DC, February, 1993.

[19] K. Fleming-Michael, Researcher studies sleep deprivation’s effect on decisions, U.S. Army Medical Research and Materiel Command, DCmilitary.com, August 2006.

[20] National Transportation Safety Board. Aircraft accident report: uncontrolled collision with terrain, American International Airways Flight 808, Douglas DC-8-61, N814CK, U.S. Naval Air Station, Guantanamo Bay, Cuba, August 18, 1993. Washington, DC: National Transportation Safety Board, 1994; NTSB/AAR-94/04.

[21] Canadian Directorate of Flight Safety, Fatigue: What You Don’t Know Can Hurt You. Flight Comment, Chief of the Air Staff, Ottawa, Canada, 3-6, Spring 2005.

[22] Battelle, An Overview of the Scientific Literature Concerning Fatigue, Sleep, and the Circadian Cycle. Report Prepared for the Office of the Chief Scientific and Technical Advisor for Human Factors, Federal Aviation Administration, Battelle Memorial Institute, January 1998.

[23] R.P.Larsen, Decision Making by Military Students Under Severe Stress, Norwegian Military Academy , Oslo, Norway, Military Psychology, 2001, (13)2:89-98.

[24] B. Sicard, Risk Propensity Assessment In Military Special Operations, Military Medicine, Oct 2001

[25] Canadian Director General Safety, Asleep at the CO Switch, Canadian Defense Force, Ottawa, Canada, Safety Digest Volume 2, 1997.

[26] US Army Combat Readiness Center, Where Should I Sleep? Lessons Learned Ground Report, February, 2007.

[27] Headquarters, Department Of The Army, Combat Stress, FM6-22.5, Washington DC, USA, June 2000.

[28] Canadian Defense Force, Epilogue: Hercules CC130327, Kabul, Afghanistan, 29 July 2003. Flight Comment – Air Safety Publication of the Royal Canadian Air Force (1)26, 2006.

[29] US Army Safety Center, The Cockpit is No Place to Sleep, Flightfax: Army Aviation Composite Risk Management Information, (34)4:6-7, April 2006.

[30] Ken Testorff , A Rude Awakening. Ground Warrior, US Navy Safety Center, Winter 2001.

[31] US Army Safety Center, Tank Night Vision Goggle Use: What You Don't Know Could Hurt You, US Army Safety Center, Lessons Learned, Ground Systems and Investigation , 2007.

[32] N.J. Wesensten, T.J. Balkin, and G. Belenky, The Role Of Sleep In Sustaining Individual And Organizational Effectiveness, Technical Report, US Army Physical Fitness Research Institute, Carlisle Barracks, Pennsylvania, USA, August 2007.

[33] AAFES Halts Sale of Ephedra, US Army Safety Center, Countermeasure, (23)10:16 October 2002.

[34] Ephedra and Ephedrine for Weight Loss and Athletic Performance Enhancement: Clinical Efficacy and Side Effects. Summary, Evidence Report/Technology Assessment: Number 76. AHRQ Publication Number 03-E021, March 2003. Agency for Healthcare Research and Quality, Rockville, MD.

6 Technology in the Creation and Mitigation of Risk

Chapter Five focused on the interactions between fatigue, human error and risk assessment in military operations. It was argued that increasing levels of fatigue increase the likelihood of human error. Fatigue also undermines attempts to assess the risks associated with those hazards. In contrast, this chapter focuses more narrowly on the impact of 'disruptive technologies'. In particular, image enhancement and infrared imaging systems help to mitigate some of the hazards that arise during night time operations. However, these technologies are implicated in a rising number of incidents and accidents. It is difficult to determine whether these mishaps occur because crews were using night vision devices or whether more accidents would have happened if crews had not been equipped with image enhancement and infrared imaging systems.

6.1 A Brief Overview of Night Vision

In order to assess the role that night vision devices play in the mitigation and in the creation of operational risk it is first necessary to introduce the underlying technologies. There are two main classes of night vision devices. Image intensification (I^2) systems enhance the lighting that is available within the existing environment. Infrared (IR) devices, in contrast, will typically use heat emissions to identify objects that cannot otherwise be detected using available light sources [1]. These systems support a wide range of military operations that would not otherwise have been possible. However, the additional capabilities provided by night vision devices also create new risks. Night operations continue to result in significantly more accidents and incidents than their daytime counterparts [2].

6.1.1 Visual Perception and Adaptation

Military personnel rely on their visual sense during all operations. Safe flight relies upon good visual acuity for landing and to identify terrain features. Drivers of land-based vehicles rely on depth perception to judge whether or not they can cross ditches. However, color vision, depth perception, and visual acuity all vary depending on which of the three different types of vision soldiers must rely on in a particular operation.

- *Photopic vision* occurs with high levels of illumination. The cones concentrated in the center of the fovea are primarily responsible for vision in bright light. High light condition will bleach out the rod cells that support peripheral vision. However, the reliance on cones produces sharp image interpretation and color vision using photopic vision.

- *Mesopic vision*, typically occurs at dawn and dusk or under full moonlight. This relies on a combination of rods and cones. Visual acuity steadily decreases with declining light. Color vision degrades as the light level decreases, and the cones become less effective. Mesopic vision is often regarded as the most dangerous if personnel do not adapt to the changing light conditions. As light levels fall, there will be a gradual loss of cone sensitivity. Operators should be trained to rely more on peripheral vision. If personnel fail to recognize the need to change scanning techniques "from central viewing to off-center viewing, incidents may occur" [3].

- *Scotopic vision* is used under low-light level environments. These include partial moonlight and starlight. Cones become ineffective, causing poor resolution of detail. Primary color perception during scotopic vision is shades of black, gray, and white unless the light source is high enough in intensity to stimulate the cones. A central blind spot, known as the night blind spot, also occurs when cone-cell sensitivity is lost. If an object is viewed directly at night, it may not be seen. If the object is detected, it will fade away when stared at for longer than two seconds.

It is important to understand the physiological mechanisms that underpin unassisted night vision. They help to identify mitigation techniques that can be sued to reduce the risk of perceptual error during military operations. For example, the human eye can adapt to low light. Biochemical reactions increase the level of rhdopsin in the rods. This controls light sensitivity. It can take between 30-45 minutes for most people to achieve their maximum acuity under low levels of light. Brief flashes, for instance from strobe lights, have little effect on night vision. However, looking at a flare or searchlight for longer than a second will have an adverse effect on most people.

A number of other factors, such as smoking and individual differences, adversely affect night vision. Night myopia arises from the way in which the visual spectrum is dominated by blue wavelengths of light. Nearsighted individuals viewing blue-green light at night typically experience blurred vision. Even personnel with perfect vision will find that image sharpness decreases as pupil diameter increases. Similarly, "dark focus" occurs because the focusing mechanism of the eye moves toward a resting position in low light levels. Special corrective lenses can be used to address this problem for individuals who suffer from night myopia. Binocular cues stem from slight differences in the images that are presented to each of the operator's eyes. Low lighting can make it difficult for personnel to perceive any visible differences. The effect is increased when objects are viewed at a distance. Low light levels also affect a number of monocular cues for depth perception. These include geometric perspective, motion parallax, retinal image size, and aerial perspective. As we shall see, the problems of depth perception play an important role in the increased risks of military incidents and accidents.

A number of training techniques can help maximize any remaining visual resources in low levels of light. For example, the following list summarizes the Canadian Army's [4] guidelines for night observation:

1. **Aim-off with the eyes** - Never look directly at what is to be seen. For example, if the eye looks directly at a pin-point of light it will not see the outline of the tank from which the light is coming.

2. **Do Not Stare Fixedly** - The eyes tire rapidly at night so an object will disappear if it is looked at for a long time.

3. **Avoid Looking at Any Bright Lights** - Shield the eyes from parachute flares, spotlight or headlights. Dim flashlights and turret lights and blink when firing weapons.

4. **Look Briefly at Illuminated Objects** - The time spent glancing at lighted objects such as maps or illuminated dials must be kept to a minimum.

5. **Do Not Scan Quickly** - Move the eyes in a series of separate movements to give the eye a chance to pick up a target which will appear much slower than daylight.

6. **Limit Time Spent Scanning** - Continuous scanning will cause the eye to partially black out. The eyes should be rested for 10 seconds every 2 minutes.

7. **If Necessary Use Eyes Individually** - If a lit area has to be observed, then protect the night vision of one eye by keeping it shut. One eye should be shut as an automatic reaction if a bright light suddenly appears.

Personnel can compensate for the limitations imposed by low light conditions either by training to make the most of their night vision or through the provision of night vision equipment.

6.1.2 Image Intensification Systems

Image intensification systems amplify low levels of ambient light. They do not ‘turn night into day’, nor do they compensate for many of the problems that affect vision in low light environments. Most image intensification systems perform poorly in total darkness. Amplification can range up to 35,000 times the available light. Higher amplification is associated with more expensive devices. However, at increased amplification there will be greater levels of distortion. The intensified image is, typically, viewed on a phosphor screen that creates a video image, on the user’s eyepieces.

Most image intensification systems are attached to the users’ helmet. Early models included relatively heavy battery packs that restricted the users’ head movements. This problem was exacerbated by the need to move the head because many devices offer a highly restricted field of vision between 40-60 degrees. A post action review of the Canadian Army's deployment in Kosovo found that "the current issue helmet and night vision goggles are not compatible and are painful to wear" [5]. Soldiers had to remove the devices to reduce the fatigue and frustration that built up during prolonged use. As we have seen in Chapter Five, rising levels of fatigue will increase the risks of human error even when assisted by night vision technology. There is also a danger that the use of these devices will lead to a form of over-confidence that can undermine overall safety. Image intensification equipment create problems in depth perception. Colour cues and binocular information are lost with many commercial systems. All of these limitations are being addressed by technological innovation. In particular, it is now possible to buy light weight and extended field of vision systems. These tend to be expensive and can be difficult to maintain under field conditions [6].

Visual acuity from night vision devices provides a vast improvement over human night vision. However, it is far from perfect. As with direct sight, higher levels of accuity are associated with closer, slower targets. The visual accuity offered by image intensification rapidly diminishes for objects over 400 feet away. Rain, clouds, mist, dust, smoke, fog all reduce accuity. For example, 'brown out' has contributed to a number of incidents where helicopter crews rely on images that are suddenly degraded by the dust that is brought up in the wash created by their rotors [3]. A recent incident involving a Canadian military helicopter in Bosnia providesa further illustration of these environmental problems [4]. Reports of adverse weather conditions initially convinced the crew to remain in Banja Luka. They calculated that if they left immediately then they could return to their base in Velika Kladusa within their eight hour flying limit. "We strapped on our night vision goggles after refueling and decided to go for it". This arguably illustrates the over-confidence that influences the perception of risk when using night vision devices. They were seven miles from their destination when they noticed that the lights on the hills were no longer where they expected them to be. They also began to lose sight of the lights ahead of them using their night vision equipment. The cloud lowered until it engulfed the hills that surrounded them. They realized that they could not go back to Banja Luka and so were forced to follow the only open valley in sight. The presence of mines from previous conflicts meant that they could not simply set down in any available field [4]. The subsequent analysis of this incident identified the risk that crews will become unduly complacent about the support provided by night vision equipment under adverse meteorological conditions.

Image intensification systems will not work if there is no light to intensify. City lights provide useful illuminations especially if cloud cover reflects the available light back onto a scene However; these light sources also pose hazards in military operations. Strong light sources will dazzle users. Looking at the moon has the same effects as looking directly at the sun under daylight lighting conditions. This creates problems when soldiers move toward a bright moon that is low on the horizon. The brightness of the 'ambient' light source degrades the intensified image. It will also cast deep shadows that can hide ground hazards, including excavated fighting positions. This creates considerable problems for drivers trying to locate emplacements using night vision equipment [7]. In ground operations, oncoming headlights pose a major hazard because drivers must often use their goggles at times when other road users rely on their vehicle lights. These light sources can dazzle the wearer of a night vision device to the point where they will not see barriers and obstacles, including equipment or people. Aircraft intensification systems are sensitive to the anti-collision lights required by FAA regulations. These are amplified to a point at which they can distract or even dazzle the wearer of an intensification system. There is also a risk that personnel will fixate on these external light sources.

Many of the problems associated with image intensification systems stem from the operational environment. Vehicle instrument lights and cockpit displays can create “washout” or halo effects. In many road-based vehicles it is possible to turn-off instrument illumination. However, it is a complex and expensive task to alter cockpit lighting systems without compromising the daytime use of the aircraft. These problems are compounded because red lights are frequently used in speedometers and engine instruments. Night vision systems are particularly sensitive to these sources. Personnel must also be trained not to use red-lens flashlights in situations where image intensification equipment is being used.

Operational risk assessments should consider the range of hazards that can arise with image intensification systems. Unfortunately, it is unclear how to conduct such an analysis. How does a hazard analysis offset the beneficial role played by external light sources against the likelihood that end users will be dazzled by strong illumination? How can we assess the risk reduction provided by intensification technologies against the misplaced complacency that has been observed in previous mishaps? In consequence, many military operations are conducted without the participants ever explicitly considering the hazards that arise from night vision equipment. Instead, there is a reliance on coping strategies including the intervention of team members when an individual is momentarily blinded by alternate light sources or confused by inadequate depth perception. As we have seen in Chapters Four and Five, such ad hoc interventions cannot be relied upon. In consequence, night-vision related accidents remain a significant concern for many military organizations.

6.1.3 Infrared and Thermal Imaging Systems

Rather than enhance light that is visible to the human eye, thermal imaging systems detect infrared radiation that is emitted by heat sources. These devices use transducers to detect thermal emissions that can then be focussed in the same way as conventional light. The difference in temperature amongst the objects in a scene is translated into a visual contrast represented by different shades on a display. Infrared systems can, therefore, be used in total darkness. They tend to be robust against the light ‘pollution’ that will dazzle users of image intensification systems. Infrared devices can also be used to 'see through' some types of fog because they do not rely on visible light.

The sensitivity of thermal imaging systems is measured in terms of degrees Celsius per optical *f*-number. In other words, it provides an indication of the temperature change that would be required to provoke a change in the image. These differences are typically in the region of 0.05-0.2 degrees Celsius. The resolution or sharpness is measured in terms of the instantaneous field of view (IFOV) in milliradians (mrad). 17.5 milliradians is equal to an angle of 1 degree in the instantaneous field of view. The lower the IFOV value is then the sharper the image and the longer the range will be. However, as the magnification of the thermal sensor increases, the field of view decreases. Operators must use scanning techniques to compensate for this limitation. Without well developed methods, it can be easy for users to overlook areas in a scene. As with image intensification systems, individuals can quickly become fatigued by the prolonged scanning that is required to use infrared systems in combat conditions.

Infrared systems create new hazards under particular environmental conditions. A wet runway may be cooled to such an extent that it appears to be further away than it actually is. High-humidity reduces thermal contrast and so will adversely affect image quality. Infrared systems cannot be used to identify precise details on remote objects, such as facial features, that are not distinguishable by different heat profiles.

Some of the hazards associated with thermal imaging systems can be mitigated through the use of infrared landing and searchlights**.** These tend to be most effective at low levels of illumination. If there are external lights then pilots tend to limit their scan to within the area directly covered by the searchlight. They have to be trained to expand their search on either side of the beam. Brownout can also occur when there are reflections from an infrared searchlight caused by the dust that is raised in a rotor wash. The heat emitted by infrared searchlights can help enemy personnel who may themselves be using night vision equipment. As with image intensification systems, individuals can quickly become fatigued through prolonged use of these devices. A recent Lessons Learned review was conducted into the initial deployment of light armored vehicles. One of four main findings was that "Long periods of using thermal optics can lead to crew fatigue…this can be overcome by having the dismounts trained on the functions of the turret" [8].

6.2 Statistical Studies of NVD Mishaps

The US Army Safety Centre conducted a number of pioneering studies into the accident rate for various forms of night operation involving rotary winged aircraft. The results are summarized in [1]. They found a lower accident rate for flights involving direct 'unaided' visual observations than for flights with night vision equipment. Such a counter-intuitive finding can be explained in a number of ways. It might be that the use of night vision equipment reduces the hazards of spatial disorientation but impairs overall situation awareness. The use of these devices can distract the crew's attention from other information systems and hence increases the likelihood of an adverse event. Equally, it might be argued that image intensification and infrared devices tend to be used under adverse meteorological and environmental conditions when accidents are more likely to occur anyway. These initial studies were conducted during the 1990s [9]. There is an urgent need for more recent analysis to determine whether these insights remain true given the changing operational demands on NATO forces.

A subsequent study of US Army's Black Hawk helicopter fleet surveyed the causes of more than 20 fatal accidents over a 27 year period. Approximately half of these occurred while pilots were wearing night vision devices [10]. However, the fact that an accident occurred when the crew were using this equipment does not imply that the incident was caused by these devices. It can be very difficult to assess the role that particular technologies play in an adverse event. In consequence, it can be difficult to determine whether or not to revise any risk assessments that are based on the operation of these systems. This is especially problematic when crewmembers may have suffered psychological or physiological trauma. They may be unable or unwilling to discuss the details of their actions in the aftermath of an accident or near-miss incident. Further problems arise because these statistical studies do not consider those accidents under direct visual conditions that could have been avoided if the crew had been provided with night vision equipment.

Some attempts have been made to conduct a more detailed analysis of the accident statistics. For instance, Ruffner, Piccione and Woodward [11] identified 160 US army accidents that were related to the use of night vision devices in ground vehicles between 1986-1996. Over two-thirds were attributable to three categories of terrain and roadway hazards: drop-offs greater than three feet (34%), ditches of three feet or less (23%) and rear collisions with another vehicle (11%). 34% involved the High Mobility Multipurpose Wheeled Vehicle (HMMWV), 18% involved the M1 Abrams Tank and 14% involved the M2/M3 Bradley Fighting Vehicle. The most commonly occurring environmental conditions that included dust (24%), blooming from light source (9%) and smoke (8%).

Braithwaite, Douglass, Durnford and Lucas [12] conducted a similar study of aviation accidents that focused on spatial disorientation caused by the use of night vision devices in helicopter operations. They argued that the various hazards of night vision devices, including the issues of depth perception and orientation mentioned in previous pages, predispose aircrew to spatial disorientation. They found that approximately 43% of all spatial disorientation mishaps occurred during flights that used night vision equipment. Only 13% of accidents that did not involve spatial disorientation involved these devices. An examination of the spatial disorientation accident rates per 100,000 flying hours revealed a significant difference between the rate for day flying and the rate for flight using night vision devices. They concluded that the use of night vision devices increased the risk of a spatial disorientation accident by almost five times.

6.3 Technology Mitigating the Risk of Mishaps

Night vision devices may have prevented many accidents. Such counterfactual arguments can be illustrated by the loss of a US Marine KC-130. The aircraft crashed into a Pakistan hillside near Shamsi airfield. There were no approach lights or navigational aids. The KC-130 was not equipped with any night vision equipment. Helicopter operations and noise restrictions prevented the crew from using their preferred approach. However, other KC-130s had landed at the same airfield without problems. The crew was experienced and rested. They had all flown into the airfield before. The official report concluded that the crew had "stopped navigating with instruments" and relied on direct visual observations during their approach [13]. Several analysts, therefore, argued that night vision equipment would have helped to avoid the accident because direct visual observations had failed to identify the hazards [14]. After the crash, the Marines began to retrofit KC-130s with night-vision equipment and a GPS linked map-based navigation system. The official report insisted that while the provision of night vision equipment would have helped the crew, it would not necessarily have prevented the accident [13].

The problems of using accident information to analyze the strengths and weaknesses of night vision technology can also be illustrated by litigation following a land-based training accident heard before the Maryland Court of Appeals [15]. A US Army Major was run over by a truck driven by two Maryland Army National Guardsmen during a training exercise. The Major belonged to an active duty unit that was evaluating the exercise. The accident occurred just after midnight, when the two guards drove their truck along a dirt road to pick up a patrol. The Major had remained seated in the roadway after he had finished evaluating another exercise. He made no apparent effort to move as the truck approached. The vehicle was driving under “blackout conditions” without headlights. Although one of the drivers had a set of night vision goggles, he was not using them. The soldiers had not received any training in their use. The Major suffered serious injuries that were exacerbated by a series of delays in his evacuation. He was transported to the wrong hospital and was eventually declared dead on arrival at the intended destination.

The National Guard determined that the Major’s death was caused by his lack of situation awareness during night vehicle maneuvers. They argued that if the Major had been alert, he would have heard the truck. The accident was also blamed on resource limitations that prevented the National Guard from training troops to use night vision equipment. In contrast, the Army rejected lack of funding and training as reasons for the drivers not using their night vision goggles. The accident was caused more by the driver’s excess speed than the Major’s inattention. The Major’s widow sued the State and the Maryland National Guard for maintaining insufficient supplies of night vision goggles and for failing to provide training to the drivers in the use of this equipment. Maryland's Court of Appeals unanimously upheld a Montgomery County Circuit Court decision to reject the $6 million lawsuit.

This ruling illustrates the difficulty of using previous accidents to justify the introduction of night vision equipment. The decision hinged on whether the court had jurisdiction over National Guard operational matters. This included the provision of particular items of equipment. To establish negligence it was argued that a jury would have to decide how many night vision goggles should have been acquired. The jury might also have to consider how such vision equipment should have been allocated, what kind of training should have been provided and when it should have been offered etc. The judges felt that this was beyond the competency of the court. This case provides a further illustration of the complexity of military risk assessments. We cannot expect the Civil legal system to provide detailed guidance on risk mitigation through improved training and use of night vision technology.

6.4 Technology Exacerbating the Risk of Mishaps

In contrast to those incidents and accidents that might have been prevented by night vision equipment, many mishaps directly stem from the provision of these devices. For example, night vision currency requirements in the US Army's Aircrew Training Manual state that aviators must fly at least one hour using night vision equipment every 45 days. A recent incident demonstrated that the minimum requirement is insufficient for many missions. A UH-60L instructor pilot had over 8,000 hours of rotary-wing experience. All the crewmembers had flown together many times in the past. Both pilots were qualified and current for the night vision goggle training mission. However, they both averaged less than 3 hours of night vision flight per month over the preceding 7 months. A US Army account of the incident argued, "If any one of the conditions — low recent experience, dust, winds, or low illumination — had not been present, perhaps the accident would not have occurred. If the aircrew had more recent experience, they would have been better able to deal with the harsh environment. If the illumination had been better, their low recent experience might not have been a factor. If the conditions had not been as dusty, perhaps the crew would not have become disoriented" [16]. This illustrates how a number of adverse factors can combine to create the conditions in which an incident occurs. In other words, the use of night vision equipment plays a necessary but insufficient role in the accident. Sufficient conditions often exist when personnel rely on these devices in extremely hazardous environmental or meteorological conditions.

Night vision devices mitigate some risks but create others. A further example is provided by an adverse event involving an officer with a motorized rifle platoon [7]. His unit was to occupy a battle position during a training exercise using an M551A1 Sheridan light tank. The officer's platoon was to move from their hiding positions to occupy prepared fighting positions. His orders included information about the safety requirements associated with zero illumination operations. The officer also had access to a compass, a map and a GPS receiver to assist with nighttime navigation. Although the officer was relatively unfamiliar with the area, the gunner had several years of experience on this range. Even so, they spent a number of hours driving around looking for their battle position. Standard operating procedures stated that the gunner should have dismounted to guide the driver when traveling cross-country in zero illumination. Instead, the officer used night vision goggles while his driver used a night sight. When they failed to find their fighting position, the officer was told to wait until first light before continuing the search. He carried on looking until the vehicle eventually overturned in the excavation. The officer was standing in the nametag defilade position and received fatal crush injuries. The Army Safety Centre argued that the crew relied too much on their night vision equipment as they searched for their battle positions. Soldiers must gain "an understanding and appreciation of the risk-management process and know that if the risks outweigh the benefits, then the mission should be a no-go" [7].

6.5 Night-Vision Accidents and Training

The operating characteristics of existing night vision systems make it important that individuals and teams are trained in the operational use of these applications. It can be difficult to master the scanning skills that are required to avoid the 'washout' and 'halo' effects that occur when image intensification systems are affected by secondary light sources. Similarly, personnel must be trained to overcome the limited field of view provided by most infrared applications. US Army driver training requirements cover the use of night vision equipment in AR 600-55. This is supported by training circulars such as TC 21-305-2 Training Program for Night Vision Goggle Driving Operations and FM 21-305. Support is provided through a range of courses designed for specific vehicles as well as more general training, including TC 1-204 Night Flight Technique and Procedures. US Army Training Circular 1-210 'Aircrew Training Program Commander's Guide to Individual and Crew Standardization' summarizes training and familiarization requirements for the use of night vision equipment. Prior to their first training flight with night vision, aviators must spend more than an hour in the cockpit of a static simulator or aircraft to familiarize themselves with a list of basic tasks including emergency procedures, night vision failure and a 'blind' cockpit drill. They must then undergo ten hours further training including: An Introduction to Night Vision Devices; Night terrain interpretation; Night Vision ground and air safety; Night tactical operations, including the impact of lighting; Night Vision navigation, including map preparation; Aircraft modification requirements for night vision flight; Vision, depth perception, and night vision orientation. TC 1-210 also includes requirements for aircrews to conduct refresher training in the use of night vision devices. One hour of refresher training is required if a night vision flight has not been completed on a particular aircraft type within the previous 180 consecutive days. There is also a requirement for aviators to conduct mission training. This involves at least ten more hours of flight using night vision devices followed by a further evaluation.

While it is possible to train personnel during particular flight conditions, it can be far more difficult to prepare operators to resist the broad range of visual illusions that complicate the operation of night vision technologies. For instance, many devices can provide an impression of a false horizon on the boundary between light and dark colored areas of sand, especially when other environmental factors, including dust and haze, obscure the true horizon. Desert conditions often also lack the visual markers and reference points that support accurate height perception. Under such circumstances, ground lights can be mistaken for the lights of other aircraft or even stars. Lack of features and relatively slow speeds may also persuade pilots that they have stopped moving even though the aircraft is actually travelling forward. In flat terrain, such as that found in dry lakebeds, infrared devices create the illusion that terrain slopes upwards at the edges. Particular problems are created when using the infrared searchlights to view other helicopters that may appear to be landing into a crater when they are landing on level ground.

Most military training materials have been informed by the hazards identified in previous adverse events. For example, a series of accidents led to a reminder being issued across the US Army that bright light from vehicle headlights and other sources will drive the goggles' gain down to the point that everything else in the field-of-view all but disappears. In addition, if the bright light exposure continues for 70 seconds (+30 seconds), some military systems will automatically turn off to protect both the equipment and the user. Similarly, officers were reminded that the natural illumination provided by the moon is often critical for image intensification systems and so missions should be planned to take into account the 15 degrees per hour change in the height of the moon as it waxes and wanes [17].

The US Army also operates systems for learning lessons about the use of night vision equipment within particular operational contexts. Insights gained about operational risks from Desert Shield and Desert Storm together with rotations in Kuwait helped to develop training materials that were put to use in more recent conflicts [18]. Desert operations in Iraq again illustrated the importance of integrating information obtained from night vision equipment with accurate data from GPS applications. In particular, operational experience reinforced the need for personnel to be trained to keep the lenses clean and the goggles stored safely when not in use. Sand and dust accounted for a higher than expected attrition rate for most units with access to these devices. Chapter Seven will return to this issue when considering the impact that extreme environments have upon the risks of military operations.

Many US aircrews were accustomed to dry lakebeds and scrub in their National Training Centre. They were less prepared for the impact of shifting sand dunes and extreme temperatures on night vision equipment. For instance, “the authorized airspeed for nap of the earth flight is 40 knots, but an aircraft flying in zero illumination at 25 feet in sand dunes should fly just ahead of effective transitional lift…Just keep in mind that at airspeeds below ETL, you may encounter rotor induced blowing sand” [18]. Operation experience also identified a number of visual illusions with night vision equipment. These devices can provide an impression of a false horizon when light-colored areas of sand surround dark areas, especially when other environmental factors, including dust and haze, also obscure the horizon. Desert conditions often also lack the visual markers and reference points that support accurate height perception. Under such circumstances, ground lights can often be mistaken for the lights of other aircraft or even stars. Lack of features and relatively slow speeds can also persuade pilots that they have stopped moving even though the aircraft is actually moving forward. These illusions can be so persuasive that individuals will still fall prey to them even though they have been trained to recognize that they can occur. Greater attention has recently been paid to team and crew coordination as a potential barrier to incidents and accidents. There is a need to synchronize crew observations and communications in order to combat some of the problems created by these illusions. Guidance can help to assign scanning responsibilities for pilots and non-rated crewmembers in different types of flight.

Chapters Seven and Eight will also demonstrate that the provision of training does not always match up to the standards that are claimed in many official publications. For instance, one of the lessons learned during the Canadian deployment in Bosnia was that more ground forces need to be trained in a wider range of this equipment. One of the participants in this deployment observed that "personnel were unable to train on the variety of Night Vision Devices that were eventually made available to us in theatre… not having this equipment available prior to deployment meant that we had to utilize valuable time to train personnel on equipment that they should have been familiar with before they arrived". Some of the equipment that they were expected to use only arrived six weeks after their deployment. However, the units were able to overcome these limitations. The Post Action review found that this equipment helped dismounted patrols in the towns and villages. The technology provided local inhabitants with a "dramatic" example of their fighting capability. This was claimed to have deterred crime and established credibility [5].

6.6 Risk Management of Disruptive Technologies

This chapter has looked beyond the advertising and hype that surrounds night vision technology. Image intensification and thermal imaging plays a significant role in many military accidents and incidents. For instance, Macuda et al. [19] have investigated the difficulties that aviators experience when using night vision systems to identify forms that are recognised by their motion. The studies of Macuda and their colleagues have shown that the relatively low image quality of many night vision systems can impair aviator performance and increase workload. Existing applications provide relatively grainy images that can prevent users from identifying depth, motion, resolution, form, size and distance information. Conversely, it has also been argued that the availability of night vision equipment would have prevented other accidents from occurring. A key conclusion is that the successful introduction of these systems depends upon a range of supporting factors. These include complementary technologies, such as GPS systems.

Risk reduction also depends upon appropriate training. This should help familiarize users with devices so that teams learn to overcome the limitations of existing technology in a range of environmental conditions. Ruffner, Piccione and

Woodward have shown that existing night vision training helps drivers to identify ditches and other road conditions [11]. It does not, however, help them to identify those depressions and other hazards that they have shown to be the cause of most night vision accidents. The accidents and incidents identified in this paper have supported many of the criticisms put forward by Ruffner et al. Several of the coalition partners in the Gulf were forced to use accelerated procurement to ensure that sufficient devices were made available to troops prior to the conflict. The UK Ministry of Defense issued an Urgent Operations Requirement action [20]. Subsequent chapters in this book will describe how this successful acquisition shortly before the conflict led to accelerated training procedures which, in turn, led to the accidents and incidents predicted by Ruffner and his colleagues.

Greater emphasis should also be placed on risk management before night vision devices are deployed in military operations. Previous chapters have characterized risk management as the process of identifying and controlling hazards. This process has been absent from the introduction of night vision technologies into most armed forces. In consequence, incident and accident reports have gradually been used to refine operational practices. This technology has encouraged troop to conduct operations that would not otherwise have been attempted and, which in retrospect, ought not to have been attempted even with this additional support. Other risks stem from the limitations of infrared and image intensification equipment; these include visual illusions and the problems associated with environmental hazards.

It is difficult to survey the risk 'landscape' in which night vision increases the likelihood of some hazards and diminishes others. This distribution of hazards changes from one operational context to another. Risk transfer is a particular concern when night vision devices are used in peacekeeping operations. Chapter one described how risk transfer occurs when hazards are mooved from one group to another during military missions. The use of infrared and image intensification systems reduces the hazards to peacekeeping troops from friendly fire but can increase the risks to the civilian population. This is illustrated by an incident in which a Canadian force killed one Somali and wounded another [19]. It was a turning point in Canadian involvement in Somalia and forced significant changes in their rules of engagement. A Reconnaissance Platoon observed two civilians walking around the wire of the Canadian Engineer's compound. The detachments had overlapping arcs of observation and fire. Infrared chemical lights were used to mark their positions in a way that was visible through night vision equipment but invisible to the naked eye. The two men fled after being challenged. They were then were shot from behind. One was immediately wounded and the other was subsequently shot dead by another part of the patrol.

Night vision equipment played a small part in this incident. The soldiers' interpretation of their rules of engagement and the leadership of the Reconnaissance Platoon were identified as primary causes. However, the subsequent inquiry did examine the decision to use night vision equipment. It was argued that if the compound had been better illuminated with conventional lighting then local civilians would have been less inclined to approach the installation. Shortly after the incident, the Engineers constructed a light tower. This was perceived to have significantly reduced the problem of petty theft. However, the shootings may also have had a deterrent effect. The key issue here is that additional lighting was not initially installed because it would have interfered with the use of night vision goggles. The risk of nighttime friendly fire incidents was perceived to be of paramount importance. The shooting showed that this underestimated the risks of using night vision equipment in close proximity to the local civilian population [21].

6.7 References for Chapter Six

[1] C.W. Johnson, The Role of Night Vision Equipment in Military Incidents and Accidents. In C.W. Johnson and P. Palanque (eds.), Human Error, Safety and Systems Development, Kluwer Academic Press, Boston, USA, 1-16, 2004.

[2] J.W. Ruffner, J. D., Antonio, D.Q. Joralmon and E. Martin, Night vision goggle training technologies and situational awareness. Proc of Advanced Technology Electronic Defense System Conference / Tactical Situational Awareness Symposium, San Diego, CA. 2004.

[3] US Department of the Army, Aeromedical Training for Flight Personnel, Washington, DC, 29 September 2000, Field Manual 2-04-301 (1-301).

[4] Canadian Air Force, A Dark and Stormy Night, Flight Comment, No 2, pp 6-7, Spring, 2002.

[5] Canadian Army Centre for Lessons Learned, Night Vision in Kosovo, The Bulletin, (8)1:6-11, April 2001.

[6] US Army, Aviation Night Vision Goggle Maintenance Documentation, All U.S. Army Aircraft, Department Of The Army Technical Bulletin TB 1-1500-348-30, Headquarters, Department Of The Army, Washington, D. C., 1995.

[7] US Army Centre for Lessons Learned, An M551A1 in the Wrong Hands, Countermeasure, Volume 29, Number 2, February 2001.

[8] New Zealand Army, Lessons Learned from Initial Deployment of the Light Armored Vehicle (LAVIII), LAV Update Number 3, August 2003.

[9] S.J. Durnford, J.S. Crowley, N.R. Rosado, J.Harper and S. DeRoche, Spatial Disorientation: A Survey of U.S. Army Helicopter Accidents 1987 – 1992, USAARL Report No. 95-25, U.S. Army Aeromedical Research Laboratory Fort Rucker, Alabama, 1995.

[10] P. Hess, Army Identifies Soldiers Killed in Crash, UPI, December 2002.
http://www.upi.com/view.cfm?StoryID=20021213-124412-7962r

[11] J. W. Ruffner, D. Piccione and K. Woodward, Development Of A Night Driving Simulator Concept For Night Vision Image Intensification Device Training. In Proc of Enhanced and Synthetic Vision Conference, SPIE 11th International Symposium on Aerospace/Defense Sensing, Simulation, and Controls, Orlando, Vol 3088. PP. 190-197, 1997.

[12] M.G. Braithwaite, P.K. Douglass, S.J. Durnford and G. Lucas, The Hazard Of Spatial Disorientation During Helicopter Flight Using Night Vision Devices. Journal of Aviation and Space Environmental Medicine, (69)11:103844, 1998.

[13] W.D. Durrett, Report into the Loss of a KC-130 at Shamsi Pakestan, January 9th 2002, US Marine Corps, San Diego, 2002.

[14] S. Vogel, Marine KC-130 That Hit Mountain Had No Night Vision, Washington Post, Sunday, February 17, 2002; Page A17.

[15] Maryland Court of Appeals, The Estate of Andrew Burris, et al. v. The State of Maryland, et al. No. 130, Sept. Term, 1999. Opinion by Wilner, J.

[16] US Army Centre for Lessons Learned, NVG Currency, A Perishable Skill — Currency is Not Proficiency, Flight Fax, Vol. 31, Number 2, February 2003.

[17] US Army Centre for Lessons Learned. Fight at Night and Survive, Countermeasure Vol 24. Number 4, April 2003.

[18] US Army Centre for Lessons Learned, Night Vision Goggles Desert Operations Lessons Learned - 13 Years in the Making, Flight Fax, Vol. 31, Number 4, April 2003.

[19] T. Macuda, R. Allison, P. Thomas, G. Craig and S. Jennings, Detection Of Motion-Defined Form Under Simulated Night Vision Conditions. Society of Photo-Optical Instrumentation Engineers (SPIE) Proceedings, 5442-36, 2004.

[20] UK Ministry of Defence, Operations in Iraq: Lessons for the Future, London, December 2003.

[21] Canadian Dept of National Defence, The Somalia Inquiry Report; Chap 5 March 5th Incident, 1997. http://www.forces.gc.ca/site/Reports/somalia/vol5/V5C38B_e.asp

7 Environmental Hazards and Risk Management

Chapter Six has argued that disruptive technologies mitigate some risks but at the same time introduce other hazards. In particular, night vision devices improve visual acuity under low levels of light. They also create problems of depth perception. Their introduction can also trigger risk transfer; NVDs reduce some hazardsfor military personnel but create new risks for civilians when they are challenged by unseen patrols in peacekeeping operations.

The following sections identify further interactions between military technologies and the environment in which they are deployed. In particular, we consider the ways in which the limitations of night vision technology are exacerbated during the brownout conditions that occur when visibility is reduced by airborne particles from helicopter downwash.

7.1 Causes of Brown-Out Accidents

Strategic and tactical requirements have forced many military organizations to operate from unprepared landing zones in arid and dusty conditions. In particular, there has been an increase in the deployment of Improvised Explosive Devices (IEDs) against land-based convoys and patrols. Rotor winged aircraft help to mitigate this threat. However, the increased use of helicopters to mitigate the IED threat has increased a number of other hazards. For example, the opportunities for pre-deployment training have been reduced. There have also been rising levels of fatigue as crews work to satisfy an increasing need for air support. Chapter Five has described the problems that this creates; fatigue impairs accurate risk assessments and increases the likelihood of human error. In such circumstances, aircrews may fail to appreciate the risks posed by brown out conditions. They may also be less prepared to cope with those conditions when particles are thrown up into the air.

7.1.1 Spatial Disorientation and Crew Resource Management

Brown out incidents increase the operational demands upon crews. High-levels of fatigue, environmental stressors including noise and heat as well as the spatial disorientation of low visibility landings combine to increase the risks of many military operations. Brown outs also expose problems in communication and decision making in the cockpit. Crew resource management (CRM) training can be used to counter these adverse effects [1]. These techniques are intended to help teams improve interactions both with their colleagues and the wider systems that they operate. Effective CRM helps to improve group coordination, for example by reducing ambiguity in communication. It can also reduce workload by encouraging personnel to share tasks during high-risk situations, for instance by working together to monitor the orientation of an aircraft during a brown-out. As mentioned before, however. increased operational demands erode the opportunities for CRM training and teams are often formed days or even hours before a mission takes place.

Most previous attention has focused on the primary effects of brown-out conditions for the crews that operating rotary winged aircraft. However, it is important not to overlook the secondary effects that these incidents have upon the airframes and the staff who must maintain them. Brown-outs expose flaws in the design and construction of airframes. Mechanical failures are triggered by the ingestion of sand. Brown-outs accelerate wear on rotor blades and gear as well as engine components and air filters. Secondary effects include the reduction of maintenance intervals and the consequent increase in the demands on support crews. High levels of maintenance workload continue to be a significant cause of other military accidents [2].

7.1.2 Operational Demands and Environmental Features

Brown-out accidents were relatively rare during the Cold War; given the small number of operations in arid desert regions. However, the importance of these mishaps has steadily increased. The decision to focus on the risks created by the interaction between night vision and 'brown out' incidents is further justified by the operational demands facing the NATO International Security Assistance Force (ISAF) in Afghanistan and coalition troops in Iraq. The operational context for these conflicts has created a requirement for formation flying to deliver troops and supplies into the field. The first aircraft to land or take–off in a formation stands a greater chance of avoiding the debris that affects their colleagues. However, in some areas even a single take-off can generate a dust cloud that extends for miles.

Brown-out accidents tend to be more 'survivable' than other aviation incidents. They, typically, occur close to the ground and at low airspeed. The UK MoD lost 16 helicopters in brown-out incidents between 2000 and 2007. Between 2002 and 2005, the US Army suffered 41 brown-out accidents. Approximately, 80 percent were during landings and 20 percent during takeoffs. The percentage of these accidents as a proportion of all Class A mishaps rose from 9% prior to the invasion of Iraq to 18% during it [3]. Since 1991, the US Army has reported more than 230 cases of aircraft damage and/or injury due to unsuccessful take-offs or landings in brown-out conditions.

Night vision technology can be used to overcome some of the spatial disorientation that results from brown-out incidents. Infrared systems are particularly useful in helping crews see through the particles that are thrown up by the downwash. Image intensification systems work less well, as these cannot penetrate the debris. However, operational demands combine with environmental conditions to exacerbate the hazards of using complex technologies, including NVDs [4]. Chapter Six has identified the problems of fatigue, of limited depth perception and of a restricted field of view that affects the operation of these systems. Night vision equipment has been associated with several different forms of spatial disorientation. These effects are exacerbated during a 'brown-out' landing or take-off when visibility is reduced by airborne particles. These particles are, typically, raised from helicopter downwash in the last 20 to 30 feet of an approach. The use of infrared imaging also can lead to overconfidence that further undermines military risk assessment.

7.1.3 Platform Specific Features

Fixed wing vehicles also suffer from the problems of 'brown out', especially during sandstorms. However, the frequency of these incidents is much lower and the consequences are typically less serious than for helicopter operations. Novel aircraft, such as the HH-60G PaveHawk and the MH-53 PaveLow, have been specifically developed to support extreme low-level operations. However, the threats created by 'brown-out' conditions are beginning to constrain the 'all-terrain' landing capability that these platforms provide. This is particularly important because the amount of debris generated in brown-out incidents is also determined by the downwash of the airframe or airframes involved in an approach. For instance, the performance characteristics of the V-22 Osprey make it particularly susceptible to these incidents. This aircraft relies on tilting rotors that increase the velocity of the downwash compared to other rotary winged aircraft such as the CH-46, which it was intended to replace. However, the precise relationship between rotor aerodynamics and downwash incidents is far from simple. For instance, the Osprey seems to be less prone to low altitude brown-out. The debris clears in the last few feet before the tilt-rotor makes a landing. Further study is required to develop a comprehensive account of the downwash characteristics of different aircraft. The Osprey observations are largely based on accounts from aircrew transitioning between the V-22 and the CH-46. Accident data provides more quantitative insights. For example, the CH-47 made 7% of all U.S. Army helicopter flight hours from February 2003 to June 2005. However, it was involved in 30% of all brown-out mishaps, 12 out of 41 in total between FY 2002 and 2005 [5, 6].

Downwash directly influences the likelihood of brown-out mishaps. A number of other design factors influence the consequences of these incidents. For example, the AH-46D Apache has a relatively narrow stance; the pilot sits in a rear section of the cockpit while the co-pilot/gunner sits immediately in front. This tandem layout makes the aircraft more susceptible to roll-over incidents in a brown-out compared to the parallel cockpit layout and broader stance of UH-60s. The Apache also provides a Forward Looking Infra Red vision system that was integrated with image enhancement systems as part of the Arrowhead upgrades. This arguably helps the aircrews to avoid the spatial disorientation associated with brown-out incidents. UH-60s only provide image intensification technology.

7.2 Brown Out Countermeasures

A range of technical and procedural countermeasures have been deployed to reduce the risks of brown out under a range of different environmental conditions.

7.2.1 Training, Tactics and Procedures

Military organisations have developed Training, Tactics and Procedures (TTPs) to reduce the risk of brown-out mishaps. For example, the UH-60 requirements include a section on Night or NVG Considerations: "A go-around should also be initiated if visual contact with the landing area is lost. Snow, Sand and Dust Considerations: If during the approach, visual reference with the landing area or obstacles is lost, initiate a go-around or instrument takeoff (ITO) as required, immediately. Be prepared to transition to instruments. Once visual meteorological conditions are regained, continue with the go-around" [7]. Training is required because there is a position close to the ground where it may be more risky to attempt a go-around rather than complete the landing. Hence, go-arounds should be initiated well before passing below any obstacles. However, brown-outs can occur in the last few feet of a descent. In other words, the aircrew must decide whether or not to abort the landing after they have passed underneath obstacles and at a time when it can be difficult to determine whether the go-around is more risky than the landing. All of these factors are exacerbated when aircrews are under fire or operating in close support of ground troops with an urgent operational need for air support. The use of night vision equipment adds further complexity because it can foster a sense of over-confidence in crews as they approach a potential landing site. This may leave them ill-prepared for the disorientation caused by an unexpected brown-out. In such circumstances, the US Army guidance makes it clear that the greatest risks arise when crews have no contingency plan and so must continue with a landing even though they are uncertain of their precise orientation with respect to the intended LZ. Aircrews must train to continuously scan for any available outside cues and for information from their instrumentation during brown-out contingencies.

The development of appropriate TTPs is further complicated by 'spikes' in the accident rate that have been identified by the US Army. Brown-out incidents are more likely to occur in the early stages of combat deployment [7]. Aircrews must quickly learn to use unprepared field sites – for example in forward arming and refuelling laagers, combat outposts etc. Eventually these sites can be upgraded with hard-standing areas using gravel, concrete and polymer coverings that are less prone to brown-out. However, aircrews cannot assume that they will be able to land on a prepared area. There is, therefore, a continuing requirement to ensure they are proficient in the monitoring skills that are essential to maintain situation awareness during brown-out conditions. Such observations reiterate the dynamic nature of risk assessment, as personnel interact with complex and challenging environments. From the perspective of an individual crew, there is an increased likelihood of brown-out incidents during an initial deployment. The risk then diminishes as appropriate skills and coping strategies are developed. However, over time proficiency can diminish either through the corrosive effect of complacency or through reduced exposure to the environmental conditions that trigger brown-out accidents.

One way of reducing the spikes that occur during initial deployment is to develop simulators that mimic the environmental conditions in which brown outs are likely to occur. However, there are considerable technical challenges in reproducing the sudden on-set of instrument meteorological conditions (IMC), especially while using night vision equipment. “Simulation is a valuable tool to aid in training aviators in the dust landing profile, and it is getting better all the time, but it cannot replace the feel, motion and characteristics of the real thing” [7].

One of the reasons why brown-out incidents have been so prominent in recent military accidents is because there has been a mismatch between pre-deployment training and mission requirements. This mismatch increases the risks that arise from the operational environment. Early US rotations in Iraq were more accustomed to the dry lakebeds and scrub of the National Training Centre. This left aircrews unprepared for brown-outs and a host of other operational conditions. They had relatively little experience of shifting sand dunes and the impact that extreme temperatures can have upon night vision equipment. In consequence, TTPs have been continually revised using operational input from subsequent mishap investigations and from lessons learned reviews.

There are significant hazards in practicing under the environmental conditions for which aircrews are not yet fully prepared. In consequence, visors, helmet bags and ‘foggles’ have been developed to restrict the vision of aircrews during exercises [17]. These help pilots to experience some of the effects of brown-outs under controlled supervision. A great deal of attention has recently focused on the integration of Night Vision Goggle Power Interrupt Devices (NVGPID) into US military TTPs. NVGPIDs help crews simulate the loss of NVG capability during brown-outs. Instructors can use the devices to induce a failure in the night vision goggles during a critical phase of a practice landing. The intention behind the NVGPID program is to help ensure aircrews “train to continuously scan, and to train the ability to rapidly adjust from outside cues to instruments” [7]. By extension, the same technique can also be used to replicate the impact of debris during take-off. The intention is to force the pilot to make use of the instruments and symbology to complete the maneuver. There are three additional benefits. Firstly, the NVGPID device is relatively cheap and simple. Secondly, instructors do not always have to fail the night vision system during practice landings; this makes it possible to mimic some of the uncertainty that arises when crews do not know whether or not a brown-out will occur. Finally, instructors can control the level of risk that is implicit within any brown-out drill outside the constraints of a simulator. Training officers can vary the stage of an approach or landing when a failure is induced. They can also integrate the NVGPID into other operational training scenarios to mimic specific approach patterns. It is far more difficult to preprogram complex simulation software to reflect the specific demands of a deployment.

Training devices, such as the NVGPID, expose crews to contexts that approximate brown outs conditions. These tools are of little use without SOPs that enable crews to combat the hazards from ground debris. One of the most effective TTPs is to keep the dust cloud behind the pilot's door during a rolling landing. This helps to ensure that the crew have a clear view of the Landing Zone (LZ) ahead of them. Crews must identify potential obstacles during a rolling approach. They must also consider the wind speed and direction to ensure that the dust cloud remains behind the aircraft. Rolling approaches also require careful coordination and planning for formation flight given that trailing aircraft can be engulfed by the debris thrown up by their colleagues during a rolling approach

Other TTP countermeasures address the CRM issues, mentioned in previous paragraphs. There is a temptation to 'stack the deck' with additional pairs of eyes during landing – for instance by requesting input from the door gunner in another platoon. However, this can increase the risk of misunderstandings. Other forms of communication failure can compromise shared situation awareness. The use of TTP solutions is further limited by one of the fundamental paradoxes of military risk assessment [8]. In order to become proficient in the communication and planning techniques that reduce the threats created by brown-out incidents, it is necessary for crews to practice these skills. However, it can be difficult to train in brown-out conditions when Standard Operating Procedures (SOPs) are intended to limit aircrew exposure to these hazards. In the decade between the Gulf War and Operation Enduring Freedom, the U.S. Army recorded over 40 cases of brown-out accidents during training. This provides a further illustration of the balance that needs to be struck between operational and training risk identified in Chapter Two.

7.2.2 Resilience Engineering

TTPs cannot address all environmental risks because it is difficult to anticipate the range of conditions that any crew will face. For instance, rolling approaches are less effective if the wind changes during a landing or if the prevailing wind prevents such an approach in the first place. Dust can be blown back into the aircraft as it approaches the LZ. There are further limitations with rolling approaches as a means of mitigating the environmental hazards that contribute to brown-out incidents. For example, ground obstacles and wires often restrict the area available within a landing zone. Other aircraft may require additional space to make their own approach. They can also create debris ahead of the potential LZ; obscuring the view of the rest of the formation. There may not be time for a prolonged rolling approach in medical evacuations (MEDEVAC), unscheduled supply drops or rapid troop transports. Such TTPs cannot be used in situations where enemy action may target the aircraft as it moves forward through the dust cloud. These factors constrain the airspeed and rate of descent needed to maintain aircraft control under brown-out conditions [9]. More acute descent and ascent profiles have been developed to minimise the hazards of brown-out. However, these manoeuvres create their own risks by placing heavy demands on the skill and proficiency of aircrews.

An alternative approach to the use of predetermined tactics and procedures is to train crews in the concepts of 'resilience engineering' [10]. Resilience can be defined as the ability to adjust in response to changes in the operating environment in order to sustain required operations under both expected and unexpected conditions [11]. This makes it particularly well suited to the complex and dynamic environments that characterise military operations.

Resilience engineering promotes self-regulating responses and coping mechanisms. It stresses the importance of encouraging adaptations in performance that promote successful operations. In other words, we can learn more by analysing the ways in which experienced crews learn to operate successfully in brown-out conditions than by investigating the reasons why accidents occur. The advocates of this approach focus on the far greater number of successful operations than the relatively small number of situations in which mishaps occur. The key concepts behing resilience engineering can be summarised as follows:

1. There is always a degree of uncertainty in the engineering and operation of complex systems. Successful operation, therefore, depends upon individuals and organizations adjusting what they do to match current demands and resources. Because resources and time are finite, such adjustments will inevitably be approximate.

2. Some adverse events can be attributed to the breakdown or malfunctioning of components. In other cases, problems arise in spite of normal system operation as the result of unexpected combinations of performance variability and environmental conditions. Most engineers are skilled in identifying and mitigating the first sort of failure. However, relatively little attention is paid to the variability of normal operations until they result in mishaps.

3. Safety management cannot be based exclusively on hindsight, from the investigation of previous failures, nor rely on the calculation of failure probabilities. Safety management must be proactive in strengthening recognized good practices as well as responding to previous mishaps. This in turn requires the development of specific methods for identifying 'good practices'.

4. Safety cannot be isolated from the core mission objectives or vice versa. Safety is a prerequisite for successful operations and mission success is a prerequisite for safety. Creating organizational or procedural distinctions between these two areas will be counterproductive [12].

Resilience engineering has not been widely applied to military operations; Chapter Three notes that most of the focus has been on understanding incidents and accidents rather than understanding successful operations. However, a number of initiatives have exploited elements of resilience engineering. For example, the US Army extended their TTP support to ensure instructors from units that are being rotating out of a combat area are then heavily involved in training their colleagues from new rotations. This was not always the case. These experienced individuals are often best placed to identify and pass on the approaches that will work best under a range of different environment hazards, including those that lead to brown-outs.

Resilience engineering raises a number of theoretical and pragmatic concerns when applied to military operations. A key argument in this book has been that many mitigation strategies reduce some risks while increasing others and that it is almost impossible to accurately assess the changing risk landscape under the pressure of military operations. Chapter Four has used the term 'compound risks' to describe these knock-on effects. It follows that aircrews cannot easily determine whether or not any particular coping strategy will increase or diminish the overall risks that they face. For example, some units in the US Army have tried to reduce the problems of spatial disorientation during brown out incidents by removing cabin doors. This increases visibility for the aircrew. Pilots and co-pilots can make more accurate direct visual observations by maximizing their field of view so that they can spot any 'breaks' in the clouds of dust and other debris. This solution has not been adopted across all units. In particular, the US Air Force TTPs have not approved the removal of cabin doors because there remain considerable concerns about the consequent loss of protection in combat areas. In order to improve their field of vision, crews increase their vulnerability to weapons fired from the ground.

7.2.3 Ground-Based Countermeasures

In addition to specialised 'brown-out' approach profiles, US Army SOPs require that aircrews use prepared landing zones whenever possible. These are mostly confined to established bases and outposts. Prepared LZs are seldom available in forward operating areas or for deliberate air assaults. Aircrews must improvise landings on dirt roads, open dry areas, or dusty mountain peaks [13, 14]. As with the removal of cabin doors, each of these coping strategies leads to further compound risks. The hazards associated with landing on unprepared surfaces are exacerbated by the difficulty of conducting detailed landing site surveys in hostile areas or where operational demands force late changes to the location of a mission. In other words, crews often do not know whether or not they will face brown-out conditions when they are tasked with a particular mission.

A range of materials have been developed to reduce the amount of debris that can be raised during take-off and landing. The US Army have laid down polyester Mobi-Mats, or 'triscuit pads', since the late 1990s. These are temporary pads that can be unrolled to provide a stable surface for rotary wing operations. However, they are heavy and can be unwieldy in the field. In consequence, the US Marine Corps have experimented with light-weight HeliMat alternatives [14]. These do not have the load bearing characteristics of the Mobi-Mats. They also wear out under a high tempo of operations. Operational deployment has, therefore, revealed the need to carry both type of synthetic surface.

More radical attempts to improve the operational environment can create a range of unanticipated hazards. For example, 'Rhino snot' polymers provide a further alternative to pre-formed surfaces and mats. These substances bind together debris prior to any landing. In order to apply these polymers, ground forces must first scrape off as much dust as possible. The area is then soaked with water, leveled and topped with gravel. Several coats of 'Rhino snot' are then applied and left to harden. Eventually, the surface breaks up to minimize longer-term environmental effects. However, the polymers offer a different set of logistic problems to those created by HeliMats and MobiMats. In order to bind surface layers, the polymers are very adhesive. This makes them difficult to handle. If clothes are contaminated then they, typically, must be destroyed. This makes the polymers very unpopular with some of the units that have to apply them. The difficulty of cleaning the equipment used to lay down the surfaces often forces ground units to reserve a small number of vehicles for this purpose. This also means that the technology cannot easily be used in forward areas.

7.2.4 Airborne Countermeasures

The operational risks associated with brown-out landings and take-offs have motivated a search for technological countermeasures. Rotors have been redesigned to reduce the likelihood of a brown-out. The US 101 variant of the Augusta-Westland EH101 has been designed with blades that are intended to push debris away from the fuselage. Traditional designs tend to propel dust towards and around the cockpit area. However, brown-out performance is one of several competing requirements for blade design and here can be trade-offs with efficiency/power, noise etc. Aerodynamic solutions to the hazards created by brown-out remain the subject of basic research [15].

It is unlikely that aerodynamic innovations in rotor design will provide a panacea for brown-out incidents in the short term. Flight information systems provide an alternative approach. For example, some MH-53's present a cross in the middle of the head-up display at 15 knots of descent. As the pilot decelerates this cross descends towards a reference box and hence can be used to monitor vertical velocity [16]. The Brown-out Situational Awareness Upgrade (BSAU) extends this approach. Vertical speed and vector information is mapped using data from radar altimeters and the Global Positioning Systems (GPS) on aircraft including the UH-60 and CH-47. Aircrews can access BSAU information using their standard head-up displays as well as through their night vision goggles. The design teams first identified information needed by aircrews to mitigate the risks of brown-out accidents. They then traced this required information back to the available input from sensor data. These sensors had to be sufficiently accurate to ensure that the application did not increase the cognitive load on aircrews when they used the symbology during a brown-out. However, US Army studies concluded that BSAU was only an initial step; “While the system proved its value during this and many other approaches, good crew coordination, briefing of go-around procedures, and power management remained critical tasks” [17]. In other words, technical innovation must be combined with effective CRM in order to address the hazards created by adverse environmental conditions.

Flight systems, such as BSAU, help pilots to monitor their attitude and rate of descent into brown-out landings. They cannot, at present, help aircrews avoid terrain features or ground obstacles. Night vision equipment can provide pilots with additional cues. As we have seen in Chapter Six, these devices also limit the aircrews’ field of view and hence may exacerbate rather than reduce the problems of spatial awareness. The underlying technologies are also susceptible to brown-out failures. Dust particles can completely obscure the narrow field of view provided by image intensification equipment, such as that installed on most Blackhawk aircraft. Airborne debris reduces the temperature profiles that are augmented in infrared systems. Further problems arise from the interaction between night vision equipment and the hazards created by brown-out incidents. For example, the FLIR (Forward Looking Infra-Red) pod and infrared countermeasure equipment have been slung beneath the HH-60G. The location of these devices makes them particularly vulnerable; “even the most experienced pilots are not immune from breaking FLIRs or rolling an aircraft due to a brown-out approach” [18].

A number of research programmes are developing enhanced night vision systems to reduce the risks created by brown-out landings. These include ‘see and remember’ applications that take a series of FLIR images of a landing zone before they are obscured by debris from the downwash. Software then recreates a pseudo-3D image for the aircrew to refer to during a subsequent brown-out. The Photographic Landing Augmentation System for Helicopters (PhLASH) has extended this ‘see and remember’ approach to image intensification systems. PhLASH combines an electro-optical sensor and infrared strobe lights to match a photograph of the ground with a coordinate on the Earth’s surface using onboard GPS. The intention is that the photograph would be taken immediately before the brown-out and hence could be ten or twenty seconds out of date during the final stages of the descent. This could create problems if vehicles or other elements of a formation moved into the LZ. It can also be difficult to obtain an accurate image of the LZ during night operations, given the limitations of image intensification and infrared [16, 17]. The Defense Advanced Research Projects Agency (DARPA) Sandblaster programme, therefore, integrates four different technologies:

1. *A radar sensor for three-dimensional scanning.* Conventional radar plots provide two dimensional overviews of a potential LZ. Phased and millimeter wave approaches can be used to build up three dimensional

representations while the radar signals penetrate the debris that causes brown-out incidents.

2. *A database to store successive scans of a potential landing zone.* The results can also be compared to pre-stored images and maps. This helps to ensure that whenever possible the radar returns can be mapped onto a known potion of the landing zone.

3. *Synthetic vision techniques to generate a representation of the LZ for the crew based on sensor feedback and the pre-stored information in the database.* The intention is that this view will restore aircrew situation awareness that would otherwise be compromised by a dust cloud.

4. *An 'agile' flight control system.* The ambitious aim of this component is to enable the helicopter to 'land itself' under low speed approaches [16].

Much work remains to be done, for example to demonstrate the utility of this approach in desert environments where sandstorms alter the landscapes recorded in spatial databases. US Air Force work in this area has focused on Laser Detection and Ranging (LADAR). In contrast to millimeter wave radar based on radio pulses, LADAR uses light sources to scan a potential landing zone. This technology has been applied in 'near operational conditions'. However, there are further technological problems. Ideally, aircrews require high resolution images (e.g., 1280x1080 pixels). However, existing LADAR sensors have low spatial resolution (i.e., 512x512 pixels). Real-time systems also suffer from the same limited field of view, around 30 to 60 degrees, that affects night vision systems. Accuracy requirements for brown-out countermeasures can be expressed in terms of centimeters at rages of 100 to 1000 ft in real time. At present the generation of synthetic images requires additional processing that prevents these resolutions being produced in real time. These limitations are being addressed by technologies that include active gated LADAR imaging and fusion of the millimeter wave radar from other areas of the Sandblaster programme. A recent Department of Defence research call has proposed the integration of LADAR technology with image intensification and infrared night vision equipment [19].

The US Army Safety Centre has stressed that these technological initiatives will not remove the need for to train crews to combat brown-out conditions. Pilots must learn the strengths and weaknesses of advanced sensing systems, just as aircrews must gain expertise in the application of infrared and image intensification equipment. The UH-60M and the CH-47G have recently been deployed to US forces and provide technological support for brown-out landings. While they do not provide the integrated sensing systems mention above, they do provide velocity vector, acceleration cursor, instantaneous vertical speed indicators, radar altimeter, and heading information on a common 'hover page'. Pilots are not forced to piece together critical information from numerous displays scattered across the cockpit. However, the Safety Center recognizes that the wider provision of this technology will require “the development of a separate aircrew training manual (ATM) task for landing without visual reference for all airframes, not just special operations aircraft” [7]. The following sections illustrate some of the problems that arise when military organizations use TTPs as a means of combating environmental risks.

7.3 Environmental Hazards and the Loss of An RAF Puma

The US Air Force has argued that landing in desert environments is the “most dangerous aspect of flying in combat helicopters today” [18]. It is difficult to provide an accurate sense of the environmental risks without considering a detailed case study. The remainder of this chapter analyses a host of operational concerns that led to the loss of a UK Puma helicopter on operational duty in Iraq during November 2007 [21]. Contributory factors included organizational issues, such as a failure to follow Standard Operating Procedures (SOPs), and human factors problems, including the difficulty of maintaining distributed situation awareness across multiple teams.

The aircraft involved in this incident formed part of a mixed formation of two Pumas and two Lynx helicopters. During the afternoon before the accident, a plan emerged to attack a series of targets under the cover of darkness. However, intelligence updates forced the Mission Leader to re-brief the formation on a revised scenario for the attacks. During the flight, the lead Lynx became separated from the rest of the formation and radio contact was lost. However, the Mission Leader believed he had correctly identified one of the targets. During an initial approach, the second Puma struck the ground and rolled over under ‘brown-out’ conditions as debris was lifted into the air from the downwash of the rotors. The aircraft caught fire shortly after impact; two passengers were trapped in the wreckage and were found to be dead by a subsequent rescue crew. The damaged Puma was destroyed in place by coalition forces.

The incident investigation examined the qualifications for the pilot handling the Puma at the time of the crash and found “It was not possible to ascertain his Night Vision Device (NVD) category, as it was not obvious in either his Log Book or his training records. It was clear that he had flown to NVD Cat B limits but there was no reference to any Cat B conversion course having taken place. Therefore although not theoretically qualified as NVD Cat B he had proven himself competent to fly to Cat B limits and the lack of a dedicated training course, although remiss, did not play a significant part in his handling of the events leading up to the crash” [21]. The training documents for the Non-Handling Person (NHP) in Puma 2 also “indicated that he had not completed the full NVD Cat B work up but he was sufficiently trained and experienced to be expected to carry out the NHP duties as required by his aircraft commander” [21]. The crewman onboard Puma 2 had completed his Full Mission Qualification workup to a ‘high standard’ but there was no record in his training folder that the qualification had been awarded. “Neither is there a record in his training folder of his award of NVD Cat B qualification”. The Non-Handling Person on Puma 1 was ‘suitably experienced and capable’ to undertake his role on the operation. However, he too had not completed his NVD CAT B training. His night tactical formation qualification had also expired. Such findings illustrate the problems of maintaining TTP qualifications when aircrews are increasingly being used to mitigate the threat from IEDs to land based operations.

The Board of Inquiry argued that ‘in-theatre’ experience made up for the lack of NVD Cat B training. This argument is supported by the observation in earlier sections of this chapter that there is an increased frequency of brown-out incidents during the initial stages of any deployment [7]. Aircrews seem to be less likely to be involved in brown-out accidents the longer that they have been deployed in environments where they encounter these conditions. However, the crews involved in this accident did not have the same level of experience in these environments. The handling pilot of Puma 2 was had considerable previous experience as a Non Handling Person with a total of 1,700 flying hours and around 830 in the Puma. The Non-Handling Person (NHP) in Puma 2 had around 430 flying hours on the Puma. However, he had only recently been deployed to Iraq and had limited opportunities to familiarize himself with the rest of his crew or to practice CRM techniques. Similarly, the pilot, non-handling person and crewman on Puma 1 had only been together for six weeks at the time of the accident. These findings are particularly significant given the emphasis that many military organizations place on mutual situation awareness and inter-crew communication during brown-out conditions [22]. RAF doctrine and course descriptions covering the operation of night vision devices also stress the need to provide aircrew not just with practical experience using image intensification and infrared devices but also with a theoretical understanding of the underlying technologies. Brown-out incidents have shown that past experience in a combat area may be insufficient to prepare crews for the particular demands that are created when their approach options are tightly constrained, for example, by enemy fire on an unprepared landing zone.

The crews of Lynx 1 and 2 lacked experience of working with Pumas. There had been no pre-deployment training between Lynx and Puma aircraft nor was there any in-theatre mixed-type workup package. The Puma force argued that the risks of in-theatre training were too great and that the operational tempo left little time for such exercises, such arguments echo the observations about training risks introduced in Chapter Two. In retrospect, additional training in the operating environment with mixed formations might have improved crew resource management both within and between aircrews. The absence of mixed formation exercises and the relative lack of familiarity between recently formed teams undermined their ability to practice the communications that are vital to maintain mutual situation awareness [21].

The MOD has taken steps to address many of these issues; for instance by conducting closer audits over the training records of RAF pilots. They have also increased the amount of practice Joint Helicopter Command aircrews receive in the ‘desert box’ rolling landing techniques, described in previous sections, as part of Exercise Jebel

Sahara. However, this accident reveals continuing areas of concern in terms of aircrew training and preparation for the interaction between brown-out incidents and the use of night vision devices.

7.3.1 Mission Planning in Uncertain Environments

The initial mission briefing provided generic information about a range of potential hazards in the operational environment, including the weather, light levels, intelligence, air tasking orders etc. It also provided an opportunity for the crews to conduct detailed formation planning for night operations, including the development of contingency plans for brown-out landings. However, the accident took place on the same day as a change in the engineering rotation. In consequence, servicing had to take place earlier than might otherwise have been expected, at the start of the aircrews' duty period. Some crewmembers had to support the engineering teams at the same time as others were taking part in mission briefs. These absences together with the uncertainty over the NVD status of crewmembers may be symptomatic of an ad hoc approach, which although it may be understandable given the operational tempo, reveals underlying concerns in the planning and staffing of missions. It also reiterates previous remarks in Chapter One of this book. Risk framing refers to the boundaries of any hazard analysis; it was not clear that the servicing requirements would have an impact on the conduct of the operation. Only in retrospect could senior officers consider the additional risks created by the duty rotation.

The final mission planning was conducted by the Non Handling Person of Puma 1. He was preoccupied in planning and so did not attend some of the mission briefings and arguably could have been better supported by members of the other crews. He could not call upon this assistance because several of his colleagues were still helping to service their aircraft. The investigators concluded that although the plans were well made; 'there was much confusion as to the exact nature of the target sets and the number of landing sites that were to be visited, suggesting that there was a great deal of confusion amongst all parties' [21].

A new mission target was identified while the crews were moving to their aircraft. This urgent operational requirement seems to have obscured the fact that the crews had not received an adequate briefing. The focus was on maximizing the opportunities provided by new intelligence rather than on ensuring that all parties understood the details of their tasking. In consequence, the Mission Leader briefed the rest of the formation over the radio. The new target required a far more demanding sortie profile than the mission that had previously been planned and briefed. The aircrews may have under-estimated the risks associated with 'in flight' briefings without detailed contingency plans, even given the need to respond to a time-limited target opportunity. These observations again illustrate the complexity of military risk assessment; time pressures and limited information compound the environmental hazards that have been described in previous paragraphs.

7.3.2 Closing on the Target

It was dusk when the formation departed their home landing site but light levels were high. A number of obstructions were spotted during the flight. These included wires that forced them to fly higher than the crews would have preferred. The Non-Handling Person on Puma 2 later acknowledged that they had experienced increasing levels of workload. Chatter on the Air Traffic and tactical radios interfered with his task of updating successive grid references generated as the target moved position. The formation closed in, a couple of miles before the last known target reference.

With around one mile to go before Puma 2 reached the target, it became clear that Lynx 1 had overshot to the South by around a mile due to an error in their navigational equipment. The over-flight alerted elements of the target forces to the potential attack. In the meantime, Puma 1 and Lynx 2 failed to establish radio contact with the missing crew. The remaining formation could now see that the correct target indication was now some 3 miles behind them to the North. The Mission Leader requested infrared ground illumination on a known location to help navigate back to the proposed landing area. He then instructed Puma 2 and Lynx 2 to join Puma 1 on a direct route to the target. This left Lynx 1 detached from the formation. The Team Leader of ground forces was also onboard the Mission Leader's aircraft, Puma 1. Together they conducted a rapid briefing on a revised approach to the target. However, the aircraft were now deployed in an unfamiliar formation without a full briefing and only the most rudimentary of contingency plans. Some of the aircrews were newly formed and, as mentioned, all lacked training in mixed formation operations.

The crew of the remaining Lynx 2 struggled to identify the target using their night vision capability during the final stages of their approach. They, therefore, decided to conduct an early overshoot, as described in the TTP section of this chapter. Meanwhile, the Mission Leader had not registered the latest intelligence updates on the location of the target and so urgently sought further clarification. He assumed that the target was now located to the South. He, therefore, altered course at a height that was below the level set for their Radar Altimeter warnings. This had not been reset after the transit phase of the mission. It, therefore, continued to generate spurious warnings. If the Altimeter warning threshold had been set at a lower height during the approach to the target then the crew might have paid more attention when the alarms were generated.

Lynx 2 now rejoined the other two Pumas having recovered from the overshoot. He remained at a safe distance to assess what they were doing. The Handling Pilot of Puma 2 was also unsure about the intentions of the Mission Leader; the target could still not be seen. The crew of Puma 2 now believed that Puma 1 was making a final approach as their speed was further reduced. Puma 1 then performed an abrupt right turn and radioed the other units that they were under fire. This was the first time that any of the units had made visual contact with the targets. Puma 1 then began approaching a field adjacent to the target area in a manner that made it clear to the crew of Puma 2 that they were about to land. The Handling Pilot of Puma 2 elected to follow the Mission Leader and come down in the same field, which appeared to be flat and stable enough to support a landing.

7.3.3 Approach to Landing

It is usual practice for Handling Pilots to announce to others in the formation that they are committed to a landing when the performance characteristics of their aircraft no longer allow for the maneuver to be aborted. However, Puma 2 made the 'committed' call during a very early stage of the approach. This made it difficult for the crew to judge the eventual problems created by the constraints on the landing zone under brown-out conditions. Their decision to make this early call was justified by their desire to support Puma 1 as it came under enemy fire.

The dust cloud raised by the down wash of Puma 1 demonstrated that ground debris would impair visibility on landing for Puma 2. However, this did not prompt the crew of Puma 2 to revise their radar altimeter settings to provide additional assurance on their descent. The late turn by Puma 1 also left Puma 2 with very limited space to land – this ruled out the rolling 'box' approach techniques that have been advocated in US and British military doctrine. As mentioned in previous sections, these rolling approaches involve a gradual descent along a preplanned forward trajectory that helps to minimize the disorientation of a rapid descent into brownout conditions. Without the space to conduct a rolling box approach, Puma 2 performed an almost vertical descent from 75 feet (23 meters). The degree of difficulty was further exacerbated by a surface wind of 5-10 knots (9-18 km/h). The handling pilot was so focused on the demands of landing the aircraft that he did not notice when one of the troops began firing on the targets from the right door of his Puma. The crewman and the non-handling pilot stated that this distracted them from their tasks.

From about 30 feet (9 meters), a significant dust cloud gathered around the descending Puma. Ground references became harder to maintain. The handling pilot stated that he was able to maintain visual references throughout the descent. However, 'they were of varying quality and mainly consisted of moving dust and straw' [21]. He did not arrest the initial descent in time and hit the ground. The resultant 'heavy landing' did not exceed the 3G limit that would have triggered the Helicopter Emergency Egress Lighting System nor did there appear to be any structural damage. The collective was not lowered and the Puma maintained around 10 degrees of pitch. Partly in consequence, the aircraft continued its forward motion. It also began a rolling oscillation that increased as the aircraft slowed. The handling pilot was concerned that the Puma would roll over. He decided to overshoot the landing without clear visual references. The handling pilot later stated that that he chose a level attitude for takeoff but did not verify this using his instrumentation. He raised the collective and felt the Puma start to climb. The low main rotor RPM audio warning sounded twice; possibly as a result of the handling pilot quickly raising the collective.

At this point, the Non-Handling Person saw a Lynx at 10 o'clock. He informed the pilot but considered that there was no chance of a collision given their relative positions. The pilot also recalls seeing the Lynx through the brown-out and became increasingly concerned that there was a danger of collision. The pilot decided to halt the climb and carry out a level transition into horizontal flight. The intention was to gain airspeed and move the aircraft away from the dust cloud. He did not check his instruments nor did he establish a visual horizon [21]. As he began this maneuver, the aircraft reentered the dust cloud and the crew lost all visual references. The Board of Inquiry argued that this disorientation prevented the crew from assessing the effects that their commands were having on the aircraft. As the pilot began to level the wings he felt an accelerated roll to the right with the noise and control motions that might be associated with the blades striking the ground. The aircraft continued moving to the right while more dust began to block out all external visual references. The crew could, however, feel the blades striking until the aircraft finally came to rest some 5 seconds after the initial impact. Both of the aircrew had their night vision devices dislodged during the landing. The emergency lighting system was activated to assist the egress from the damaged aircraft. The aircraft caught fire shortly after impact. Two passengers were trapped and later found to be dead by a rescue crew.

7.3.4 Environmental Hazards

The subsequent Board of Inquiry discounted aircraft technical failure and aircraft performance as potential causes. They also excluded enemy action; sabotage and friendly fire. The Board did, however, identify a range of environmental hazards that contributed to the accident.

Meteorology: The crew of Puma 2 experienced significant downwind during their approach. This led to a loss of lift and a higher than anticipated rate of decent, earlier than would otherwise have been expected. The initial heavy landing was, therefore, the consequence of an uncorrected increase in the rate of descent caused by this downwind component. Meteorological conditions also had a direct impact on the brown-out. As the crew approached the landing zone, they might have expected the dust cloud to form behind them given their descent profile. However, the downwind component created brown-out conditions below the aircraft at a much earlier point in the landing than might have been anticipated. The wind also blew debris ahead of the aircraft making it much harder for the crew to judge their rate of descent and attitude. Finally, the investigation argued that the downwind component exacerbated the Puma's tendency to over-rotate forward during transition and led to a nose down attitude that increased the rate of descent.

Light Levels and Night Vision Device Performance: The Board concluded that "The Op training directive states that all crews should be both NVD Cat B and Night

Tactical Formation qualified prior to Basic Mission Qualification (BMQ) training. The Handling Pilot, Non-handling Person of Puma 2 and Non-Handling Person of Puma 1 were not correctly qualified to NVD Cat B before their BMQ training. A review of qualifications is underway". However, they also argued that the performance of night vision equipment and ambient light levels were not contributory factors in this accident [21]. The sun set approximately one hour before the crash and the crews reported that ambient light levels were workable. The sun's afterglow was visible in the second Puma's 9 or 10 o'clock position but it was not mentioned as a distraction in testimonies after the accident.

It is noticeable that the Board of Inquiry did not consider the operational strengths and weaknesses of the night vision devices that were available to the crews. This was beyond their remit. In contrast, separate hearings in Coroner's courts increasingly criticize the UK MoD for failing to adequately consider the operational performance of the equipment that they provide [23, 24, 25]. Coroner's hearings give families of the injured and bereaved valuable opportunities to voice their concerns over military procurement. However, their criticisms often lack the detailed engineering and technical input that is required to develop constructive proposals and avoid future failures. There is an urgent need to develop procedures by which the findings of Board of Inquiry can be extended to maximize the lessons learned from previous accidents in a manner that is both technically convincing and which elicits the support of all stakeholders, including both surviving personnel and the relatives of any casualties. This is increasingly important when many defense suppliers only take a passing interest in the ways in which their equipment actually performs under operational conditions. Many night vision developers continue to ignore the operational 'lessons learned' from incidents such as the loss of the Puma.

The Dust Cloud: The Board of Inquiry treated the dust cloud as a distinct issue from the light levels and the performance of night vision equipment. It was argued that light levels did not contribute to the accident, even though the crew was wearing NVG's for which they did not have the full CAT B training. The approach was conducted into a 'significant' dust cloud that robbed the handling pilot of visual references; "Despite the crew's utilization of the latest UK NVD technology they ended up being close to the ground but unable to see the surface due to dust" [21]. This sentence illustrates how the Board viewed night vision technology as a means of mitigating the risks created by brown-out conditions without considering the consequent hazards of increased spatial disorientation.

The loss of the Puma stemmed in part from the disorientation of the crew. The Handling Pilot initially reported that he lost visual references at around 15 feet (4.5m) on final approach. However, he subsequently contradicted this statement. It is clear that he experienced some difficulty in judging his rate of descent and after the first impact was 'flying blind' within the debris that was raised by the rotor wash. He could not, therefore, judge the extent of the subsequent roll. This contributed to his decision to overshoot. His attention was focused on external cues rather than monitoring his instrumentation. This made it difficult for him to obtain adequate visual references so that he could judge the rate of climb. The crew was able to glimpse the Lynx but this was also in motion. Any references would be relative to the trajectory of that aircraft and could be very misleading. The crew, therefore, lacked the necessary information to identify the effects of any attempts to transition forward. Arguably, they could not determine whether they were ascending, descending or turning [21].

Terrain: The landing area was relatively flat, however, it was crossed by a rectangular grid of irrigation ditches around 2 feet deep (0.6m) and smaller furrows of around one foot in height. It was very dusty. There was a significant risk that an aircraft might strike one of these ditches. The subsequent investigation argued that “if a thorough reconnaissance of the field had been carried out, these features would have been noticed and an appropriate landing would have been chosen to avoid any run on, making oscillations (following a ground strike) unlikely” [21]. Of course, any decision to reduce the run-on would have correspondingly increased the likelihood and consequences of a brown-out by further constraining the use of the rolling box approaches that have been described in the opening sections of this Chapter. This argument again stresses the need to look in more detail at the complex interactions that arise under military operations; where a change in tactics might reduce exposure to one potential risk while at the same time increasing the likelihood of other hazards. By trying to avoid the terrain hazards through a vertical descent, the aircrews would increase the dangers of a brown out landing.

The landing area was seen by the crew very late in the approach of Puma 2. There was also pressure to land when they observed the tracer close to Puma 1. The handling pilot may, therefore, have felt very constrained in terms of the potential areas in which he could complete a landing. This led him to follow a non-standard vertical approach profile that was ‘inappropriate in dusty conditions as height judgment is very difficult and references are very difficult to maintain” [21]. In consequence, the handling pilot lost the cues necessary to arrest the descent.

7.3.5 Operational and Command Hazards

The previous section identified a host of environmental factors that contributed to this accident. These combined to expose deeper operational and command weaknesses.

Poor Supervision: Puma 2’s handling pilot had not passed an appropriate Cat B NVD training course. The Board argued that this might have reflected a potential problem in crew selection procedures. One possible consequence of this decision to assess NVD competency within crew selection was that the Board rejected the handling pilots NVD training as a contributory factor ‘in itself’. Similarly, the non-handling person’s lack of training was also considered narrowly in terms of the insights it provided into the supervision of crew composition rather than the ‘systems issues’ in terms of the interaction between terrain, meteorology, NVD operation and approach trajectory. The observation that the non-handling person's logbook indicated NVD CAT B competency when he had not completed the desert environment qualification was, therefore, dismissed as a contributory factor by the Board even though this undermined the effectiveness of many of the TTP techniques identified in earlier sections of this Chapter.

The Board argued that the Handling Pilot's concern to reduce the inexperienced Non-Handling Person's workload, by taking over the tactical radio net etc, may have contributed to the accident. The Non-Handling Person in Puma 1 was found to be 'incorrectly qualified' for the mission having an out of date 'Night Tactical Formation' qualification and not possessing the NVD Cat B qualification. Again, these omissions were not found to be contributory factors except that they showed the crews were working at, or beyond, their operational capacity; "the fact that all 4 crewmembers were working very hard meant that no one took stock of the situation and no one was balancing the risks that were taken" [21].

If the lack of NVD qualifications had been identified as a contributory factor in this accident then many crews would have been grounded until they completed the courses that would become prerequisites for subsequent missions. This would have created heavy burdens on those crews that did possess CAT B qualifications at a time of rising operational demands. We must consider whether the risks of deploying personnel without CAT B NVD training outweigh the operational benefits of tasking them to use this technology on missions that have significant tactical importance for ground forces? This question extends well beyond the Puma case study. The development of innovative technologies, including multi-sensor fusion for the visualization of brown-out approaches, increases rather than reduces the need for appropriate training. In the future, leaders will still have to determine whether to deploy troops who have not completed the TTPs that have been developed to support the use of advanced technology in complex combat environments. If they decide that these TTPs are essential requirements, then there will be an inevitable increase in workload for those personnel who have already met the necessary training requirements. If they decide to overlook TTP deficiencies then there is an increased risk associated with the use of advanced technologies by crews that lack the necessary exposure to training, tactics and procedures.

Operational Pressure: It is hard to underestimate the importance of operational pressure as a factor in the decision to task this mission to the Puma and Lynx crews. There was an urgent need to get the mission underway and this eroded the time that would otherwise have been available for mission planning. Changing intelligence also forced late revisions to the plans. There is a suspicion that had the mission been successful, leaders would have been commended for improvisation. In the circumstances, however, it is clear that a re-brief might have helped crews consider likely contingencies during the approach to the landing sites. The inquiry argued that after the loss of the lead Lynx, the formation did not know the disposition of the target and hence 'operational pressure both real and perceived was a contributory factor'.

Inadequate Authorization Processes: The authorization of missions provides a process of checks and balances that are intended to safeguard military personnel. However, the formal mission approval process must also provide leaders with sufficient flexibility to respond to changing intelligence; environmental factors; resource constraints etc. The standard format in place at the time of this accident was deliberately designed so that approval did not need to be written out in full for every sortie. Instead, pro forma authorization sheets were used. In this sortie, they were signed at such an early point that the authorizing officer could not discuss the limits or nature of the task. It was, therefore, difficult for the authorized captain to explain those critical mission constraints to the rest of the crews. Many military organizations now have an expectation that leaders will explicitly request briefings or 'resets' when they are unsure of essential mission parameters. In contrast, the authorization sheets asked the crews to complete any tasking without caveat or recourse to the chain of command. The authorization process had evolved under operational pressures to the point where "it removed the final check of understanding and confirmation of crew suitability for the task at hand" [21].

Inadequate Briefings: The failure of the authorization process to establish mission parameters and guide crew composition was compounded by operational pressures. Together these factors constrained the briefing process that is intended to act as a foundation for mission safety. The briefings described missions that were never flown; changing intelligence forced successive revisions to the plans. Even so, senior personnel were missing from the briefings in order to complete other tasks, including aircraft servicing. This removed an opportunity to provide guidance to the less experienced crew members and, theoretically, alter the deployment and composition of the teams. Quick Battle Orders (QBOs) were used to brief the crews in-flight. These may have been ambiguous – for instance over whether Puma 1 or the remaining Lynx was the mission lead. The QBOs were not passed on to the reserve Puma's 3 and 4. This is a significant omission given that the Deputy Leader was in Puma 3. The reserve Pumas also carried more experienced crews who might then have realized the complexity and risks of what was being proposed. These communication problems were exposed by the environmental factors that led to the brown out conditions at the LZ.

Formation and Deployment: The task of communicating Quick Battle Orders was complicated by radio problems within the formation. This was said to be a common occurrence – something that itself is a priority 'lesson' from this accident. After the mission it was unclear whether messages were not received, or whether they were missed by crews dealing with high workload or whether some of the crews had the volume turned down to reduce distractions. Such uncertainty again underlines the need for a more systematic review of communications within these formations.

As noted previously, Lynx 1 missed the target area and divided the formation. This created uncertainty for Puma 2's handling pilot about the position of the missing Lynx as he attempted the overshoot. It also created potential confusion amongst all elements by undermining the formation and mission brief. Crews could no longer rely on de-confliction plans between the Lynx and Pumas. The eventual deployment was based on Quick Battle Orders using an untried combination of one Lynx and two Pumas. The nature of the QBO's, the communications problems and the failure to brief all crews on intelligence updates about the location of the target added to the risks associated with this formation. The Board summarized these findings by arguing that 'there was a significant breakdown in Crew Resource Management across the formation with a low standard of leadership and 'followership' being displayed throughout" [21].

Inadequate Standard Operation Procedures (SOPs): After the mishap it was argued that the crew of the Puma had accepted a role that was not described within the existing SOPs. This contributed to the lack of clarity in mission objectives and tactics that was observed in previous paragraphs. In particular, the emerging plan did not identify an Initial Point (IP). In formation flying, these act as a rendezvous and help to ensure that aircrews approach a target along an agreed route from a known location. Initial Points also help to coordinate a series of final checks, including making adjustments to the radar altimeter warnings. These warnings are initially set en route to a target at a level that ensures they are not triggered every time the aircraft crosses raised ground. However, they are then reset for the descent into a landing zone.

The crew of Puma 2 never agreed on the IP and hence they flew a beyond the transit phase without having set the Radar Altimeter (Rad Alt) to 25ft (7.5m) for the final approach contrary to the Standard Operation Procedures (SOPs) for Puma dust operations. After the accident, it was found that the Handling Pilot directed the Rad Alt audio warning should not be reset for approaches as a matter of course. This decision was not questioned by the rest of the crew and the same policy also seems to have been adopted by others in the squadron. The subsequent board noted that 'this was not the view of the 22 Squadron training staff who believed it should be set at 25 feet for all dust approaches, without exception' [21].

This contradiction between official SOPs and everyday operations illustrates the complexity of military accidents. The decision not to reset the Rad Alt warning contributed to this accident. However, the crews' actions were also motivated by a desire to reduce intrusive and distracting warnings. There are further human factors concerns when spurious alarms significantly increase the workload on crews approaching a landing site. Local practices diverged from SOPs in a number of other ways. For example, Minimum Safe Heights were not commonly calculated for this area of operations. The accident also found examples where there were no SOPs to support crew operations. In particular, the individual SOPs for Puma and for Lynx aircraft did not describe what should be done during joint operations. This created considerable mutual uncertainty; neither knew the procedures associated with their colleague's platform.

7.4 Operational Tempo and Risk Exposure

This chapter has argued that strategic changes in military operations can expose personnel to a range of hazardous environments. The contingencies and characteristics of asymmetric warfare in Iraq and Afghanistan have increased the demands on rotary winged aircraft both for troop deployment and supply missions. Changes in insurgent technology, including the use of remotely detonated IEDs, have also increased the use of night operations in areas where there are few prepared LZs [23]. These factors have combined to expose personnel to the environmental conditions in which brown outs are likely to occur.

Increased exposure to adverse environmental conditions helps to explain an increasing number of brown out related mishaps. We have also argued that many of the incidents arise from problems in the creation and delivery of Training, Tactics and Procedures (TTPs). The US Army have observed spikes in the frequency of incidents during the initial phases of any deployment; these may be associated with the difficulty of training personnel in the environmental hazards that they will face in a combat area. Simulation and training tools can be used to address these issues but problems arise when operational pressures prevent many personnel from completing courses before their deployment.

Many of these factors contributed to the loss of a UK Puma on operational duty in Iraq. This mishap was triggered by the crews' loss of situation awareness that, in turn, was undermined by a number of environmental hazards. However, the immediate events leading to the accident stemmed from a wider range of latent issues. These included operational pressures, problems in the supervision and approval of missions, inadequate briefings and a failure to consider the risks associated with mixed formation operations.

The official Board of Inquiry into the loss of the Puma revealed a number of issues that, although they were not identified as contributory factors, form a stark contrast with the doctrine and practices in other military organizations. It was not possible for the investigation to use the existing logs and training records to determine the Night Vision Device category of the handling pilot of the aircraft involved in the crash. He had flown in operations requiring NVD Cat B conditions but there was no reference to any conversion course intended to bring him up to this level. Similarly, the Non-Handling crewmember of the Puma had not completed the full NVD Cat B training. Nor was there any record in his training folder that he had completed his Full Mission Qualification. It is difficult to argue with the Board's conclusion that the lack of NVD training was either a cause or contributory factor. They insisted that the operational performance of the crew demonstrated that they could perform to NVD Cat B levels. However, it seems clear from the initiatives in other military organizations that more could be done to train crews for the demands created by brown-out conditions. These initiatives will never be effective unless better records are kept of the training that aircrews have received. These records must be used to inform mission tasking.

The causes of many brownout incidents can be traced back to the operational tempo in Iraq and Afghanistan. Incomplete training records are symptomatic of the common pressures on UK and US forces to take on significant operational demands with finite resources. Ultimately, these pressures are a greater threat than those associated either with night vision operations or with the environmental conditions that trigger brown-out incidents.

7.5 References for Chapter Seven

[1] C.W. Johnson, Reasons For The Failure of CRM Training. In K. Abbott, J.-J. Speyer and G. Boy (eds.), HCI Aero 2000: International Conference on Human-Computer Interfaces in Aeronautics, Cepadues-Editions, Toulouse, France, 137-142, 2000.

[2] US Air Force, National Helicopter Experts Gather to Discuss Aerodynamic Solutions for Brown-out, Wright Patterson Air Force Base, USA, 2007. http://www.wpafb.af.mil/news/story.asp?id=123058808, Last accessed February 2009.

[3] G. Jennings, Down in the Dirt: Helicopter Brown-outs, Jane's Defence: Air Forces, 13th February 2008.

[4] US Army Centre for Lessons Learned, Night Vision Goggles Desert Operations Lessons Learned - 13 Years in the Making, Flight Fax, Vol. 31, Number 4, April 2003.

[5] Project on Government Oversight, Brown-out Accidents Plague CSAR-X: Controversy Surrounds Air Force Selection. June 21, 2007.

[6] US Army, Army Aircraft Mishaps/Accidents and Percentage of Chinook Accidents, cited in Aerospace Daily & Defense Report, March 7, Volume 221, Issue 44, 2007.

[7] R. Gant, Night Vision Goggle Power Interrupt Device (NVGPID): A Simple Device to Train Crucial Skills, US Army Aviation, 22-26, April/May 2007. .

[8] C.W. Johnson, The Paradoxes of Military Risk Assessment. In A.G. Boyer and N.J. Gauthier (eds.) Proceedings of the 25th International Systems Safety Conference, Baltimore, USA, International Systems Safety Society, Unionville, VA, USA, 859-869, 0-9721385-7-9, 2007.

[9] U.S. Army Combat Readiness Center, Brown-out on the Battlefield, US Army FlightFax, May 2005.

[10] E. Hollnagel, D.D. Woods and N. Leveson (eds.), Resilience Engineering: Concepts and Precepts, Ashgate, Publishing, London, UK. 2006.

[11] C.W. Johnson, A. Herd and M. Wolff, The Application of Resilience Engineering to Human Space Flight, In Proceedings of the International Association for the Advancement of Space Safety, Huntsville Alabama, NASA/ESA, 2010.

[12] J. Leonhardt, E. Hollnagel, L. Macchi, B. Kirwan, A White Paper on Resilience Engineering for ATM, EUROCONTROL, Brussels, Belgium, 2009.

[13] U.S. Army Combat Readiness Center, Brown-out on the Battlefield, FlightFax, May 2005.

[14] R. Whittle, Combat Operations: Matting Down Brown-out, Engine Wear. Rotor and Wing Magazine, Wednesday, 1st August 2007.

[15] US Air Force, National Helicopter Experts Gather to Discuss Aerodynamic Solutions for Brown-out, Wright Patterson Air Force Base, USA**,** 2007.

[16] C. Martin, In The Thick of It All: Surviving the Brown-out, Canadian Air Force, Flight Safety: Debriefing Newsletter, June 2008.

[17] U.S. Army Combat Readiness Center, BSAU: Improving Aircrew Situational Awareness During Brown-out, Flightfax, (32)10, October 2004.

[18] U.S. Air Force, J. Sherer, Brown-out. US Air Force Safety Center, , 2nd November 2008. http://www.afsc.af.mil/news/story.asp?id=123085802, Last accessed February 2009.

[19] US Department of Defense, Improve LASER RADAR (LADAR) Image and Data System Processing with Multi-Sensor Fusion in Vertical Lift Visual Degraded Environments, Navy Solicitation SBIR 2008.2 - Topic N08-147, Washington DC, USA.

[20] C.E. Rash, Flying Blind, Aviation Safety World, Flight Safety Foundation, 44-46, December 2006.

[21] Royal Air Force, Board of Inquiry into an Accident Involving Puma HC1ZA938 on 20 November 2007 while on Operation in Iraq, UK Ministry of Defence, London, UK.

[22] U.S. Army Combat Readiness Center, Night Vision Goggles Desert Operations Lessons Learned - 13 Years in the Making, Flight Fax, Vol. 31, Number 4, April 2003a.

[23] BBC. Hercules Safety ‘Still Lacking’, 2007. Available on http://news.bbc.co.uk/1/hi/uk/6406203.stm, last accessed February 2009.

[24] BBC. MoD Criticised over Soldier’s Death, 2008. Available on http://news.bbc.co.uk/1/hi/england/tyne/7318813.stm, last accessed February 2009.

[25] BBC. Mother Questions Soldier’s Armour, 2009. Available on http://news.bbc.co.uk/1/hi/england/leicestershire/7884270.stm, last accessed February 2009.

[26] C.W. Johnson and L. Nilsen-Nygaard, A Systemic Approach to Counter the Threat to Public Safety from Improvised Explosive Devices (IEDs), Proceedings of the 27th International Conference on Systems Safety, Huntsville, Alabama, USA 2009, International Systems Safety Society, Unionville, VA, USA, 2009.

8 UAVs and the Military Hazards of Political Decision Making

Environmental hazards can often be mitigated by the introduction of new technologies. For instance, the risks posed to crews by brown out conditions can be reduced by the introduction of Unmanned Airborne Vehicles (UAVs). These systems can be deployed in both surveillance and, increasingly, in offensive roles against remote targets. There are significant benefits to be gained from this technology [1]. Not only does it reduce the risk to air crews, it can also extend the duration of operations by enabling shift systems to be introduced for ground control teams. This, in turn, helps to reduce the risk to ground forces.

Politics plays a critical role in establishing the context in which military operations take place. It follows that any analysis of military risk assessment must consider the hazards that are created or mitigated by political decision making. In particular, the following pages consider the operational problems that were created by a commitment to support the rapid deployment of Unmanned Airborne Systems (UAS) into Afghanistan. Political demands forced the introduction of this technology against deadlines that reduced the amount of time available for the procurement and testing of these systems. These limitations also arguably undermined the intended operational benefits that would otherwise have been obtained from the use of these systems in complex and hostile environments [2].

8.1 UAV Deployment in Afghanistan

Operation ATHENA began in August 2003, when Canadian forces returned to support the International Security Assistance Force (ISAF) around Kabul, Afghanistan. Over five successive, six month rotations, troops conducted foot patrols and surveillance in cooperation with other ISAF units. Their aim was to provide a visible military presence, improve the intelligence and situation awareness of local organizations and, in turn, support the Afghan National Assembly. This deployment ended in November 2005 with the withdrawal of the Canadian squadron from ISAF. Their base at Camp Julien was closed and the focus of operations was transferred to the Kandahar region.

Operational Arguments: The demands placed on Canadian forces during ATHENA made it an ideal testing ground for the deployment of UAVs. There was a clear need to provide ground forces with tactical and operational information as a means of mitigating the increasing threats from insurgent operations. Not only was it felt that the intelligence provided from UAVs would help to reduce the impact of any attacks, for example by providing advanced warnings of unusual activity. It was also felt that the presence of UAS would help to reduce the likelihood of attacks by providing a visual reminder of their surveillance capability.

A number of further arguments motivated the deployment of UAS. The diverse and changing demands on ISAF units made it difficult to coordinate the deployment of conventional air resources. The previous chapter has identified some of the hazards that can arise when aircrews are briefed and then re-briefed as mission objectives are revised over time. The difficulty of changing targets in-flight can be significantly reduced when remote, ground based operators are provided with additional communications and intelligence sources that are, typically, not available to aircrews who are on flight to a remote target.

As mentioned in the introduction, a further motivation for the deployment of UAVs was that they could be operated from unprepared landing zones without the risks to air crews that are associated with brown out operations. Many UAS can be launched using catapult technology from the back of trailers. If necessary, they can be deployed close to the ground forces that require their assistance. Many UAVs also employ novel landing technologies, including parachutes and airbag systems. This enables them to be retrieved many miles away from a prepared launch site. The consequent operational benefits should not be underestimated, given that ISAF was operating in high, mountainous terrain and it was not always possible to fly directly back to an operational base. These innovative recovery techniques, therefore, helped to significantly increase UAS endurance.

Political Arguments: Previous chapters have argued that military procurement is a political as well as an operational process. The acquisition of new technologies must be justified either in terms of efficiency or increased effectiveness. This, in turn, implies that new systems must support either cost reduction or capacity enhancement. In either case, their procurement can create significant political benefits for the administrations that oversee their procurement. Capability enhancement is important because it, typically, helps to support the foreign status of a nation. For instance, new capabilities can enhance the influence of a state within military coalitions. Cost reduction through the development of novel military technology provides domestic benefits; increasing the confidence of tax payers that military expenditure is ‘under control’. Problems can arise when the capability improvements or cost reductions help to obscure some of the consequent operational risks that can arise from the deployment of new technologies.

In August 2003, the Canadian defence minister, therefore, announced the acquisition of an Unmanned Airborne System (UAS) consisting of four UAVs, two control stations and support facilities. When announcing this procurement, the Minister explicitly justified their acquisition in terms of risk mitigation for the ISAF ground forces. “In military terms, UAVs will decrease the risk to troops in Afghanistan. The security threat is a big concern for all Canadians, especially those serving in Kabul, and I want to ensure that they have the necessary equipment for the operation.” This purchase was also intended to meet a further political obligation to the North Atlantic Treaty Organization (NATO) for Canada to possess a UAV capability by 2004. As we shall see, however, the initial enthusiasm for the deployment of UASs was soon tempered by the organizational, technical and environmental demands that ATHENA placed on the equipment and its crews.

8.1.1 Political Pressure for Unforecast Operational Requirements

The first group of UAVs was purchased as an ‘unforecast operational requirement’. Most military organizations have created similar processes to support the rapid acquisition of supplies. They provide a means of responding to the changing needs of modern military operations, which as we have seen, are often conducted in environments that are very different from training bases. Unforecast Operational Requirements also provide military organizations with means of replacing technology that fails to meet the demands of a particular operation. They can also be used to rapidly acquire additional items that have proven to be particularly successful. However, these rapid procurement mechanisms are also associated with increased risks. They avoid the additional checks and oversight that help to protect traditional acquisition. There can also be particular problems when new systems are deployed without opportunities to adequately train troops in their operation. Chapter Six provided examples when it described the numerous hazards that emerged from the urgent deployment of additional night vision equipment in support of US operations in Iraq.

It has been claimed that the entire process from tender to deployment for the UAVs deployed by Canadian forces in support of Operation ATHENA took only seventeen weeks in late 2003. The Canadian Forces Director of Flight Safety subsequently remarked ‘the high risks associated with deploying a new system directly into the extreme operational environment of Kabul, Afghanistan had been identified prior to the deployment. The overriding operational requirement for this capability in theatre resulted in the acceptance of this risk’ [3].

The UAS chosen for the ATHENA operation was built by a French company and had five primary components:

1. The air vehicle based on a delta-wing design and a push propeller. The UAV had a wing-span off just less than seven meters and a top speed of around 80 knots. The maximum take-off weight of these UAVs was 330 kgs with a 45 kg payload. This innovative arrangement provided a high degree of maneuverability;

2. The Orientable Line-of-Site payload that provided imagery on remote targets to the ground control team. Line of sight communications is required between the GCS and the UAV. Once ‘line of sight’ is lost, the UAV returns to a pre-programmed flight sequence for up to 15 minutes. The intention is to provide an opportunity for ground teams to re-establish communications. However, if no further contact is made then the vehicle will initiate recovery through the deployment of the parachute;

3. The ground control station (GCS) that operated the UAV. The Ground Control System had three working positions: the Mission Planner

coordinated current and future operations and reported to outside agencies; the Air Vehicle Operator controlled and monitored the vehicle; the Payload Operator performed similar functions for the imaging equipment. The Mission Planner and Air Vehicle Operator workstations were identical and provided additional redundancy in the case of failure. In addition to the three working positions originally supported by the design, the ATHENA deployment also made use of an Air Vehicle Commander. This was, typically, an air force pilot or navigator. The commander did not have a control position but was responsible for monitoring the GCS screens of the Mission Planner and Air Vehicle Operator. This use of four-person rather than three-person crews was developed to meet concerns about Canadian military 'airworthiness requirements' during the deployment [4].

4. The communications infrastructure that linked data between the GCS, the UAV and outside agencies,

5. The ground support elements including a catapult launching system, as well as maintenance resources. Recovery involved the deployment of a parachute and a number of airbags. This provided flexibility and endurance benefits that were identified in previous sections of this Chapter.

The following sections use a number of mishaps that occurred during the ATEHNA deployment of the Canadian UAV system to illustrate some of the operational hazards that arise when new military technologies are deployed to meet a range of political objectives.

8.1.2 UAVs, Risk Erosion and the Loss of First Person Liability

The first major incident involving one of the Canadian UAVs led to category 'A' damage: the aircraft is destroyed, declared missing or sustains damage beyond economic repair'. This mishap occurred while the UAS was still undergoing in-theatre certification against the tight deadlines imposed for deployment [3]. On the day of the accident, the in-flight section of the test had been completed without difficulty. The Air Vehicle Operator (AVO) issued commands to start recovery. During this process, the engine must be shut down. This triggers the opening of the parachute door and releases a compressed spring, which deploys a drogue chute into the air stream. The force on the drogue extracts the main chute that in turn triggers the inflation of air bags under the nose and each wing.

At the time of this incident, winds around Camp Julien were measured at 3m/sec. This was sufficient to create 'standing eddies' in the lee of the Queen's Castle hill which overlooked the landing area. The eddies caused an instantaneous climb of 12m in half a second as the UAV passed through them. The forces created by the climb exceeded the escape velocity of the drogue-spring mechanism. In consequence, the main chute did not deploy and the airbag sequence was not triggered. Instead, the

UAV maintained a 7-degree nose-high pitch as the on-board computers waited for parachute deployment. Airspeed fell to the point where the UAV entered a glide mode as it passed over the Queens Palace and line of sight communications were lost.

The operators failed to consider that the release springs on the drogue chute would be insufficient to trigger the main parachute. This, in turn, led to the loss of control as the UAV passed beyond 'line of sight' in glide mode. A number of contributory factors combined to make these events more likely. The lack of bilingual reference material, of training flights over 1,500MSL and of Crew Resource Management techniques all combined to prevent the operators from accurately assessing the risks from local meteorological conditions.

In order to understand the reasons why these factors complicated the operation of the UAV, it is necessary to consider the longer term, political and organizational causes behind this mishap. Arguably, there was the mistaken assumption that UAV operations pose significantly less risks than conventional aviation and, consequently, require a much lower skill set. This is partly due to the erosion of 'first person liability' in which the crew of a conventional aircraft risk their lives to the airworthiness of the vehicle that they operate. This assumption, in turn, explains the rapid procurement of a UAV capability to support the Canadian involvement in ISAF as well as the need to meet the NATO commitment to acquire UASs. It was appropriate to conduct a rapid procurement because the deployment of UAVs would not endanger the lives of any aircrew. It is hard to image the same timetable for deployment being accepted for any conventional military aircraft.

The rapid acquisition created considerable time constraints that contributed to the operational risks of deployment. The UAS were deployed without comprehensive test and acceptance programmes. Similarly, the crews lacked training and documentation. It is difficult to underestimate the significance of this lack of training. The training of UAV crews is more important than it is in conventional aviation because the ground based operators must not only control the flight profile of their aircraft but they must also maintain situation awareness over unreliable communications networks using systems that appear extremely primitive in comparison to those available to the crews of conventional aircraft. In many of the mishaps described in this book, the training of ground crews for UAS fell short of even the reduced requirements for UAV operations.

8.1.3 UAVs, Human Factors and Remote Situation Awareness

A second major incident also resulted in Category A damage [4]. The crew were conducting their second flight after a 61 day layoff. They were practicing a range of recovery procedures at successively lower altitudes. The aim was to initiate the procedure earlier and earlier in the approach to provide additional time to track the in-bound leg of the flight. On the fourth circuit, the UAV hit terrain while descending in a final turn onto the in-bound approach. In this case, Canadian Military investigators identified the lack of Standard Operating Procedures (SOPs); the absence of a Standard Maneuver Manual; inadequate crew standard procedures and the lack of standard crew terminology had exacerbated the crews' lack of experience in the operational environment. These factors combined to create the context in which the ground team decided to further reduce the approach altitude on the fourth circuit. The lack of SOPs also explains why the Payload Operator had skewed their camera at 90 degrees to acquire the recovery area so that they had less opportunity to identify any potential collision with the mountain.

Further opportunities to predict the potential collision were lost by the ground crew's decision to set the automated altitude warning at 200 rather than 300m above ground level (AGL). This reduced the number of spurious alarms that were generated during routine flights in mountainous terrain. However, it also delayed the automated alarm so that the crew only received the warning a very short time before the collision. The large number of spurious alarms may also explain why the crew habitually ignored the aural warning associated with the altitude alarm.

The crews' apparent lack of situation awareness was exacerbated by the Airborne Vehicle Operator's decision to display engine monitoring information on their workstation rather than the altitude screen that might have provided additional cues to the potential danger from rising terrain. This decision can, in turn, be explained by the way in which the manufacturer's documentation stressed the need for the AVO to continually monitor engine parameters, for example to ensure correct fuel mixtures. However, this engine monitoring information was of limited value during this recovery stage of the flight.

Although all of these different contributory factors seem to have an operational focus, it is important to stress that they can be traced back to longer term political and organizational hazards. The tight deadlines associated with the UAV deployment prevented the development of specific SOPs and associated procedural support. The simulator time and other forms of training before operations focused on low-lying flat ground that was very different from the conditions encountered around Kabul. This may, in turn, explain why the crew failed to recognize the risks of a collision with terrain and why they did not focus on the altitude screen during the fourth recovery circuit. Once in the field, it can be extremely difficult to address the many operational problems that are created by the rapid procurement of complex military systems. These problems are compounded by the lack of necessary doctrine, either in the form of Standard Operating Procedures for a particular region or in the form of supporting techniques such as Crew Resource Management.

8.1.4 Hazards Created by UAS Configuration

A third mishap led to a category 'B' incident; the aircraft sustained damage to major components requiring the vehicle to be shipped to a 3rd line repair facility but where the overall structural damage was assessed to be within economical repair [5]. This occurred during a training exercise to familiarize a new crew. This and the previous incidents reinforce a point made in Chapter Seven about the need to monitor the dynamic nature of risks within military rotations. UAV operations illustrate the same increased level of risk during an initial deployment as was observed for US Army operations in brown out conditions.

Shortly after take-off the UAV entered a shallow descent into a populated suburb of Kabul. The AVC noticed that the UAV was producing insufficient thrust to sustain flight and so ordered an emergency recovery before the vehicle reached Kabul. However, the parachute deployed at too low an altitude for it to fully slow the vehicle before impact with the ground. Insufficient power was produced by the UAV because the number 1 cylinder carburetor's fuel mixture was too rich. This, in turn, was due to a lean mixture preset screw being advanced beyond the recommended ¾ turn, probably during routine maintenance.

The incorrect setting for the lean mixture preset led to a gradual fall in power during subsequent flights over the two days before the incident. This reduction was not noticed by the operating and maintenance teams partly because they did not usually record differences in the Engine Gas Temperatures between the cylinders. Such differences can be used to diagnose potential problems in the engine settings. A manufacturer's Service Information Letter (SILs) had recommended that these values should be analyzed. However, this document had not been received in theatre by the time of the accident. The manufacturer's service bulletins that described launch profiles for the Kabul area did not specifically consider these engine management issues in detail.

The relatively harsh operating conditions meant that the UAV was routinely being launched on the edge of its performance profile. The crews lacked the necessary documentation to judge whether the UAS was correctly configured to meet the prevailing environmental conditions. Even if they had been provided with this data, the lack of doctrine governing sleep and duty rotations for UAV crews may have prevented the effective use of guidance material; two of the crew had only had 4.5 hours between two duty periods. This observation builds on arguments about the role of fatigue in military risk assessment introduced in Chapter Five.

A number of further operational factors can be identified in the causes of this mishap. For instance, the UAV never entered the climb phase that might have provided the crew with the opportunity to alter the fuel mixture. However, there was insufficient time for them to complete any adjustments in the short interval before the crash. It is uncertain whether the available techniques could have been used to resolve the fuel mixture problems before the collision.

As with the previous two incidents, most of the causal factors can be linked back to the political decisions to deploy the UAVs at relatively short notice. The manufacturer had no time to work with the operational teams. Hence there was a lack of appropriate performance data and associated operational documentation. These problems were compounded by the difficulty of distributing the limited information that was available. Without the time necessary to prepare and disseminate these additional sources of information, it is little surprise that operational staff could not accurately configure the UAS. This prevented them from accurately assessing the mission risks that were posed by their operating environment.

Previous studies have remarked that the general standards of maintenance tend to be lower for unmanned as opposed to manned vehicles both in military and in civil systems [2]. One reason for this is that UAS are frequently seen as experimental, they are subject to frequent modifications that can be necessary to tailor the vehicles to operational requirements and to novel environments in a manner that is not typical for manned systems. The feeling of 'corporate responsibility' that often characterizes teams of co-workers who maintain conventional military aircraft may be less apparent in some UAV operations. The experience gained from Operation ATHENA suggests that maintenance requirements should be strengthened to ensure that there is no sense of complacency in the management of these systems.

8.1.5 Hazards Created by UAS Airworthiness Problems

The final incident reported during the Canadian UAS deployment in support of ISAF resulted in category 'C' damage; 'the aircraft must be flown to a contractor or depot facility for repairs, repairs are carried out by a mobile repair party, or a major component has to be replaced' [6]. During this incident, the crew lost communication with the UAV while it was 15 kilometers from the recovery zone. Attempts to restore communication failed and the UAV went into an autonomous recovery mode, landing in a residential area. The initial loss of 'line of sight' control occurred when the UAV descended to 3000m AGL. A mountain ridge interrupted signal transmissions. Communications were regained once the crew had implemented their emergency checklist and the vehicle had entered into an autonomous recovery mode.

A second interruption then took place. This occurred when the UAV was operating at 3,350m AGL, well beyond the required line of sight. However, this mishap was not due to the loss of communications. Subsequent analysis revealed that there was a 55 amp spike immediately before a voltage drop in the on-board systems. This loss in voltage is similar to that experienced during an engine shut-down but the avionics seemed to indicate that power was still being generated by the UAV. The subsequent investigation found numerous faults in the vehicle, one of these included the improper installation of the bracket that helped to retain an alternator cable. This, in turn, left the cable free to rub against a retaining nut and hence create the short circuit that would have interrupted electrical power to the UAV.

The manufacturer's guidance material provides a detailed diagram to illustrate the 'correct' installation of these components. This documentation together with the evidence of other maintenance problems, suggested that there were 'systemic problems' in the field maintenance of UAVs during ATHENA. These problems can be traced back to the operational demands that were created by the political decision to deploy the UAS' within extremely tight timescales. There is evidence that maintenance training was 'rushed' when the UAV infrastructure was being exposed to unforgiving operational environments.

The problems of maintaining air worthiness for UAS operations is typically of a wider class of hazards, which are easily overlooked in efforts to meet higher level political and organizational objectives. Responsibility for the resolution of such problems is often devolved to lower levels within the military procurement process. It is then often difficult for the individuals concerned to feedback operational difficulties back to the politicians and civil servants who first directed the acquisition. This creates a recursive problem. The more often politicians rely on urgent procurements without assessing operational hazards then the more likely they are to by-pass the processes that safeguard military acquisitions.

8.1.6 Wider Political Risks of Unforecast Operational Requirements

It is important not to focus simply on the technical impact of political decisions to rapidly deploy innovative technologies. Chapter Nine will argue that conventional risk assessments for UAV operations have been far too narrow in their scope. This section of the book will describe the numerous hazards that have to be faced by the units that must retrieve these vehicles when they are lost in hostile territory. UAS are not, typically, deployed in benign situations. Hence the reaction forces that are sent to secure a lost UAV are by definition going into 'high risk' environments. The consequences of such retrieval missions are seldom considered in the pre-deployment risk assessments prepared by manufacturers and suppliers.

A number of further problems were created by the political decision to deploy UAVs without adequate risk assessment. The lack of reliability illustrated by the case studies, in this chapter, increased the workload of military investigators. They had to piece together the causes of failure that undermined complex airborne and ground-based systems. Each of the short descriptions provided in previous sections was the product of many months of interviews, technical analysis and reconstruction [7]. This diluted the resources available to other operations, especially as the political investment in the UAV deployment made it a high priority to ensure the success of future operations by learning as many lessons as possible from a growing number of mishaps [8].

UAV operations also carried a political cost that is often underestimated by the proponents of these systems. In particular, they often act as a focus for local opposition. The civilian population faces the obvious hazard of being struck by one of these vehicles. Even the micro-scale UAVs carry sufficient kinetic energy to cause fatalities. In consequence, most civilian regulators segregate these systems so that they cannot be used over populated areas. The UK Civil Aviation Authority's CAP

722 and the US Federal Aviation Administration's (FAA) 08-01: Unmanned Aircraft Systems Operations in the U.S. National Airspace System (NAS), as well as EUROCONTROL's Spec-0102 on the Use of Military Unmanned Aerial Vehicles as Operational Air Traffic outside Segregated Airspace all impose rigorous constraints that limit the operation of these systems outside combat areas.

In addition, there are significant risks to the local population from the 'collateral damage' that has been caused by the use of combat UAVs. These have been exacerbated by problems in target identification using aerial reconnaissance. These issues often are not adequately considered during procurement. In consequence, ground forces are continuing to face the local resentment created amongst the civilian population by the deployment of UAS technologies in many areas of Iraq, Afghanistan and Pakistan [9].

It is important to reiterate the parallels that exist between the rapid deployment of UAVs and night vision technology, analyzed in previous chapters of this book. In both cases, there were strong political and operational arguments in favor of these systems. In both cases, urgent procurement orders were used to avoid some of the procedural safeguards in conventional military acquisitions. However, the deployment of these technologies created a host of unintended consequences or consequential risks that had not previously been envisaged. For example, previous chapters have described a number of confrontations in which peacekeeping forces have fired on civilians who were surprised when patrols emerged from the darkness. Similarly, the impact of UAV attacks has created local resentment that has undermined some of the efforts of ground forces to establish cooperation in conflict regions. Night vision devices have also exposed their users to a number of additional hazards – for example, many drivers of military vehicles have been involved in fatal accidents when they fail to identify ditches that cannot be distinguished by differences in their heat signature. Similarly, UAS operations have increased the risks for the ground forces that are tasked to retrieve these vehicles when they are lost in hostile areas. Many of the problems during ATHENA arose because the crews failed to anticipate the hazards that arose from the operational deployment of UAVs [10, 11]. Unless these deeper problems are addressed then it is likely that we will continue to acquire systems that endanger the lives of those who operate and support them while they are 'debugged' in the field.

The operational experience during ATHENA forced senior levels on the Canadian defence forces to reconsider their use of UAVs. The initial focus was on replacing the platforms that were involved in the incidents in this Chapter. However, this did not happen immediately. It took many months before there was the political recognition that action needed to be taken. Many of the individuals and teams involved in the initial procurement recognised the need to take a more systematic view of UAS integration. There was a need to develop a plan for the acquisition of a family of UAVs ranging from small, lightweight devices that could be carried with infantry units through to heavier and more complex systems with advanced sensing technologies up to combat-enabled weapons. This strategic view included UAS within a wider vision of a network-enabled force. In consequence, the Canadian Forces created a Joint Project Office (JPO) and a UAV Roadmap supported by a ‘Battlelab’ that provides facilities for the validation of these new technologies. Arguably the most eloquent response to Operation ATHENA is the creation of a new eleven stage process to guide the acquisitions identified from the roadmap [12]. This measured process is in strong contrast to the Unforecast Operational Requirements that were described in the opening sections of this Chapter.

8.2 National Security and the Risks to Civilians

The previous section identified the operational hazards that can be created by political decisions to deploy innovative technologies in military operations. The focus was on tracing the operational risks that stemmed from the political decision to meet a NATO commitment to provide the Canadian Defence Forces with a UAS capability inside the ISAF coalition. It is important to stress that the hazards of political decision making extend well beyond military operations. The following sections, therefore, demonstrate that similar concerns arise from the decisions to deploy military technologies in civil contexts.

8.2.1 Overview of the Nogales Predator Mishap

In the early hours of 25th April, 2006, a Predator UAV crashed northwest of Nogales International Airport, Arizona. Although it landed in a sparsely populated residential area, there were no injuries but there was substantial damage to the aircraft. The Predator B is a turboprop aircraft with redundant, fault-tolerant avionics. It can be flown by a remote pilot or autonomously. It was designed as a long-endurance, high-altitude platform with a wingspan of 66 feet, a maximum weight of 10,000 pounds and a maximum speed above 220 knots.

The National Transportation Safety Board (NTSB) coordinated the immediate investigation of the mishap [13]. As with the Canadian UAV accidents, it was possible to identiofy a number of immediate operational and technical causes for the accident. It was triggered by the ‘lock-up’ of a ground control system that then forced the pilot to change from one pilot payload operator position (PPO-1) to another (PPO-2). The NTSB concluded that the loss of the Predator was caused by the pilot’s failure to use an appropriate checklist when switching control from PPO-1 to PPO-2. In making this change, he forgot to alter the position of the controls in the new position. This resulted in the fuel valve inadvertently being shut off, which in turn starved the engine.

The UAV was owned by the US Customs and Border Protection (CBP) agency but at the time of the crash was being operated under contract. In other words, the pilot was employed by an external contract to work for the CBP. This commercial relationship is explained by a political imperative to rapidly increase the CBP’s use of unmanned surveillance aircraft to improve security along the United States’ southern borders.

The use of UAVs along the Southern border was part of a wider initiative to reduce and ideally halt illegal immigration. The Secretary of Homeland Security pioneered the Secure Border Initiative (SBI). He coordinated the construction of a network of ground based sensors, together wih more than 1,500 camera towers and a distributed computational infrastructure known as the Secure Border Initiative Network (SBINet). The political objective was to provide "a virtual fence. … It's going to be a smart fence, not a stupid fence — a 21st century fence, not a 19th century fence" [14]. The deployment of UAVs formed a key component within the Secure Border Initiative. The Nogales incident, therefore, provides an eloquent example of the civil risks that can arise from political pressure to extend the application of military technology.

8.2.2 Political Pressure for the Civil Use of Military Systems

The civil application of military UAVs raises a host of safety concerns. These include the need to mitigate potential hazards to both the public on the ground and also to other air space users. In the United States, the Code of Federal Regulations Title 14 covers the rights of way of aircraft. However, this includes terms such as 'see and avoid' that make little sense in the context of unmanned operations. In consequence, the FAA issued Interim Operational Guidance 08-01 on the integration of UAS into civil air space in March 2008. This document is used to determine if UAS may operate within the U. S. National Airspace System (NAS). Members of the FAA Aviation Safety Unmanned Aircraft Program Office and Air Traffic Organization apply 08-01 to evaluate applications for a Certificate of Waiver or Authorization (COA) whenever an organization makes a request to begin or renew UAS operations. This guidance covered the deployment of the Predators to monitor the US-Mexico border within the Secure Border Initiative.

At the time of the Nogales accident, the CBP UAS was operating under a COA that reflected its role within the Department of Homeland Security. The Departments of Defense or Homeland Security may decide that the operation of a UAV is necessary to protect national interests. In such circumstances, 08-01 recognizes that the FAA might approve an application that would not otherwise be acceptable. However, the applicant must identify and accept all risks that arise from any authorization. For example, Section 6 of 08-01 addresses airworthiness requirements and stipulates that all UAS must be in a fit state to conduct operations in the NAS. In particular, there is a requirement that the components of the system be maintained and conform to "the same airworthiness standards as defined for the 14 CFR parts under which UAS are intended to be operated" [13]. The desire to meet the political imperative to secure the Southern borders created a situation in which "At the time of the accident, CBP was unable to certify to the FAA that BP-101 was airworthy. Because of national security issues and past experience with similar UASs, the FAA temporarily waived this requirement for the issuance of the Certificate of Waiver or Authorization (COA) to operate in the National Airspace System (NAS)".

One of the reasons why the FAA waived section 6 was that the CBP had been directed to start flying the Predator B programme at short notice as part of the Secure Border Initiative, just as the Candian Defence Forces were ordered to meet tight deadlines for the introduction of UAS capability into Operation ATHENA. Not only did national security interests support the accelerated introduction of military technology for civil applications, they also placed constraints on public scrutiny even after the accident. The CBP asked that the NTSB did not release specific information in the accident report, for instance about the terms of the COA, which might provide details about the operation of the Predator.

8.2.3 Civil and Military UAV Airworthiness Requirements

The CBP COA provided exemptions from some of the requirements under FAA 08-01 for civil operators; these stated that COA applicants must apply for a special airworthiness certificate and that they must submit all the necessary data to demonstrate that the UAS is "designed, built, and maintained in a safe and airworthy condition". The Nogales accident illustrates how far the CBP strayed from these civil requirements under their exemptions for national security. An important factor in the underlying causes of the accident was that work arounds were routinely accepted to enable safety-critical operations to continue [15, 16]. As in the Canadian examples, maintenance procedures were often poorly documented and there was a lack of information about corrective actions.

The high number of previous failures and the inadequate maintenance actions may also have reflected deeper problems in the risk assessment practices that were intended to guide the operation of the CBP UAS programme. In other words, many of the problems that characterized the lack of airworthiness in military UAS operations were also apparent in civil applications. The Nogales accident was triggered by the lock-up in the ground control system. A review of a computer log showed nine previous lock-ups in the three months before the mishap. Two of these occurred before launch on the day of the accident. The maintenance logs did not record any attempts to correct another incident that had occurred six days before. The NTSB, therefore, concluded that these incidents had become normal or routine. They were corrected by cycling the power rather than by addressing the source of the problem.

It can be argued that some of these practices reflected the heavy demands that were being placed on the CBP to sustain and increase their UAV operations in the face of strategic and political concerns over the integrity of the border. These practices may also have come about as a result of the need to integrate CBP activities with support functions from a number of sub-contracting companies. There was also a political desire under the Bush administration to use commercial relationships as a foundation for National Security in the aftermath of the 2001 attacks. This created a distribution of responsibilities where it was not always clear whether the CBP or the sub-contractors were responsible for monitoring the safety margins that were intended to support continued UAS operations within the National Air Space. NTSB investigators could not find any explicit process for testing the UAS after maintenance. For instance, there was no method for checking whether the workaround of rebooting the ground control system console had any undesired side-effects prior to launch. The investigators concluded that "neither the CBP nor its contractors had a documented maintenance program that ensured that maintenance tasks were performed correctly and that comprehensive root-cause analyses and corrective action procedures were required when failures, such as console lockups, occurred repeatedly. As a result, maintenance actions could not be relied upon to be effective or repeatable, which is a critical factor in ensuring airworthiness" [13].

8.2.4 Public Use, Regulation and Sub-Contracting

Tracing the root causes of the Nogales accident, the subsequent investigation returned to the 'public use' provisions under which the COA was granted – the use of this term is instructive because it represents an important distinction from the national security reference that is retained within FAA 08-01. The relevance of these exemptions is summarized in the observation that the CBP was acting both as the regulator and also the operator of the UAS. It is for this reason that they were charged both with implementing an effective maintenance plan but also monitoring that level of effectiveness:

> "...because the CBP UA operation is considered public use, the FAA is not responsible for overseeing many aspects of the CBP's UAS program. The CBP must fulfill the roles of the regulator (which normally conducts oversight of operators) as well as the operator of its UASs; thus, the CBP must not only establish an effective maintenance program plan for its UA operation but also must monitor its implementation" [13].

This overloading of responsibility is, typically, not permitted in safety-critical industries. Many previous accidents have stemmed from confusion between regulation and operation [17]. These issues might have been resolved if the CBP had chosen to adopt a more focused regulatory role, while their sub-contractor had taken a narrow operational responsibility. However, the investigations argued that the CBP lacked the necessary resources, especially maintenance and engineering expertise, to oversee the work of their sub-contractor. They were not in a position to determine whether or not the external agency was employing sufficient mitigation to offset the risks that their operations potential posed to other air space users and to the wider public. If the contingencies of public use and national security are to continue to justify exemptions from FAA 08-01 requirements then it seems clear that agencies such as the CBP need to recruit sufficient expertise to discharge their regulatory role. The risk assessment and hazard analysis practices that characterized the military use of UAS within operation ATHENA are clearly not acceptable within the civil systems that govern National Air Space.

A more difficult political question is whether such waivers ought to be allowed in the first place given the clear threats to public safety that arose from the regulation of UAVs both before the Nogales crash and with the continued provisions for exclusions embedded in FAA 08-01. Any subsequent revision of the FAA guidance should require that ‘public use’ and ‘national security’ exemptions must be supported by detailed descriptions of the regulatory processes that will be used to monitor airworthiness in lieu of the FAA. These are likely to be less onerous than the provisions for civil operators even if they are more demanding than those expected for military operations. However, there is a need to maintain public confidence in the safe operation of UAVs when they are operating in the National Airspace System.

8.2.5 Risk Management in Civil Applications of Military Systems

FAA 08-01 requires that formal risk assessments are conducted to justify UAV operations. This can be difficult given the lack of long-term operational experience with many UAS applications. The quantitative aspect of these assessments may have to be based on expert judgment rather than the performance of complex safety-related systems in comparable operations. Most of the experience gained with these systems has been during military operations where many of the details of previous mishaps remain classified. This makes it difficult to assess the likely consequences of particular failures. Similarly, most military organizations will not publicize the frequency of UAV operations that they conduct. This prevents civil operators from using previous incidents to calculate the probability of future mishaps.

Previous sections of this Chapter have argued that political requirements have enabled operators to use national security arguments instead of the risk mitigations that would otherwise be expected. However, this *does not* exempt the operators from conducting such a risk assessment in the first place. In other words, it is acceptable for national security considerations to justify an increased level of acceptable risk – providing an agreement is reached with the FAA Administrator. However, it remains an FAA requirement that an explicit risk assessment is conducted in the first place to identify the hazards that are being accepted.

The Nogales accident demonstrated that the operators lacked any clear plan to mitigate the risks associated with the operation of the UAS under degraded modes of operation. The lock-up on PPO-1 was only one example of several other deficiencies. For instance, another unresolved component problem had disabled the satellite communication control function of PPO-2. The investigators argued that one potential mitigation would have been to introduce a minimum equipment list or a deviations guide. These documents can help operators to identify spare parts that should be retained in order to help engineers promptly respond to any failures that do occur. However, there were remarkably few parts at the CBP Predator facility. This may have constrained the opportunities that maintenance technicians had both to intervene in but also to diagnose those failures that did occur. Without minimum equipment lists, it was difficult for the operators or the CBP acting in their regulatory role to assess the residual risk that might arise if further components should fail during flights adjacent to controlled air space and residential areas.

8.2.6 The Hazards of Autonomous Operation

Many military organizations are developing autonomous and semi-autonomous platforms. These systems can be deployed with minimal intervention from human controllers. For instance, the US Marines tested a range of autonomous 'follow me' support vehicles during the 2010 Rim of the Pacific (RIMPAC) exercises. The US Army has also experimented with a range of Autonomous Sniper Systems including rotary winged platforms. These platforms raise a host of additional ethical and security questions that do not arise for most UAV operations where Ground Control Systems supervise the operations of the remote platform. However, the Nogales accident illustrates some of the risks of autonomous operation. The US public were exposed to these risks as a by-product of the political imperative to commence UAS flights without a more sustained analysis of the hazards involved.

Several sections of the FAA08-01 guidance refer to the need to create and maintain lost link profiles for UAS operation. These provide a flight-plan that can automatically be triggered when the UAV detects that it is no longer in communication with a base station. For example, 08-01 requires that "In all cases, the UAS must be provided with a means of automatic recovery in the event of a lost link. There are many acceptable approaches to satisfy the requirement. The intent is to ensure airborne operations are predictable in the event of lost link".

Following the Nogales accident, NTSB investigators found that there were three lost link profiles stored on the ground control system. Only one of these could be active. However, the pilot could change their selection during an operation in response to changes in the area in which the UAV was being flown. For the Predator involved in this mishap, the profiles were typically intended to ensure that the UAV would turn to a lost-link heading, climb for approximately 50 seconds on full power at 105 knots in order to gain time and help reacquire the signal. The UAV then establishes a waypoint 2.5 nautical miles from the location where the link was first lost on the heading established for that profile. When this waypoint is reached or after half an hour, the vehicle will fly to a series of predetermined locations and altitudes. If contact cannot be re-established then the Predator will crash when the available fuel is exhausted.

Considerable care is required in creating and maintaining autonomous lost link profiles. FAA Guidance 08-01 Section 8 notes the dangers of UAV operations over populated areas; "It is the applicant's responsibility to demonstrate that injury to persons or property along the flight path is extremely improbable...UAVs with performance characteristics that impede normal air traffic operations may be restricted in their operations". UAS operations should avoid routes with heavy traffic or with open assemblies of people. These can only be approved in emergency or relief situations if 'the proposed mitigation strategies are found to be acceptable'.

UAV operators must, therefore, conduct a formal risk assessment with associated safety arguments to demonstrate that the residual hazards are "extremely improbable". In contrast, the NTSB investigation argued that there "was no standardized safety-based method for determining the routes for the lost-link flight path and that inadequate consideration was given to ensuring the flight path did not include flight over population centers, property, or other installations of value". The lost-link profile followed by the Predator on the day of the accident was unnecessarily complicated. It was also argued that the pilots were uncertain about the actual flight path of the UAV following the loss of communications with the vehicle. The investigation also found that the UAV would crash along the route specified in the lost link profile. This created considerable uncertainty about the potential location for any 'landing'. In future, it was recommended that lost link profiles lead to a safety zone.

The Nogales mishap provides more general insights into the potential hazards of autonomous operations in both civil and military systems. FAA 08-01 states that although all UAVs will have an element of autonomous operation; it is a requirement that there should be pilot in the loop capability before they can be allowed outside restricted air space. The loss of the CBP predator illustrates the problems that operators face in predicting the autonomous behavior of complex systems when direct control is interrupted. This accident provides important warnings about potential failure modes involving more innovative platforms such as the US Army sniper systems. Together with the Canadian experience of UAS operations out of Camp Julien, the Nogales accident also helps to illustrate the potential hazards from political support for early deployment.

8.2.7 Communications and Mitigation in Autonomous Operation

One of thje most direct means of mitigating the risks associated with autonomous and semi-autonomous systems is to warn colleagues whenever control has been lost. Section 8 of FAA Guidance 08-01 establishes communications requirements for the operation of UAS inside the US National Airspace System. Pilots must have immediate radio contact with relevant Air Traffic Management (ATM) facilities at all times if the UAV is being operated in class A or D airspace or under instrument flight rules. Prior to the Nogales accident, the CBP should also have notified the FAA and ATM of any changes to their lost link profiles within a COA. These updates would have helped to coordinate any response to an emergency loss of control. However, the changes to the lost link profiles had not been communicated to these other agencies. The NTSB, therefore, argued that there was a real potential for an in-flight collision. The UAV created a significant hazard for other users of the National Airspace System as the platform assumed autonomous operation following the loss of control.

After any loss of communications, the CBP COA required that the pilot in command immediately informed ATM of:

1. The UAS call sign.
2. UAS IFF (Identification, Friend or Foe) squawk.
3. Lost link profile.
4. Last known position.
5. Pre-programmed airspeed.

6. Usable fuel remaining (expressed in hours and minutes).
7. Heading from the last known position to the destination of any lost link emergency mission maneuver.

However, there was no communication between Albuquerque Air Route Traffic Control Center and the UAV ground crew about the lost link profile, as required by the COA. This lack of communication was compounded by the loss of power to the UAV following the console lock-up that triggered the accident. UAV functionality was seriously compromised as it began to rely on battery power. The aircraft shut down its satellite communication system and the transponder. If the transponder had continued to work with mode C altitude data then ATM might have been able to track the course of the UAV and warn other airspace users. An important finding from the Nogales crash was the need to modify UAS design so that transponder functionality should continue even under degraded modes of operation. Arguably, this ought to be an explicit requirement within 08-01; the importance of transponder information would seem to be more critical than in other forms of aviation.

A key objective for the coordination between UAS pilots and ATC is to ensure adequate separation between aircraft within the same air space. In order to help meet this requirement, the UAV was only authorized to operate in temporarily restricted airspace. Any other aircraft wishing to operate within this area had to contact ATC before entering. For the CBP operations, the UAV’s restricted operating air space extended along the southern border from 14,000 to 16,000 feet (4250m-4850m) Mean Sea Level (MSL). However, the loss of power prevented the UAV from maintaining its altitude. The Predator breached the lower limit of the restricted zone. The investigators, therefore, argued that the UAV was operating autonomously in unprotected airspace until it crashed. ATC contacted the Predator’s pilot after they lost contact with the vehicle and the transponder had stopped working. However, the pilot did not inform them that the UAV had descended below the 14,000 feet MSL. At this point, the pilot or the Air Traffic Control Officer (ATCO) should have declared an emergency and taken measures to alert traffic in the area. They should have warned neighboring Air Traffic centers to monitor the missing vehicle. The ATCO could also have started efforts to increase the level of surveillance on the UAV, for instance by contacting the Western Area Defense Sector to gather information using their height finding radar. None of these things happened following the loss of contact with the UAV.

The loss of satellite communications and the transponder reflect a deeper concern over the rapid deployment of military technology in civil operations. Air Navigation Service Providers are facing a host of problems in maintaining levels of safety against increasing commercial and political pressure to allow the integration of UAS into civil (‘controlled’) air space. However, the military heritage of most platforms has let to the widespread use of composite materials with aerodynamics that are deliberately intended not to be visible on conventional radars. These issues and the many problems faced by Air Traffic management in support UAV operations have remained largely invisible to the political groups that have promoted the introduction of these systems either for national security concerns or as a means of reducing the costs associated with conventional forms of aviation.

It is important not to underestimate the impact that the Nogales accident had upon many of the agencies involved in the operation of UAV’s by both military and civil agencies. The Navy and Air Force Safety Boards took a direct interest in the findings from the accident investigation and also monitored the changes that were made to the PPO and systems software in the immediate aftermath of the Nogales crash. The accident also had an indirect effect on the teams that participated in the drafting of the US Department of Defence, Unmanned Systems Safety Guide for DOD Acquisition. However, the impact was arguably not reflected in the final draft of this guidance. Although it directly mentions the Chernobyl mishap, the loss of the Space Shuttle Columbia and the Ford Pinto design flaws, it does not directly mention the lessons that Nogales provides for the acquisition and control of advanced military technologies.

8.3 Common Political Hazards in Civil and Military Systems

This Chapter has argued that political pressure for the introduction of novel technologies creates new hazards in military and civil operations. The opening sections analyzed the causes of a number of mishaps involving Unmanned Airborne Systems (UAS) platforms within the Canadian contribution to International Security Assistance Force (ISAF) as part of Operation ATHENA. The need to provide ground forces with additional support against insurgent forces combined with a political desire to meet previous NATO commitments. In consequence, procurement staff were faced with accelerated deadlines for an Unforecast Operational Requirement. The entire process from tender to deployment took approximately seventeen weeks in late 2003. This undermined the usual risk mitigation processes intended to identify hazards from the acquisition and deployment of new systems.

Many of the operational problems created by political pressure for the rapid introduction of UAS capability reiterate arguments introduced in Chapter Seven. There was a mismatch between the environment in which the systems were developed and the rigors imposed by high altitude operations around Kabul. Many of the subsequent mishaps might have been avoided if there had been more opportunities for operational staff to work with the suppliers. There was insufficient time for Canadian Defence Force personnel to identify appropriate configurations for the UAS engine management systems. Similarly, the platforms were pushed into service with little time to draft standard operating procedures or maintenance practices, including pre-flight testing requirements. In consequence, local civilians were placed at risk from flights that crashed in populated areas. The large number of mishaps placed considerable burdens on military incident investigation teams that supported the ATHENA deployment. Maintenance staff struggled to keep up with successive revisions to operating procedures as more and more lessons were learned in an ad hoc manner following the loss of each platform. Ground forces were placed at additional risk when they had to retrieve missing UAV’s. These are not isolated observations; the following chapter will describe some of the hazards faced by UK and Afghan troops sent to retrieve UAVs in counter insurgency operations.

It took a considerable amount of time before the operational experience during ATHENA triggered changes in longer term procurement practices. Growing pressure from teams in the field helped to persuade senior military personnel that a more strategic approach was needed. In consequence, the Joint Project Office (JPO) commissioned studies that followed the pattern established by the US Department Defense Unmanned Aircraft Systems Roadmap 2005-2030. The Canadian JPO initiative helped to scope out future roles for unmanned systems, ranging from miniature surveillance aircraft up to remotely operated weapons platforms. A key requirement was that future UAV operations should be closely integrated with future visions of a network-enabled force. It remains to be seen whether sufficient operational experience can be used to temper political pressures for the rapid deployment of such ‘future visions’.

The second half of this chapter has identified the hazards created by political pressure for the rapid deployment of military technology to support civil operations. In particular, previous sections have described how senior officials promoted the deployment of UAVs within the Department of Homeland Security’s Secure Border Initiative (SBI). The US Customs and Border Patrol used external contractors to meet this requirement, following the administration’s emphasis on partnerships between Federal and commercial organizations. Political support for the deployment of advanced technologies following the attacks of 2001 helped to shape a regulatory regime that could dispense with many of the FAA’s requirements for UAS operations. National security and public use provisions led the CBP to fill a dual role in operation and regulation. However, they lacked the necessary expertise to determine whether or not their sub-contractors had taken sufficient steps to mitigate the risks to the public and to other air space users. The subsequent NTSB investigation identified systemic problems in the supervision of maintenance procedures. It also identified a host of lessons for the future operation of autonomous systems in military and civil applications, following the loss of direct control by the contractor’s ground staff. Unless we learn the lessons from this incident and the Canadian deployment of UAVs within Operation ATHENA then there is a significant risk that political pressures will continue to blunt the edge provided by new technologies.

8.4 References for Chapter Eight

[1] K.W. Williams, A Summary of Unmanned Aircraft Accident/Incident Data: Human Factors Implications. (December 2004), DOT/FAA/AM-04/24, Office of Aerospace Medicine.

[2] C. Patchett and V. Sastry, A Preliminary Model of Accident Causality for Uninhabited Autonomous Air Systems and Its Implications for their Decision

Architectures, 10th International Conference on Computer Modelling and Simulation, IEEE Computer Socirty, 487-492, 2008.

[3] Canadian Forces Flight Safety Investigation Report (FSIR), CU161 Sperwer Unmanned Aerial Vehicle (UAV), File 1010-CU161003 (DFS 2-6), 17th November 2003, Camp Julien, Kabul, Afghanistan, 17th November 2005.

[4] Canadian Forces Flight Safety Investigation Report (FSIR), CU161 Sperwer Unmanned Aerial Vehicle (UAV), File 1010-CU161005 (DFS 2-4), January 2004, Camp Julien, Kabul, Afghanistan, 18th November 2005.

[5] Canadian Forces Flight Safety Investigation Report (FSIR), CU161 Sperwer Unmanned Aerial Vehicle (UAV), File 1010-CU161002-1 (DFS 2-3), 20th March 2004, Camp Julien, Kabul, Afghanistan, 10th April 2007.

[6] Canadian Forces Flight Safety Investigation Report (FSIR), CU161 Sperwer Unmanned Aerial Vehicle (UAV), File 1010-CU161004 (DFS 2-3-2), 20th March 2004, Camp Julien, Kabul, Afghanistan, 20 February 2007.

[7] C.W. Johnson, A Handbook of Accident and Incident Reporting, University of Glasgow Press, Glasgow, Scotland, 2003. http://www.dcs.gla.ac.uk/~johnson/book

[8] J.R. Fitzsimonds and T.G.Mahnken, Military Officer Attitudes Toward UAV Adoption: Exploring Institutional Impediments to Innovation, Institute for National Strategic Studies, National Defence University, Washington DC, USA, 2007.

[9] J.C. Dawkins, Unmanned Combat Aerial Vehicles: Examining the Political, Moral, and Social Implications, MSc Thesis, School of Advanced Air and Space Studies, US Air University, Maxwell Air Force Base, Alabama, USA, June 2005. Available on https://research.maxwell.af.mil/papers/ay2005/saas/Dawkins.pdf

[10] C.W. Johnson, The Operational Strengths and Weaknesses of Military Night Vision Equipment, Defence Management Journal - Yearbook 2004, 72-75, PCSA International, Newcastle Under Lyme, UK.

[11] C.W. Johnson, The Paradoxes of Military Risk Assessment, In A.G. Boyer and N.J. Gauthier, Proceedings of the 25th International Systems Safety Conference, Baltimore, USA, International Systems Safety Society, Unionville, VA, USA, 859-869, 0-9721385-7-9, 2007.

[12] S. Wheatley, The Time is Right: Developing a UAV Policy for the Canadian Forces Centre for Military and Strategic Studies, University of Calgary, Canada, 2004. http://www.cda-cdai.ca/cdai/uploads/cdai/2009/04/wheatley04.pdf

[13] NTSB, Safety recommendation A-07-70 through -86: Loss of a Type—B Predator 10 nautical miles northwest of Nogales International Airport, Nogales, Arizona, April 25, 2006. Washington DC, USA, October 2007.

[14] C. Bronk, Managing The U.S.-Mexico Border Problem, The James A. Baker Institute For Public Policy, Rice University, USA, August 2007. http://scholarship.rice.edu/bitstream/handle/1911/20482/WWT_US-Mexico.pdf?sequence=1

[15] C.W. Johnson and C. Shea, The Hidden Human Factors in Unmanned Aerial Vehicles. In R.J. Simmons, D.J. Mohan and M. Mullane (eds.), Proceedings of the 26th International Conference on Systems Safety, Vancouver, Canada, 2008, International Systems Safety Society, Unionville, VA, USA, ISBN 0-9721385-8-7, 2008.

[16] C.W. Johnson and C. Shea, The Contribution of Degraded Modes of Operation as a Cause of Incidents and Accidents in Air Traffic Management. In Proceedings of the 2007 International Systems Safety Society Conference, Baltimore, USA, 2007.

[17] C.W. Johnson, Failure in Safety-Critical Systems: A Handbook of Accident and Incident Reporting, University of Glasgow Press, Glasgow, Scotland, ISBN 0-85261-784-, 2003.

9 Military Risk Assessment in Counter Insurgency Operations

Military operations inevitably involve hazards that do not arise in most civilian occupations [1]. The need to conduct complex, multi-agency operations, often at night and to tight deadlines creates pressures that have few parallels. Chapter Seven has argued that local terrain, meteorological and climatic features further complicate military actions. Limited knowledge, contradictory information, the need to provide flexible orders and also allow for local initiative creates additional challenges. These pressures are exacerbated by the deliberate attempts of opposition forces to exploit perceived weaknesses. The consequences of mistakes and violations are also very different from those in most other endeavors. In consequence, there are significant barriers to the application of civilian risk assessment techniques in military operations.

The previous Chapter has shown that these problems are compounded by the need to procure and then maintain innovative technologies, including Unmanned Airborne Vehicles (UAVs) and network enabled capabilities [2]. Vendors, politicians and military planners are often motivated to deploy technologies before they are fully mature. These decisions are justified by the need to establish an ‘edge’ over opposition forces and hence reduce the risks to individuals in the field. However, new technology often acts as a forcing function that reveals underlying weaknesses in the composition and resourcing of military units [3]. Personnel have to develop coping strategies and ‘work arounds’ when innovative systems fail in unpredictable ways. This process of ‘making do’ exposes forces to increased levels of risk that are often not appreciated by the advocates of innovative network infrastructures or weapons platforms [4].

9.1 The Hazards of Counterinsurgency

This chapter extends these arguments by considering the risks of counter insurgency operations [5]. These conflicts occur in situations where the norms of ‘declared combat’ do not apply:

> “A counterinsurgency campaign is a mix of offensive, defensive, and stability operations conducted along multiple lines of operations. It requires Soldiers and Marines to employ a mix of familiar combat tasks and skills more often associated with non-military agencies. The balance between them depends on the local situation. Achieving this balance is not easy. It requires leaders at all levels to adjust their approach constantly. They must ensure that their Soldiers and Marines are ready to be greeted with either a handshake or a hand grenade while taking on missions only infrequently practiced until recently at our combat training centers.” [6].

Counterinsurgency operations tend to be more complex than conventional conflicts. They are characterised by political and religious convictions that sustain a direct and high level of violence. Figure 13 illustrates the cyclic nature of counterinsurgency operations. Combatants, typically, seek to destabilise a regime. They widen existing conflicts and, thereby, create opportunities to seize control. One way of doing this is to provoke an excessive response from conventional forces that alienates the local population. This sustains the insurgency and creates new opportunities for the violence that will trigger a further reaction. This leads to a cycle of attack, reaction and local alienation.

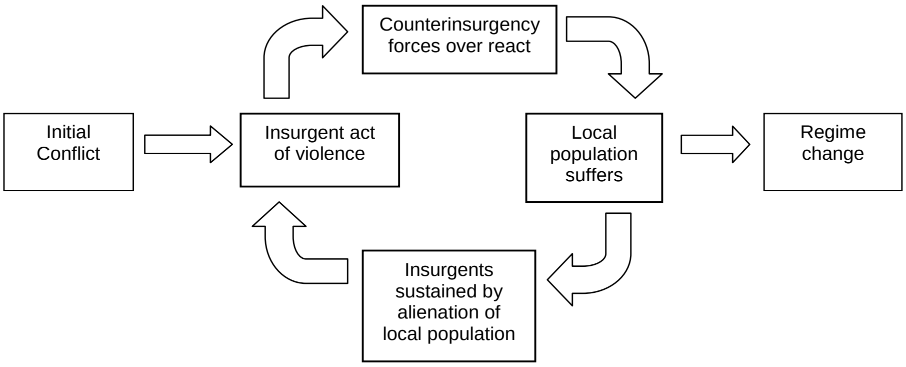

Counterinsurgency forces over react
Initial Conflict
Insurgent act of violence
Local population suffers
Regime change
Insurgents sustained by alienation of local population

Fig. 13. The Counterinsurgency Cycle

In order to break this cycle, counterinsurgents must often avoid the use of excessive force in order to avoid alienating the local population. This will deprive insurgents of the support hat will sustain further operations. However, this element of restraint will often place conventional forces at additional risk. Within NATO, rules of engagement are typically drafted to help commanders respond in proportion to the threat posed by enemy action. The constraints imposed by different rules of engagement limit the actions that units may take to mitigate these threats. Even with these rules of engagement, it is extremely difficult for troops on the ground to identify when to call upon additional support – there is a fine balance to be judged between the risks to local populations and the potential threat to counterinsurgency operations.The use of insufficient force can lead to a unit being overrun. The use of excessive force can lead to unnecessary civilian casualties that in turn fuel further conflict.

Paradox of Force Protection in Counterinsurgency:

In conventional combat, success can be gauged by the degree of protection that is provided for friendly forces and the extent to which security is denied to opposition units. In contrast, counterinsurgency operations typically gauge success by the degree of protection that is provided to the local population. This creates a number of paradoxes. For instance, it may be necessary to increase the risks to a counterinsurgency operation in order to reduce the hazards faced by the local population. One way of reducing the risks to counterinsurgency forces is for them to remain in their bases so that they are not exposed to hazards including Improvised

Explosive Devices (IEDs). However, they may lose contact with their civilian supporters. Insurgents may accuse them of avoiding confrontation. They can take the initiative in gathering local support. Conversely, the risks to counterinsurgency may be reduced by taking more aggressive measures by conducting spot checks and patrols with substantial deployments. This again risks alienating the local population. They may see counterinsurgency units as part of an occupying force that restricts their civil liberties. A proportionate approach ideally persuades local populations that they share the risks with counterinsurgency operations. This is intended to help civilians understand the reasons for patrols and checkpoints when they are necessary. It also forms part of a wider policy that focuses on local populations taking a greater and greater responsibility for their own security.

Paradox of Force and Risk in Counterinsurgency:

The previous section has argued that risk reduction strategies in counterinsurgency operations often avoid short term hazards but increase the longer term risks to successful operations. Excessive force or the use of additional force protection measures can alienate the local population and provide indirect support for insurgent forces. US Army doctrine summarises these observations:

> "Combat requires commanders to be prepared to take some risk, especially at the tactical level. Though this tenet is true for the entire spectrum of conflict, it is particularly important during counterinsurgency (COIN) operations, where insurgents seek to hide among the local populace. Risk takes many forms. Sometimes accepting it is necessary to generate overwhelming force. However, in COIN operations, commanders may need to accept substantial risk to de-escalate a dangerous situation" [6].

As the frequency and violence of insurgent activities drops, it is necessary for counterinsurgency forces to accept even more risks. Any increase in security is likely to be accompanied by increasing expectations of reduced military activity. Local populations are less and less tolerant of intrusions that seem unjustified by the perceived level of threat from insurgent attacks. Hence, a key stage in any counterinsurgency operation is to determine when the risks justify the gradual transition from active military intervention to policing actions.

Risk Informed Counterinsurgency Operations:

Chapter Two argued that military risk assessments can create tactical vulnerabilities if enemy forces can identify patterns of behavior and decision making that are encouraged by particular hazard analysis techniques. For example, ambushes may be prepared in expectation that a risk averse force will commit more and more resources in response to an initial attack. These same observations also create opportunities for counterinsurgency operations. For example, insurgent forces face a number of obvious hazards. They are, typically, in a significant minority compared to conventional units [5]. They may also be isolated from the local population by the religious and political beliefs that motivate their actions. This creates significant vulnerabilities to intelligence activities and to infiltration by counterinsurgency operations. Much can be gained by studying the ways in which enemy forces address the risks that the face. For instance, insurgents often find it difficult to establish a base for their operations. If they select a remote area then this will distance them from the local populations that they seek to influence and can increase their vulnerability to external surveillance using network centric techniques and sensing systems described in the previous chapter. If they try to set up a base that is too close to major centers then this will increase their changes of infiltration; “bases close to national borders can be attractive when they are beyond the reach of counterinsurgents yet safe enough to avoid suspicions of the neighboring authority or population. Timely, resolute counterinsurgent actions to exploit poor enemy base locations and eliminate or disrupt good ones can significantly weaken an insurgency”. [6]

9.1.1 Case Study in the Hazards of Counter Insurgency

This chapter not only justifies the arguments made in the previous section, it also illustrates concerns that were first raised in Chapter Eight. This previous chapter argued that the proponents of Unmanned Aerial Systems (UAS) focus on the benefits that this technology provides for ground operations. They do not consider the risks to those who have to retrieve UAVs when they are lost in enemy territory. In contrast, the following pages describe the retrieval of a British Unmanned Airborne Vehicle (UAV) from Helmand Province in Afghanistan in June 2006. During this counterinsurgency operation, a member of the UK armed forces was killed by a single bullet wound to the head.

In October 2005, the 18th Battery of the 32nd Royal Artillery Regiment in the British Army were tasked to conduct an operational trial of the Desert Hawk miniaturized UAVs. This trial led to the purchase and deployment of this UAS to Afghanistan in 2006 with elements of the 16th Air Assault Brigade. The intention was to provide UK forces with a ‘step-change’ in tactical situational awareness and to improve force protection for deployed troops. On the afternoon of the 11th June 2006, the 18th UAV Battery was using the Hawks to observe a suspected Taliban position near Sangin in Helmand. Shortly before 17:00, the operator reported that the UAV had ‘fallen out of the sky for some reason’ [7]. The Battery commander reported to the Ops Room at the Combat Outpost near Sangin and requested that a patrol be dispatched to recover it.

The Difficulty of Assessing the Value of New Technologies:
The decision to order the recovery had to offset the potential risks from the operation against the benefits of retrieving the Hawk. The cost of each UAV is relatively low, new bodies for the air vehicles are around $300 each. The payload and platform also had a relatively low security classification. However, the operators were anxious to know the reasons why the Hawk had come down. The recovery task fell to members of an Operational Mentoring and Liaison Team (OMLT) assigned to work with the

Afghan National Army at the Combat Outpost [8]. After the event, opinions varied amongst the OMLT as to whether or not the recovery mission was worth the risk.

The subsequent investigation argued that informal discussions had already taken place about the risks of mounting any operation to recover a missing Desert Hawk. There appears to have been a verbal understanding between the 18th Battery and the OMLT that the UAV should be retrieved. The UAV programme was still in its test phase and there were concerns that the insurgents should not discover the capabilities of the vehicle should one be lost. After the incident, the Major in command of the outpost argued that "their loss wasn't going to be particularly painful or a real drama but it was my understanding if one went down we should try, within reason, to bring it back".

Incomplete and Contradictory SOPs in Counterinsurgency Operations:
Given the complex and dynamic nature of counterinsurgency operations, it is impossible to draft Standard Operating Procedures (SOPs) that cover every possible contingency. A copy of the procedures for the Desert Hawk UAV on the wall in the OPS room in the Combat outpost stated that 'if a UAV ditches or lands short of the recovery point it should be recovered' it also warns that 'the recovery of a UAV should not be attempted if there is a risk to live'. The apparent confusion over whether or not the UAV had to be retrieved illustrates the difficulties that arise in military risk assessment. The SOPs stated that the UAV should be retrieved but without 'risk to life'. It is difficult or impossible for unit leaders to guarantee that any counter insurgency operation can be conducted without 'risk to life' and leaders were provided with little specific guidance on how to conduct such an assessment. In these circumstances it is hardly surprising that there was some uncertainty about whether or not a mission should be conducted to retrieve the missing Desert Hawk.

The Difficulty of Assessing Appropriate Force in Counterinsurgency:
Previous sections have argued that counterinsurgency operations raise a number of complex dilemmas or paradoxes that undermine standard forms of military risk assessment. Using too little force can create risks for conventional forces. It creates obvious operational and tactical vulnerabilities if insurgents can mass their forces quickly enough to launch a coordinated attack on a small force. In contrast, the deployment of additional troops can create an attractive target for ambush attacks and for IEDs. This can also serve to alienate the local population.

The Major in charge of the Combat Outpost approved the redeployment of a Patrol to help retrieve the UAV using 4 lightly armoured Snatch Land Rovers and a Weapons Mounted Installation Kit (WMIK) Land Rover equipped with a General Purpose Machine Gun. Afghan soldiers were also carried in Light Transit Vehicles. These initial units were equipped with Bowman High Frequency radio systems. They left the Outpost at 18:07 with last light around 19:00. This was the first time that the unit had deployed on patrol together and the first time that any Patrol from the OMLT had crossed the Helmand River. However, the participants were eager to conduct the operation. Chapter Seven has introduced some of the communications and coordination issues that arise when new teams are first deployed in combat operations. There is no evidence that Crew Resource Management techniques had been used to support OMLT operations as a means of mitigating these risks [9].

The Difficulty of Assessing the Risk of Insurgent Attacks:
The commanders of counterinsurgency operations must, typically, rely on past experience and on local intelligence when assessing the risks associated with any potential operation. In this case, there was evidence to suggest that there was a relatively low threat from insurgents in the area. The only previous incident had been the discovery of an Improvised Explosive Device close to the local Police post [7].

Unfortunately, risk assessments are seldom as clear cut as might appear from the previous paragraph. In areas such as Sangin it can be difficult to isolate different threat levels within particular regions. Insurgents frequently move position, trying to blend in with the local population as best they can. Even if the initial assessment was that there was a low level of risk around the combat outpost, it is important to recall that the UAV had been deployed to verify reports of Taliban activity. If this intelligence was correct then the relatively low initial risk assessments should have been revised. Without information from the UAV, however, it was impossible for units in the field to be sure that insurgents were operating in the area.

Trading the Risks of Known Routes against the Hazard of Massed Insurgent Attacks:
The retrieval patrol eventually chose a route that used the crossing point for the river, which the Desert Hawk had previously been monitoring for Taliban activity. This provides a specific example of the problems associated with assessing risk exposure in counterinsurgency operations. Alternate routes would have significantly delayed the retrieval mission. These delays would have enabled insurgent to mass their forces and provide an opportunity to prepare for the patrol when it eventually arrived at the crash site. Instead, the local commanders took the decision to rapidly ford the river at a location where there had been suspected Taliban activity but that reduced the amount of time for insurgents to coordinate their response.

A key theme of previous chapters has been that the dynamic nature of military operations often forces initial risk assessments to be revised in the face of new demands either from the environment, from higher levels of command or in response to enemy action. In this case, the vehicles were unable to cross the Helmand River. This increased the risks associated with the retrieval mission as the original force had to be split up. 15 troops were left behind as a rear party while 21 members of the retrieval patrol continued on foot.

When the patrol reached the suspected crash site, locals informed them that the UAV had been driven off in a pick-up truck. The advance group conducted a cursory search of nearby compounds and moved back across the river. They then began to receive reports on the Bowman HF radios that the Taliban were massing at the bazaar in Sangin. They collected their vehicles and began to return along the same route that they had followed to the crash site. As before, this involved a difficult decision about the risks associated with retracing their route back to the Combat Outpost. This was identified as a lower risk than using an alternate, circuitous approach than would have led them closer to the reported insurgent activity.

The Problems of Local Decision Making in Counter Insurgency:
Around 20:12, the retrieval patrol came under attack from small arms fire and rocket propelled grenades. The Bowman was used to inform the Ops Room at the Outpost; 'Contact, Wait'. After receiving the contact report a Quick Reaction Force (QRF) was told to 'Stand To' [7]. When the Major arrived in the Ops Room, he initially considered making a formal request for support to the Helmand Reaction Force. Instead, he decided to support the patrol with the resources at his disposal. In retrospect, this might be interpreted as a mistake. However, it is easy to criticize any risk assessment with the benefit of hindsight. It is also important to recall that the retrieval patrol had not requested additional support when they had come under attack. This decision is typical of the dilemmas introduced in the opening sections of this chapter; counterinsurgency operations inevitably expose conventional forces to levels of risk that cannot easily be assessed. Increasing levels of force can provoke a response that is just as dangerous as a lack of force in combat operations.

The QRF consisted of just less than 50 troops from the UK, US and Afghan armies in nine vehicles. The QRF members carried an SA 80 (A2) rifle, except for one sergeant who took a General Purpose Machine Gun (Light Role). Six members of the group had night vision equipment. All wore helmets and body armor. Many carried short range Personal Role Radios. However in the haste to deploy, the vehicles were not equipped with Bowmans; which were kept in a secure store. The Major in charge stated to the subsequent Board of Inquiry 'I was just going to take that risk and get out there rather than just faff around' [7]. This response illustrates the time pressures that complicate military risk assessment; it can be difficult to persuade personnel of the need to consider potential hazards on any mission, especially when comrades may be in danger. However, the Major's comments also illustrate particular problems for risk assessment in counter insurgency operations. As mentioned before, there are strong reasons to act as swiftly as possible. Additional delays provide insurgent groups with the opportunity to mass more of their forces against regular units.

Risk assessment techniques encourage military personnel to consider the potential hazards that could complicate each operation. However, the initial briefing of the QRF did not discuss what might go wrong nor did it propose any contingency plans. This has strong parallels with the lack of contingency planning that was identified as a contributory factor in the Puma-Lynx operation, described in Chapter Seven. As in this case study, enemy action forced rapid changes in an initial plan. These changes pushed ISAF units onto the 'back foot'. They were forced to react to insurgent operations and were, in consequence, left without any fallback plans.

It is difficult to convey the urgency that characterized the deployment of the QRF in support of their colleagues. Some members set off with no idea of where they were going. The rush to assist the retrieval patrol partly explains why the driver of one of the HMMWV's set off with the ignition keys for two of the other vehicles in his pocket. The HMMWV became entangled in barbed wire as it left the Combat Outpost. The driver eventually returned with the keys. In the meantime, the other HMMWV, two of the Snatch Land Rovers and two Afghan National Army vehicles left without noticing that the other vehicles had been delayed. From now on this group is referred to as Quick Reaction Force (QRF1). These events illustrate a 'Catch 22' problem for risk assessment in counter insurgency operations. The need to provide a prompt response and the difficulty of operating at night arguably increased the risks associated with the QRF's mission. These factors also made it more difficult for unit leaders to conduct any form of objective risk assessment. They chose to focus their attention on coordinating their response before additional insurgent forces could be deployed.

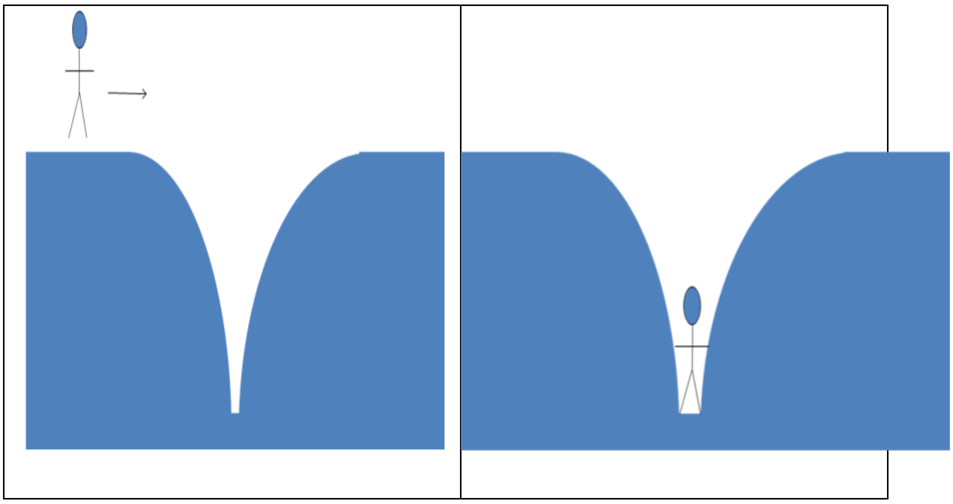


Fig. 14. The Accident Pit

9.1.2 The Accident Pit in Insurgent Operations

Figure 14 provides a sketch of the 'accident pit'. This helps to characterize some of the problems that occur when military operations must continually revise their assessment of the risks that they face. Initially, there appeared to be relatively little risk. The original retrieval patrol left the combat outpost in the belief that there was unlikely to be Taliban activity in the area. When they reached the crash site they discovered that the UAV had been taken and learned of forces massing near the bazaar. They rapidly revised their assessment of the risks, moving the figure in the previous diagram towards the edge of the pit. The deployment of the Quick Reaction Force along the same route as the previous patrol now further increased the likelihood of insurgent attack. The sense of urgency undermined attempts to coordinate their intervention or to develop contingency plans. By analogy, the mission was now heading to the bottom of the accident pit and there were very few options for escape.

The perceived need to provide prompt assistance was complicated by the fear that the route followed by the original Patrol was now covered by insurgent fire. Alternate routes through Sangin added a considerable distance to the journey and there was no information about whether or not these alternatives would also be targeted by the Taliban. The Major therefore set off along the original route with an initial plan to negotiate contact with the retrieval patrol using the personal radio systems. This decision reiterates previous points about the difficulty of accounting for risk exposure in military decision making, in this case preferring the hazards associated with a known route to the potential risks of a longer journey close to areas that were the centre of insurgent activity.

As mentioned previously, the Quick Reaction Force was divided into two sections as some of the drivers could not find the keys to their vehicles and one of HMMWVs had become trapped in the barbed wire surrounding the combat output. The first section, QRF1, eventually turned onto a narrow track where they were forced to stop. Although it was now almost 21:00, there were relatively good ambient light conditions. The members of QRF1 dismounted and discovered that some of the vehicles were missing. The subsequent briefing lasted about two minutes and established the Order of March. One person from each vehicle remained to protect their means of escape. The rest set off along a foot path bordered by drainage ditches.

On the Need for Training in Counterinsurgency:
A young man on a motorbike was stopped and sent to the back with Afghan National Army soldiers; however, he claimed not to have seen anything and the relative quiet of the march led the Major to assume that the insurgents had begun to withdraw. Shortly afterwards the forward members of the team noticed three men acting suspiciously. Two moved into the wood line and the third seemed to take cover behind a hay bale. One was observed to use a radio. The men moved off into a farm compound and QRF1 resumed their patrol. These events might have urged a more cautious approach. For example, the Order of March could have been changed to ensure that the General Purpose Machine Gun was closer to the front and that the unit leader was able to gain an overview of the rest of the patrol. However, QRF1 decided to ‘press on and make contact’. This may reflect the OMLT’s lack of experience in counter insurgency operations. Additional specialized training might have helped the unit leader to identify the potential hazards faced by QRF1 as it searched for the original patrol [10, 11].

The track was bordered by a drainage ditch inside a wall on its southern edge. There was another mud wall on the northern side that opened into a field with a bund line or embankment running from north-east to south-west. The Major used his personal radio to inform the rest of QRF1 that he had heard whispers some 30m (100 feet) ahead. Another member of QRF1 used his Common Weapons System (CWS) image intensifier to observe 12-15 people with small arms. The Major then shouted 'British Army, Stop or I fire'. Accounts vary as to the immediate events following this; however, the volume of fire directed at QRF1 was higher than they managed to return [7]. At this time, the members of QRF1 were either prone or kneeling. During this initial contact, a Captain who had volunteered for the mission was fatally wounded from a bullet to the head.

There then followed a period of approximately 5 minutes characterized by general confusion. Some members of QRF1 could not return fire in case they hit other members of the patrol. The Major decided to take the Captain's body back to the vehicles; this involved pulling him through a drainage tunnel while the others provided covering fire and used grenades. Some members of the party wanted to leave the Captain. Assistance could not be called from the vehicles because the personal radios were omitting a loud tone and could not transmit. QRF1 eventually managed to get back to their vehicles with the body of the Captain.

9.2 Immediate Causes of the Incident

Standard Operating Procedures (SOPs) can describe the steps that units should take in order to mitigate the potential risks that they encounter during counterinsurgency operations. For example, they often include minimum equipment lists that specify the communications and weapons systems to be used on particular missions. QRF1 did not have an SOP covering this support operation. The rapidly changing nature of the OMLT deployment meant that there was little opportunity to draft this guidance. There were also problems in providing basic IT for documenting SOPs, there was a lack of printers. In consequence, QRF1 deployed without a number of checks that might otherwise have been expected from the use of SOPS in counter insurgency operations. Team members were unclear about their role and objectives. They set off without having agreed upon the route to the retrieval patrol. There was no discussion of the contingency plans that might be used if opposition was encountered. They left the Combat Outpost without installing the Bowman radios. The subsequent Board of Enquiry argued that had the Major been able to use the HF radio system to communicate with the first UAV patrol and the Ops Room in the Combat Outpost then he might have been alerted to the hazards of attack from insurgent forces. He would then have been more aware of the risks being taken when he pressed on with the deployment of QRF1 [7].

9.2.1 Inadequate Briefings on Insurgent Risks

The lack of SOPs was compounded by *the limited nature of the briefings* both at the Combat Outpost and after QRF1 had left their vehicles. These briefings could have reviewed some of the decisions that contributed to the mishap. The leader of QRF1 went to the front in the Order of March. This deprived him of a tactical overview during the insurgents' attack. It may also have prevented him from communicating effectively to individuals at the back of the unit. More detailed briefings might have provided an opportunity to review the distribution of night vision equipment between the members of QRF1. This reiterates some of the concerns that were raised about training in the use of these innovative technologies in Chapter Six. There were also design limitations with their equipment. The patrol commander had to rely on a monocle device that was designed for US forces and could not be mounted to a British helmet. There was, therefore, no way for him to both observe and fire at the same time. These arguments again illustrate the complexity of military risk assessment. The time taken for additional briefings might also have provided insurgent forces with the opportunity to group against the members of QRF1 as they dismounted from their vehicles.

9.2.2 Lack of Night Vision Equipment

The risks associated with counterinsurgency operations can be influenced by the availability of appropriate equipment. The OMLT Chief of Staff had written to the Helmand Task Force Headquarters on several occasions before the incident expressing his concern over the lack of resources in his units. In May 2006, he had requested a list of 'mission essential equipment' for force protection. This included 48 Head Mounted Night Vision Goggles. These are the monocles that are, typically, worn around the neck by British troops. He had also requested 10 Common Weapons Systems which provide an image intensification facility mounted on the SA80 (A2). Chapter Six has argued that the provision of these systems can reduce some of the hazards associated with night operations and, at the same time, increase other risks. However, the Chief of Staff argued that 'neither the task being undertaken by OMLTs, nor the operational risk being taken, should...be underestimated; it is essential that teams are properly resourced' [7]. However, the mission essential equipment list was not sent to the right unit. This led to a 25-day delay. By the time of the incident, the request was approved but had still to be resourced.

9.2.3 Lack of Appropriate Firepower

This case study provides a detailed example of the problems in ensuring an appropriate response to counterinsurgency operations, identified within US Army Field Manual 3-24 [6]. Previous sections have argued that it can be difficult to determine the degree of force required to complete a particular mission without exacerbating local relations or presenting an undue target for insurgent forces. However, QRF1 lacked sufficient firepower after they had come into contact with the insurgents. This was compounded by the decision to dismount, leaving many of the heavy weapons on their vehicles. QRF1 might have benefitted from Underslung Grenade Launchers as well as additional Light/General Purpose Machine Guns. These omissions were particularly important, given the probability that insurgent forces would be carrying rocket propelled grenades.

The subsequent Board of Inquiry argued that the provision of additional weapons would have taken resources from other units in the Helmand area. Tracer rounds would have helped in the extraction of the patrol; although this ammunition had been delivered to the OMLT it had not been brought forward to the Combat Outpost. Although the provision of these items need not have prevented the fatality; they would have significantly reduced the risks to the remaining members of QRF1 as they fought their way back to the vehicles.

9.2.4 The Ambiguities of Counterinsurgency Operations

Counterinsurgency operations are complicated by the problems of distinguishing insurgents from members of the local population. The opening pages of this chapter have also identified the risks of alienating the local population through excessive force or through unintended attacks on civilian targets. Immediately before opening fire, the leader of QRF1 shouted 'British Army, Stop or I fire'. This may have been motivated by a desire to reduce the likelihood of civilian casualties. However, other members of the unit had already reported seeing a group carrying small arms. The shouted warning might also have been intended to reduce the risk of fratricide given that they had to locate members of the original patrol. Irrespective of the causes, the subsequent investigations argued that any delay between the warning and opening fire provided the enemy with enough time to respond aggressively [7].

9.3 Longer Term Risks in Counterinsurgency

The previous sections have identified a number of short term causes that contributed to the loss of life in this incident. The lack of SOPs can be traced back to the austere conditions in the Combat Outpost and to the lack of IT equipment throughout the OMLT. The members of QRF1 failed to appreciate the risks that they faced because they had not been trained in counter insurgency operations. The lack of appropriate firepower was the result of insufficient equipment being distributed across the OMLT. The following sections identify longer term causes that undermined counterinsurgency operations in Helmand.

9.3.1 Insufficient Personnel with Counterinsurgency Expertise

There were insufficient troops for a dedicated Quick Reaction Force to be continually on stand-by at the combat outpost. Instead, the team had to be formed on an ad hoc basis. This task was complicated because the OMLT support group was not drawn from infantry units; they came from the Royal Logistics Corps, Royal Electrical and Mechanical Engineers, Adjutant's General Corps etc. This may have frustrated attempts to establish a more coherent approach to contingency planning and to the risk assessments that might otherwise have anticipated some of the difficulties that led them into the accident pit, illustrated in Figure 14.

The lack of experience in counterinsurgency operations can be traced to both strategic and political decisions. The British Army had decided to staff the OMLT with full-time soldiers; their task was to support individual platoons within the Afghan National Army. In contrast, the US Army chose to develop Embedded Task Teams from reserve and National Guard units. Their support was concentrated at company level. The British Army also had to meet this greater demand on their personnel from within the 3,150 soldiers that the UK Secretary of State for Defence had previously announced to Parliament. This reinforces comments made in Chapter Eight about the impact of political decision making on the hazards of modern warfare.

9.3.2 Insufficient Support from Local Forces

The causes of the incident can also be traced back to differences in emphasis over the threat and force structure in the region. The US and Canadian emphasis was on 'full-spectrum' combat operations. In contrast, the UK Helmand Task Force focused on redevelopment and capacity building for the Afghan forces. These activities were intended to be a precursor to withdrawal. The OMLT played a pivotal role in this capacity building, acting as mentors for the Afghan National Army. However, it can also be argued that the focus on reconstruction left the OMLT ill-prepared for the 'mission creep' that led to their deployment in counter insurgency operations.

Many of the OMLT members were surprised to learn that they would have to fight alongside Afghan soldiers. There had been an assumption that they would only be involved in training and reconstruction activities. The lack of clarity over the role of the OMLT was reflected in their pre-deployment training. This lasted 2 weeks, well short of the 6 weeks recommended by some senior officers and was not well matched to the operating environment in Helmand. Concerns over the mismatch between training and operations have been a recurring theme of this book. The operational elements of the OMLT arguably received insufficient training about the hazards that faced them because the strategic planning for their deployment did not recognize the risks that would arise from their role in counter insurgency operations [12].

9.3.3 Inadequate Equipment

The 7th Para Royal Horse Artillery coordinated the planning for the OMLT. They, in turn, requested vehicles and communication support from Headquarters, 16th Air Assault Brigade. This left HQ with two choices; either to redistribute resources from other units in Helmand or to issue an Urgent Operational Requirement (UOR). Chapter Eight has identified the ways in which such initiatives can undermine the risk mitigation processes that safeguard other forms of military procurement.

The UK Ministry of Defence and the Treasury were unwilling to commit funds for UORs until there was a formal political announcement. The Secretary of State delayed the Helmand deployment for almost 2 months. He was anxious to ensure that the mission objectives could be met within the 3,150 manning cap. He was also keen to secure further commitments of support from other NATO members. Hence, the political desire to mitigate the longer term strategic risks associated with UK commitments in Afghanistan led to a short term increased in the operational risks for the OMLT as necessary resources were delayed in procurement.

9.4 Strategic & Tactical Risk Assessments for Counterinsurgency

Strategic and tactical constraints led to the gradual transformation of the OMLT from a reconstruction and training force into what amounted to a counter insurgency unit. Longer term problems, therefore, stemmed from the military decision making processes that underestimated the hazards created by ‘mission creep’.

9.4.1 Was the Strategic Risk Assessment Adequate?

During the first weeks of deployment for the Helmand Task Force it became clear that the focus had shifted from stabilization and reconstruction to counter insurgency. This led to a considerable drain on resources with priority being given to groups such as the Joint Helicopter Force rather than the OMLT. Tactical satellite communications, night vision equipment, machine guns were all allocated on a ‘whole fleet management’ principle in which risk assessment was used to determine those units with greatest need.

As we have seen, however, it is very difficult to assess the risk of insurgent activity for particular units in operating environments that are as complex and dynamic as Helmand. The lack of resources, including night vision devices and grenade launchers, hindered the original UAV retrieval patrol and the subsequent Quick Reaction Force. From the perspective of senior commanders, it is hard to anticipate these detailed local requirements. It is important to recall that there had been relatively little insurgent activity in the area around Sangin, with the exception of an IED used against the local police station. Such observations reiterate a point made in Chapter Three – it is difficult to base future predictions on previous incidents in military risk assessment. In some contexts, the relative lack of previous insurgent activity can be taken as an indication of increased risk of future attacks rather than any guarantee of future quiescence.

9.4.2 Was the Tactical Risk Assessment Adequate?

The Chief of Staff of the OMLT was concerned that higher levels of command in the Helmand Task Force did not understand the resource requirements for joint OMLT and Afghan National Army operations. In particular, he felt that their operational deployment would require additional support from the rest of the task force. The lack of armored vehicles, night vision equipment and heavy machine guns was compounded by under-staffing. He requested that Helmand Task Force HQ address these issues before the joint force was deployed ‘as a matter of urgency to help mitigate the significant risk being taken by the UK OMLT for this operation’.

This again illustrates the difficulty of using tactical risk assessment to guide counterinsurgency operations. Although individuals recognized the hazards from the changing nature of the OMLT deployment, only limited steps could be taken to mitigate the risks. Additional equipment was provided but this was insufficient to meet the operational needs. Political constraints also delayed procurement and placed strict limits on the deployment of additional personnel. At the same time, the rapidly changing nature of the environment in Helmand created a context in which it was particularly difficult to assess the threat level in at any particular time in any particular region. The assessments for Sangin fluctuated from ‘high’ through to ‘benign’ within days. The changing role of the OMLT illustrates a classic coping response in which highly motivated teams did their best to with the resources at hand.

9.5 From Counterinsurgency to Counterterrorism

This chapter has identified some of the problems that complicate risk assessment in counter insurgency operations. In particular, it can be difficult to balance the need to maintain the security of conventional forces while at the same time ensuring the support of the local population. Counter insurgency operations often involve significant risks, as troops cannot isolate themselves within their bases if the objective is to combat insurgent attempts to destabilize the prevailing regime. This creates a number of paradoxes. For example, there is a need to take increasing risks as the situation improves. Local civilians expect a transition to policing operations as the threat of insurgency declines.

The complexity of risk assessment in counterinsurgency operations has been illustrated by a detailed account of an incident in which a member of the UK armed forces was killed during an OMLT deployment. This fatality also served to illustrate some of the hazards identified in Chapter Eight as it occurred during a mission to retrieve a missing Unmanned Aerial Vehicle (UAV). Problems began when senior officers in the Combat Outpost failed to accurately assess the likelihood of insurgent activities in their area even though the UAV had been tasked to verify reports of Taliban movements. These were compounded by a lack of Standard Operating Procedures; these were not available because there was insufficient staff and a lack of basic word processing resources in the combat outposts. The difficulty in assessing the likelihood of insurgent activity, in turn, helps to explain why the initial patrol crossed the Helmand River at the same point that had been under observation by the Desert Hawk and why they then returned along the same route.

Subsequent sections in this chapter use the Accident Pit to explain why a relatively stable situation quickly deteriorated into combat activities in which a Quick Reaction Force was responding to insurgent activities rather than following agreed contingency plans. Once the original patrol had come under attack, ad hoc procedures were used to form a rescue mission even though this had not been requested from the first patrol. The perceived urgency of responding and the lack of SOPs together with limited experience in counterinsurgency operations contributed to a confused and disjointed response. Some elements of the QRF were delayed because they could not find the keys to their vehicles. Bowman tactical radios were not fitted before they left. The remaining members of the QRF were deprived of necessary firepower by the need to leave their vehicles behind as they moved towards the initial patrol. At each stage, the opportunities to conduct a more considered appraisal of the potential risks was undermined by the perceived need to hurry to help their comrades and to deny the Taliban any further opportunity to mass insurgent forces.

This incident illustrates the practical and theoretical barriers to the use of risk assessment in military operations. The dynamic and time critical nature of the mission, the need to ‘make do’ with limited resources and the strong desire to help colleagues fulfill mission objectives makes it unlikely that formal approaches to risk assessment would have provided strong benefits to the teams involved in this incident. At the tactical and strategic level, many individuals were aware of the hazards being faced by the units in the field. However, political constraints, resource limitations and the difficulty of predicting the level of threat posed by local insurgent operations all combined to frustrate the mitigation of those risks. Unless these wider issues are resolved then there is little prospect that the proponents of military risk management will realize the benefits that they anticipate.

This chapter has focused on some of the particular problems that complicate risk assessment in counterinsurgency operations. Unfortunately, many of these observations can be extended to the threats posed by domestic terrorists. In both cases there is a need to identify a proportionate response to uncertain threats while retaining the support of the local population. The following chapter, therefore, extends our analysis of counterinsurgency operations to consider the particular risks posed by Improvised Explosive Devices for military personnel and civilian populations.

9.6 References for Chapter Nine

[1] C.W. Johnson, Paradoxes of Military Risk Assessment, In A.G. Boyer and N.J. Gauthier, Proc of 25th Int. Systems Safety Conference, Baltimore, USA, International Systems Safety Society, Unionville, VA, USA, 859-869, 0-9721385-7-9, 2007.

[2] D. Russell, N. Looker, L. Liu, and J. Xu, Service-Oriented Integration of Systems For Military Capability. IEEE International Symposium On Object/Component/Service-Oriented Real-Time Distributed Computing, In Press. Orlando, Florida, 2008.

[3] J.L. Drury, L. Riek, N. Rackliffe A Decomposition of UAV-Related Situation Awareness, Proc 1st ACM SIGCHI/SIGART Conf. on Human-Robot Interaction, Salt Lake City, Utah, USA, 88-94, 2006, ISBN:1-59593-294-1.

[4] P. Houghton, Potential System Vulnerabilities of a Network Enabled Force, 9th International Command and Control Research and Technology Symposium, 131, Copenhagen, Denmark, UK paper 11, 2004.

[5] S. Metz and R. Millen, Insurgency And Counterinsurgency in the 21st Century: Reconceptualizing Threat And Response, Strategic Studies Institute, U.S. Army War College, Pennsylvania, USA, November 2004.

[6] US Army, Counterinsurgency, Headquarters, Department of the Army, Field Manual 3-24, December 2006.

[7] UK Ministry of Defence, Board of Inquiry Report into the Death of Capt J Philippson, Helmand Province, Afghanistan, 11th June 2006. London, U.K. 13th February 2008.
http://www.mod.uk/DefenceInternet/AboutDefence/CorporatePublications/BoardsOfInquiry/

[8] NATO, fact Sheet: NATO's Operational Mentor and Liaison Teams (OMLTs), Brussels, Belgium, October 2009.

[9] K. Mearns, R. Flin and P. O'Connor, Sharing 'Worlds of Risk': Improving Communication with Crew Resource Management, Journal of Risk Research, (4)4:377-392, October 2001.

[10] J.S. Corum, Training Indigenous Forces in Counterinsurgency, Strategic Studies Institute, U.S. Army War College, Pennsylvania, USA, March 2006.

[11] W.E. Sponsler, Striking the Balance Between Training High-Intensity Conflict and Counterinsurgency: Maintaining Full Spectrum Dominance in the US Army. MSc Thesis, Marine Corps Command And Staff College, Quantico Virginia, USA, 2008.

[12] B. Hoffman, Insurgency and Counterinsurgency in Iraq, The Rand Corporation, National Security Research Division, OP-127-IPC/CMEPP, Virginia, USA, June 2004.

10 Military and Civil Risks in Counter-IED Operations

The previous chapter has identified a range of problems that complicate the application of risk assessment techniques in counter insurgency operations. It is difficult to determine the appropriate degree of risk to expose troops to when they have to prevent insurgent attacks and at the same time retain the confidence and support of the local population. Isolating forces within checkpoints and bases can increase levels of force protection but will distance personnel from the people that they are trying to protect. The following pages build on this analysis and identify the additional complexity that arises from domestic terrorism. Many of the issues resemble those for overseas counter insurgency operations except that there can be far greater political, social and ethical constraints on police, security and military organizations operating inside their own national borders.

The role of risk assessment in counter terrorist operations has become an acute topic of interest since the attacks of 2001. Prior to this date, much attention was focused on what have been termed ‘failed states'; where terrorism was seen as a response to the lack of conventional political forms of expression [1]. Violence was seen as an alternative to other forms of political expression. A further strand of work focused on the risks of state sponsored terrorism. This concern was motivated by events including the US hostage crisis in Iran; “Possible military actions range from rescuing hostages to neutralizing terrorist camps and making direct strikes against targets verified as the infrastructure for state-sponsored training and support complexes of terrorist groups, The military response is part of a larger strategy that seeks to maximize the risk of punishment for terrorists and their sponsors and supporters while minimizing their potential rewards” [2].

10.1 C-IED Risk Assessments in Military and Civil Contexts

After the attacks of 2001, it was clear that a risk based approach had to be extended to consider a far broader range of potential targets. During his Senate confirmation hearing, Department of Homeland Security (DHS) Secretary Michael Chertoff asserted that “DHS must base its work on priorities driven by risk”. However, a subsequent report by the Congressional Research Service acknowledged that “While the practice of risk analysis may be advanced in the insurance and financial industries, it is relatively less developed in the homeland security field. Although there are numerous reasons that account for this dynamic, two primary reasons include (1) the dynamic nature of terrorism and ability of terrorists to adapt to successful countermeasures, and (2) the lack of a rich historical database of terrorist attacks, which necessitates a reliance on intelligence and terrorist experts for probabilistic assessments of types of terrorist attacks against critical assets and/or regions” [3]. The Rand Corporation reached similar conclusions; “Ultimately, efficient allocation of homeland security resources would be determined based upon assessment of the cost effectiveness of alternative risk-reduction opportunities. After potentially first addressing obvious and easily mitigated risks, this requires understanding the cost effectiveness of different types and amounts of investment. Neither the methods nor the data are available to answer questions about the effectiveness of available risk-reduction alternatives or to determine reasonable minimum standards for community preparedness" [4]. These studies mirror many of the concerns raised about military risk assessment in previous chapters of this book.

The remaining sections of this chapter present a high level model that has been used to apply risk assessment techniques in homeland security operations. Subsequent paragraphs explain how this has been used to guide both strategic and tactical decision making in counter terrorism operations. It is important to stress that these approaches, in common with the Rand report and the Congressional Research Review, provide a framework for high-level policy. They lack the 'rich historical' case studies that provide more detailed operational insights into the nature of the threats or potential countermeasures. The closing sections of this chapter, therefore, extend our analysis to consider the interaction between civil and military risk assessment in Counter Improvised Explosive Device (C-IED) operations. This focus on C-IED operations is justified by the scale of the threat that is proportional to the perceived threat posed by these devices. The US DoD and Congress have provided $15.9 billion to fund counter-IED programmes between 2004 and 2010. These initiatives have been coordinated by the Joint IED Task Force (JIEDDO), which has recently requested $3.5 billion for FY2011 alone. The UK Centre for the Protection of National Infrastructure and the US Department of Homeland Security's (DHS) Critical Infrastructure Programme have also focused on increasing the resilience of civil society against IED attacks. In May 2008, the DHS allocated $3billion to secure US critical infrastructure and transportation systems [5]. Much of this work has been directed by risk assessment techniques; "By better understanding terrorist tactics, first responders and private sector partners can improve their capabilities to stop terrorist attacks in the planning phase, thereby reducing the risk of IED attacks…The Office for Bombing Prevention focuses on reducing the risk of IED attacks by: Coordinating national IED awareness programs; Analyzing counter-IED requirements, capabilities, and gaps. A Multi-Jurisdiction Security Plan integrates the capabilities of multiple emergency services providers in areas that have many local jurisdictions, and its IED security plan outlines specific bombing prevention actions that reduce vulnerability and mitigate risk" [6].

10.1.1 Strategic Similarities

There are strong similarities between the manner in which risk assessment techniques guide US military strategy and the application of these methods to guide Homeland Security. Chapter 2 has described how the Chairman of the Joint Chiefs of Staff submits to the Secretary of Defence an annual assessment of "the nature and magnitude of the strategic and military risks associated with executing the missions called for under the current National Military Strategy". These assessments guide military expenditure; for example, by directing funding to programmes that address the greatest threats identified in the Chairman's Annual Risk Assessment.

The Department of Homeland Security has followed a similar approach in their strategic Risk Assessment Methodology. The allocation of funding to counter terrorism programmes has been split. 40% of the funds have been targeted for statutorily mandated initiatives. The remaining 60% were allocated to projects that mitigated the risk of attack, calculated in proportion to the population at stake:

[Risk = Population].

In the immediate aftermath of the 2001 attacks, it was not possible to identify a coherent methodology for assessing the risks posed to particular strategic targets. Instead, it was assumed that threats were proportional to the population, which might be affected by a potential attack. Over time, this formula was refined by assessing whether or not there was perceived to be a threat towards a particular target, whether that target formed part of a critical infrastructure and finally the population density:

[Risk = Threat + Infrastructure_Criticality + Population_Density].

Initially, risk computations were performed in an additive manner. In other words each of the factors in the equation was assumed to be of equal importance. Today this has been further amended to support risk-based computations that are built on a weighted product of threat, consequence and vulnerability:

[Risk = Threat * f(Vulnerability, Consequence)]

The threat contributes 20% to any risk assessment and is derived from detainee interrogations and other intelligence sources. The product of vulnerability and consequence together make up the remaining 80%. These, in turn, are derived from a measure of the economic importance of any potential target, the population density and a measure of the national strategic significance of any infrastructure elements [3].

Chapter Two has identified problems that complicate the use of risk assessments in guiding military strategy. For example, hazard analysis is often focused on previous threats. In consequence, we are often best prepared to fight the previous war and not the next. Similar concerns can be raised about the application of risk assessment techniques within Homeland Security. Terrorists are arguably less likely to attack a previous target even though most civil forms of risk assessment would associate an increased likelihood with hazards that have occurred in the recent past. A host of further paradoxes affect the risk-based allocation of funds within a counter terrorism portfolio:

> “Since its inception, DHS’s risk-based formula for distributing funds to state and local communities has been a source of frustration for members of the federal, state, and local governments and those who assess post-9/11

> counterterrorism program implementation efforts. Some homeland security observers suggest that it is unrealistic to expect grant levels to continue to increase as U.S. budget concerns weigh on future appropriations. Others might note that as at-risk jurisdictions continue to shore up previously known vulnerabilities they will require less federal funding due to a lowering of their risk profile" [3].

This section has identified a number of similarities between the strategic use of risk assessment for resource allocation in homeland security and in military planning. In both domains, civilian techniques have been extended from banking, insurance and engineering to support procurement decisions in domains that also have to consider the active threat posed by insurgents and terrorist organizations. Concepts such as hazards, likelihood, consequence and vulnerabilities provide a common framework around which it is possible to structure complex decisions in a manner that is transparent to most stakeholders. However, the strategic use of risk assessment techniques extends common weaknesses to both military planning and homeland security. It is often impossible to derive objective measures for any of the values that are required in order to perform risk calculations. It is also difficult to determine the extent to which any subsequent investments in increasing civil resilience or military procurement will reduce the risks of future attacks.

10.1.2 Tactical Similarities

Just as there are strong similarities between the strategic use of risk assessment in military planning and in homeland security, further parallels can be drawn at a tactical level. Improvised Explosive Devices (IEDs) remain a significant threat both to military operations and to civil protection. These weapons caused approximately 60% of all American combat casualties in Iraq. In Afghanistan, they have been responsible for 50% of US combat casualties [7]. IEDs have led to more than a seven fold increase in the number of International Security Assistance Force (ISAF) personnel wounded between 2007 and 2010. In consequence, a number of training programs have been developed to reduce the risks to military personnel:

> "The integration of IEDs into theater immersion focuses on a variety of areas to include interdicting an IED far in advance of its use; force protection measures to keep Soldiers protected; cultural immersion to readily gain valuable intelligence; pattern analysis to identify areas and times of risk; methods to reestablish control and shape the battlefield; and battle drills to close with and destroy the enemy after an attack is launched" [8].

IEDs remain a weapon of choice not only for attacking military targets but also for making political statements and for attracting media attention. The rising threat posed by asymmetric warfare parallels the increasing focus on the threats posed by terrorism for homeland security. It should not, therefore, be surprising that the UK has adopted a risk-based approach to counter IED (C-IED) training amongst the owners and operators of commercial targets, including shopping malls, factories and entertainment venues. The UK National Counter Terrorism Security Office (NaCTSO) created Operation Argus to persuade retailers that even though the threat from the IRA had diminished, they continued to be potential targets for future terrorist attacks. NaCTSO have stressed within Argus that "the responsibility for the implementation of protective security measures following a vulnerability and risk assessment may fall on an individual site owner/manager, an area manager or business development manager with a security remit within a larger organisation" [9]. Such initiatives are important in raising awareness about potential threats and existing vulnerabilities. However, there is limited scope for intervention by retailers and shop owners. It can also be difficult for intelligence and security agencies to schedule time to meet every potential target. In consequence, a proportionate approach had to be adopted in which the level of expected engagement was intended to reflect the likelihood of any threat. The principle mechanism for supporting such risk assessments is through the Vulnerability Self Assessment Tool (VSAT). This provides a tactical extension of the strategic risk assessments, described in the previous section, that the Department of Homeland Security have used to direct their counter-terrorism budgets:

> "… [VSAT] consists of 33 questions and should take no longer than 30 minutes to complete and will provide you with an understanding of what you need to put in place thereby reducing your vulnerability to a terrorist attack. The online assessment will also produce a report of recommendations. In order to protect your business interest and commercial sensitivities the report will only be available to the individual user. The overall assessment result will be displayed without showing numerical values but will assist you in prioritising where you may need to make improvements". [9]

Previous sections have identified strong similarities in strategic use of risk assessments to guide military planning and counter terrorism programmes for homeland security. This section has identified further tactical similarities, risk assessments have been used to inform C-IED training in both domains. The following section extends this analysis to consider further parallels at an operational level.

10.1.3 Operational Similarities

A number of factors make it difficult to assess the risk posed by IEDs in domestic terrorist attacks and in counter insurgency operations. One of the difficulties is the ubiquitous nature of these devices and the ease with which components can be obtained. Many components, especially microelectronics, can be reused from a range of consumer devices. At the same time, informal information exchange networks trade information about the composition and effectiveness of IEDs. Most but not all of these groups use Internet technologies, for instance using mobile data services, that cannot easily be suppressed by national security agencies.

The exchange of instruction manuals as well as operational feedback, including videos of successful attacks, helps terrorist organizations to rapidly evolve their operational planning in the face of strategic and technological countermeasures. This enables the perpetrators to exploit feedback mechanisms that closely resemble those employed by military lessons learned systems discussed in Chapter Two. This creates a form of conflict in which each side attempts to learn from the other in order to reduce the risks to their own personnel and maximize the harm to their adversaries.

Traditionally, counter terrorism initiatives have focused on the identification of disaffected groups as well as the development of detection and jamming systems. Few security services believe that they will always be able to disrupt terrorist groups before an attack can take place. Similarly, technological countermeasures offer limited mitigation for the risks posed by IEDS. The development of jamming devices has led to the increased use of suicide bombers and to the use of decoy devices in multiple coordinated attacks. There is now an increasing recognition that we cannot address individual aspects of the problem in isolation – hence detection and disruption of devices must be supported by initiatives to mitigate the consequences of successful attacks.

10.2 Risk-Based Approaches to C-IED Operations

In order for ‘risk-based approaches’ to be successful, it is critical that we learn as much as we can about previous IED attacks. By identifying common patterns, it is possible to develop scenarios that can be used in planning for the detection, disruption and mitigation of future hazards. The remainder of this chapter uses a generic development model to structure a risk-based approach to military and civil C-IED operations.

10.2.1 The IED Development Model

Figure 15 illustrates some of the stages that have been observed in IED attacks. This describes a cycle in which the dissemination and publication of reports about ‘successful’ detonations may help to recruit further attackers. As we have seen, however, there is no straightforward relationship between the frequency of previous threats and future attacks. Subsequent sections will show that many terrorist groups deliberately vary their modes of attack to exploit perceived vulnerabilities. They will innovate in ways that undermine attempts to use previous incidents as a means of anticipating future threats. However, by studying common phases in earlier attacks, including those areas where terrorist groups have altered their pattern of attack in response to additional security measures, it is possible to anticipate the likelihood and consequences of future attacks. Subsequent sections will also describe how risk-based simulations can be used to identify a range of future scenarios that depart from the specific details of previous incidents.

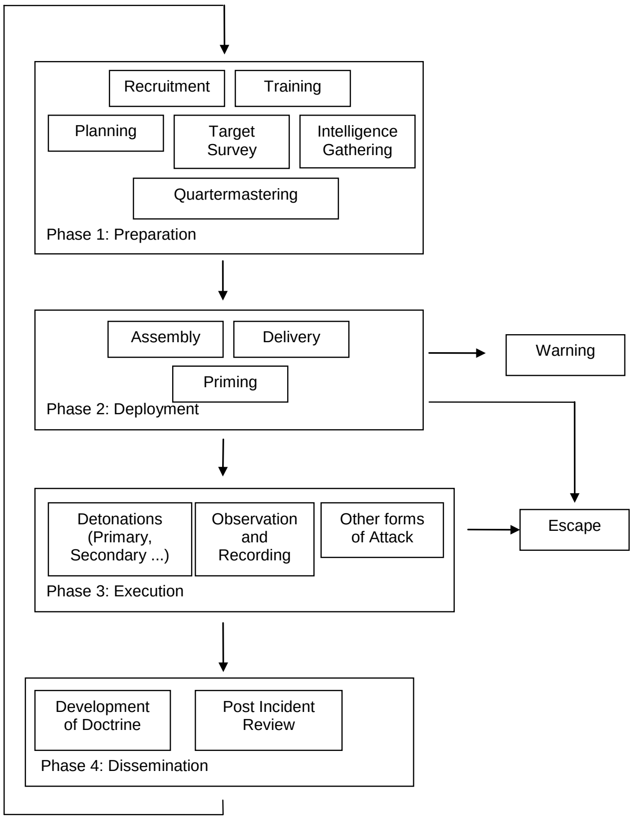


Fig. 15. The IED Development Model

Previous chapters have identified the difficulty of deriving objective estimates for the likelihood of different modes of attack in conventional military operations. In C-IED operations, these problems are complicated by the diversity that characterizes each phase of Figure 15. For example, the hazards posed during the execution phase are determined by the scope of the attack. IEDs have been used against individuals, groups and increasingly entire districts. There is often an inverse relationship between the consequences of an attack and the likelihood of detection. For instance, attacks against individuals may be harder for security forces to detect than those aimed against larger targets, although this is not always the case. Further diversity arises when IED attacks can be planned by a single disaffected individual through to regional and international conspiracies. They can exploit high-technology components without metal parts or they may involve crude adaptations of simple pipe bombs.

The diversity of IED attacks makes it hard to determine the best way of allocating C-IED resources to disrupt each stage of the development process. Further complexity stems from the partial ordering of the activities inside each phase of Figure 15. From preparation through deployment to execution and dissemination, there are a host of more detailed activities that can occur in parallel or in any number of different sequences.

It is difficult to predict the likelihood of an attack because many of the activities in the generic IED development model have benign explanations. Security forces cannot easily distinguish information gathering or target reconnaissance from tourism, an interest in photography etc. The previous chapter has argued that military commanders must balance possible threat against the opportunity to build trust and confidence when troops interact with local populations in counter insurgency operations. Similarly, security and police forces must balance the need to disrupt potential terrorist attacks against the alienation that can occur when arrests are not followed by successful prosecutions. Further problems arise in determining whether any terrorist activity forms part of a wider conspiracy, in which case it may be necessary to delay arrests until more information is gathered about the wider members of a terrorist group. The interval between the initial detection of a threat and the decision to act will inevitably place both the public and officers at a potential risk as more evidence is gathered.

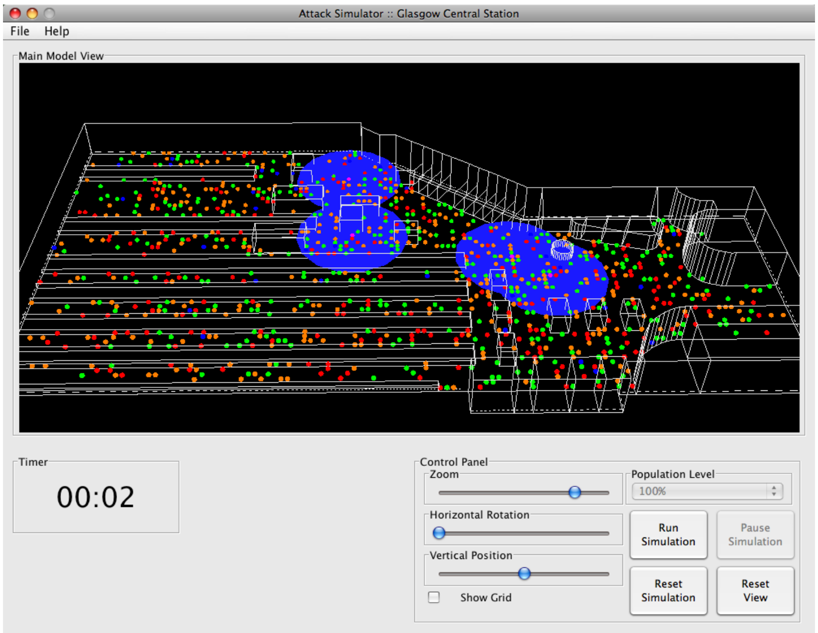
It is equally difficult to anticipate the consequences of a potential threat. Many police forces now monitor dozens of disaffected individual and groups. Most of these are alienated from wider aspects of society. In some cases, criminal offences may have been committed by, for example, incitement to engage in terrorist acts. However, there is often very little evidence of direct involvement in the deployment or execution of an attack. In such circumstances, security forces risk further alienation by preemptive arrests with little likelihood of conviction for more serious offences. Technological and tactical innovations also make it difficult for security and police forces to anticipate the scope of an attack when there is more evidence on active threat. For example, the detonation of a primary device may be used to lure members of the public or emergency service personnel into secondary explosions. The events in Mumbai have also shown that IED s can be integrated into more sustained assaults using conventional weapons. 9/11, as well as the London and Madrid bombings illustrate the difficulties associated with consequence assessment for the range of terrorist attacks that are being planned by dissident groups. The Head of Mi5 recently issues a statement in which he argued that “Risk can be managed and reduced but it cannot realistically be abolished and if we delude ourselves that it can, we are setting ourselves up for a nasty disappointment”. The remaining sections of this chapter justify this argument.

10.2.2 Using Previous Hazards to Simulate Future Attacks

Much attention has focused on reducing risk in the first and second phase of the IED development trajectory. The intention has been to detect recruitment activities or to deploy technological countermeasures immediately after the delivery and priming of a device. However, a key argument in this chapter is that risk-based approaches to counter-IED programmes should take a broader view given the relative difficulty of preventing recruitment and the limited success in the deployment of electronic countermeasures in many areas of conflict. One way of doing this is to study the wider development trajectory of previous attacks; from planning through execution to the response of emergency personnel and the subsequent innovations introduced by both security agencies and terrorist groups.

Counter terrorism risk assessments cannot simply be based on backwards looking studies of previous attacks. In order to anticipate future threats, it is important that we can identify vulnerabilities before they are exploited. This forward looking perspective is critical if we are to avoid the ‘failure of imagination’ that was referred to by the 9/11 Commission [10] and the Intelligence and Security Committee investigation into the London bombings [11].

One way of reducing this element of surprise is to study attack patterns in other countries.Figure 16 presents the interface to computer simulations that have been developed to identify what could happen if IED tactics were transferred from Iraq or Afghanistan to attack the civil population around the globe. This particular example is based on the busiest railway station in the UK outside of London, with peak weekday occupancy of more than 15,000 people. Each year a table top exercise is held. This involves more than 50 staff from the station, transport police and the train operating companies. The exercise is designed to prepare for possible attacks and to help refine the procedures in place for dealing with them. The intention is that this tool can be used by staff to support these annual exercises, for instance, by working through the inter-agency response to a range of different scenarios. In this instance, suicide bombers can be identified by the circles that represent the potential targets caught in any blast. The number of people who might be injured changes for each bomber as they and the other passengers move throughout the station concourse in real-time. The size of the blast and fragmentation areas can be varied to allow for larger and smaller devices given the type of explosive used.



Timer 00:02

Population Level 100%

Fig. 16. Interface to an IED Simulation (Ack: L. Nygaard Nilsen)

An important aim behind the use of these simulations is to reduce the consequences of an attack by helping security agencies apply insights gained from previous IED incidents. In order to do this, it is important to learn as much as possible from the large number of attacks that have occurred in different parts of the world. For instance, the scenario shown in Figure 16 is based on two previous incidents. The first involved a suicide attack on Mustansiriyah University in Iraq during January 2007. A car bomb was detonated at one of the two entrances to the site. This led to a partial evacuation that drew crowds to the other exit where a suicide bomber detonated a secondary device. This is not an isolated incident. Hours before, a second coordinated attack took place in a second hand motorcycle market in the Shia Bab al-Sheik neighborhood of Baghdad. The first blast drew onlookers and the emergency services, who were then hit by a second explosion moments later. The use of advanced computational techniques, including software simulations, helps to drive a risk-based approach to C-IED operations that is based on the consequences of previous attacks but which also considers the likelihood of more innovative attack patterns.

10.3 The Challenge of C-IED Risk Assessment

It is important not to overlook the caveats that affect the successful application of innovative approaches, such as the use of risk-based simulations. It can be difficult to sustain public support if this or similar techniques helped to reduce the likelihood or mitigate the impact of terrorist attacks. ‘Success’ can encourage a sense of complacency that underestimates the scale of the threat faced by security forces in many areas of the world. It is for this reason that we must continually study global patterns in IED attacks.

10.3.1 The Global Nature of the Problem

A number of problems frustrate attempts to assess the risks posed by IED attacks. There are obvious logistic challenges; in order to anticipate possible patterns of future attacks it is important to gather as much information as possible about innovative techniques being deployed in existing civil and military conflicts. It is difficult to obtain information about attacks that occur in remote areas of the world. This, in turn, makes it difficult to anticipate the ways in which particular techniques might be propagated to other regions. For instance, it is possible to obtain a post incident report into the January 2009 suicide car bomb that set fire to a tanker in Kabul; killing four civilians and an American soldier. One reason for this is that it occurred close to a US base and the German embassy. Far less is known about a similar attack that occurred only days later when a civilian died and six people were injured in Nangarhar province. The operational constraints of ‘insurgent’ areas make it far more difficult to conduct detailed investigations, especially in areas outside the Afghan capital. This lack of information frustrates attempts to anticipate patterns that might characterize future attacks.

Further problems stem from the number of IED attacks that occur each year. At almost the same time as the two Afghan blasts mentioned above, an African Union peacekeeper was killed and another injured by a roadside bomb in Mogadishu, Somalia. Less information is known about the tactics and technology used in this incident because African Union forces lack the resources to conduct the same level of examination as the US military. The global nature of the problem can also be illustrated by a series of attacks that all took place within a four week period. Anti-government Thai militants killed one person with a bomb buried in a roadway. Four people were wounded. Some three weeks later, an IED killed twenty people and injured eighty more at a bus stop in Sri Lanka. A police officer was killed by an IED close to the Venezuelan Chamber of Commerce in Caracas. In another incident, two members of the Indian security forces were killed and several others wounded by an IED. There were more than ten IED attacks in India alone within a four week period in 2010. The frequency of these incidents, together with the geographical distribution frustrates attempts to create an international database similar to those that are used to collate ‘lessons learned’ and inform risk assessments across the aviation safety community. There are profound differences in the quality of information that is available to intelligence agencies following IED attacks. In consequence, we must gradually identify common patterns that reflect changing tactics and technologies from partial accounts of a subset of all the IED attacks that occur across a wide range of conflicts.

10.3.2 Accounting for Many Different IED Technologies

Simulators, such as that illustrated in Figure 16, can help security services train for IED attacks. The flexibility of software simulation tools creates new challenges in selecting appropriate scenarios from many different attack patterns. Risk assessment techniques help to map out different threats. For example, software simulators can be used to model high probability attacks, similar to those that have been attempted in previous incidents in neighboring states. They can also be used to simulate more innovative or novel attack patterns, based on intelligence reports from other areas of the globe.

Simulations can also be used to prepare for scenarios with very different consequences, ranging from relatively small pipe bombs through to the deployment of multiple, synchronized vehicle-based devices. This diversity can be illustrated by the range of devices being used in Iraq. Many of the IEDs deployed in this country are formed from simple platter charges. They are constructed from several kilograms of plastic explosive pressed into a similar mass of flat metal, typically steel. This will propel the platter into a target with an approximate velocity of 1,800 m/s at up to 50m. For other targets, Explosively Formed Penetrators have been deployed. In these devices, the force of a blast helps to form a penetrating projectile that can be effective more than 80m from the target. Cylindrical shaped charges can be tipped with a concave metal disc, usually copper. Variants on this type of device have been successfully deployed against Abrams M1A2 tanks. US Army field manual FM20-32 provides a useful starting point for the development of counter terrorism simulators because it provides an initial taxonomy for improvised explosive devices. It distinguishes between high-explosive, artillery-shell antitank devices, platter charges, improvised Claymores, grapeshot antipersonnel devices and barbwire antipersonnel devices. FM20-32 focuses on devices that have been used against organized military units. All of these IEDs have also been used on civil populations in different parts of the globe.

10.3.3 Assessing the Risks from Different Explosives

The threat posed by an IED is, in part, determined by its power. This, in turn, can be characterized by the blast and fragmentation that it produces. These parameters are determined by the quantity and quality of explosive. In some areas, terrorist and insurgent groups must improvise ‘home brew’ explosives from off the shelf ingredients. However, the risks posed by IEDS have been significantly increased when devices are constructed from military munitions stolen from supply lines. For instance, Chinese and North Koreans forces massively underestimated their need for mines in response to the defensive tactics used by UN forces during the Korean War. They, therefore, improvised a series of ‘battlefield devices’ many of which relied upon mines that had been lifted from UN positions. The same techniques were also widely employed by the Viet Cong during the Vietnam conflict [12]. 33% of U.S.

casualties in Vietnam were caused by mines including IEDs that used trip wires and rubber bands to detonate grenades [13].

The reuse of munitions in IEDs illustrates the 'systemic nature' of the threat. Rather than focus narrowly on technological countermeasures once a device has been planted, in many cases risks can be reduced by securing supply lines. This point was recognized in a recent report by the US Government Accountability Office. They argued that the DoD planning for Operation Iraqi Freedom had incorrectly assumed the Iraqi army would rapidly be convinced to provide security for their stockpiles of conventional munitions once they had surrendered. In consequence, a large number of conventional munitions were 'looted'. These munitions were subsequently used in the majority of IEDs deployed against allied forces. The GAO concluded that "…DOD's actions generally have emphasized countering the use of IEDs by resistance groups during post-hostility operations… GAO also concludes that this situation shows both that Iraqi stockpiles of munitions may not be an anomaly and that information on the amount and location of an adversary's munitions can represent a strategic planning consideration for future operations. However, without joint guidance, DOD cannot ensure that Operation Iraqi Freedom lessons learned about the security of an adversary's conventional munitions storage sites will be integrated into future operations planning and execution" [14]. These problems are not isolated to Iraq or Afghanistan. TNT and the C4 compound were used in the 2008 attack on the Islamabad Marriot.

10.3.4 Assessing the Risks of Large Scale Attacks

In many areas of the world, there is a relatively low risk from the use of IEDs based on military grade explosives. However, there is little room for complacency. More than a ton of fertilizer-based explosive was used in the 1992 IRA attack on the Baltic Exchange in the City of London. This killed 3 people and caused £350 million of damage. A similar quantity of 'home brew' compound was used in the 1993 attack on Bishopsgate in the same City, injuring 40 people and caused damage totaling more than £1 billion. This IED was hidden in a construction truck and left a crater more than 40ft wide and 20ft deep. A half ton bomb under South Quay station caused £85 million of damage to London's Docklands in February 1996. A ton and a half of improvised explosive was used against Canary Wharf tower in November 1992, but the detonator failed to ignite the main charge. A slightly larger lorry-based IED injured more than 200 people in Manchester city centre in June 1996. This remains the largest bomb to explode in the UK since the Second World War and was parked under a shopping center some two hours before detonation. It was subsequently estimated that up to 50,000 square meters of retail space and nearly 25,000 square meters of office space had to be reconstructed.

The risk posed by ‘home-made’ explosive charges extends across the globe. Three quarters of a ton of a fertilizer-based compound was detonated in the underground car park at the World Trade Centre in 1993. A more destructive form of explosive was used in the Oklahoma City bombing. The ease with which the two conspirators were able to amass more than 2,300 kg of explosive-grade ammonium nitrate fertilizer and 600 liters of liquid nitromethane is an instructive lesson for security agencies in many countries. Similar compounds were used in the Bali bombings of 2002 that killed more than 200 people and in multiple attacks on US embassies in August 1998 killing 224 people. One of the key insights from this enumeration is the continuing risk that these weapons pose to civil society. The insights derived from the consulate bombings of the 1990s still did not yield enough counter measures to prevent the 2002 attack on the US embassy in Karachi where a truck based fertilizer bomb killed 12 and injured 51.

Software simulations help building occupants and owners to visualize the threats from IED attacks. They help to focus attention by showing them what might happen if they were attacked using the size of devices that were deployed against Bishopgate or the Federal Buildings in Oklahoma. It can be far more convincing to develop training scenarios where the impact of the IED is based on a device that has already been detonated in London or New York rather than Baghdad or Beirut. However, this is not a panacea. Previous IED attacks demonstrate that terrorist groups have achieved considerable success even when stakeholders are forewarned that they are potential targets. Well protected sites have been rendered vulnerable by increasing the force of the device or by changing tactics, for example from delay based detonators to suicide attacks. The 2003, Al Qaida-inspired attack on the British Consulate in Istanbul and the HSBC bank killed 30 people even though security personnel were aware of the risks. Suicide bombers detonated a mixture of ammonium nitrate and fuel oil in pick-up trucks.

10.3.5 Assessing the Risks of Medium Scale Attacks

Large scale devices usually require quartermasters to coordinate the acquisition and storage of materials before an IED can be assembled. This arguably mitigates some of the risks posed by the relatively high impact of these devices; there are more opportunities for security forces to detect the build-up of components. This is another reason why it is so important to study the early stages of the IED development trajectory; to learn how the perpetrators of previous attacks were able to acquire their materials.

Higher risks can be associated with multiple, medium scale IEDs. They are often far harder to detect because they can be improvised from limited quantities of legitimate components with very little prior planning. The unpredictable risks associated with these attacks can be illustrated by attempts to detonate IEDs in London and at Glasgow Airport. The initial plan was to load two cars with gas canisters on the back seats; together with nails and petrol in the trunk. Mobile phones provided improvised detonators. The cars were driven from Scotland to London in June 2008. The first vehicle was parked outside a nightclub. The second was parked a few streets away; with the possible intention of catching people in a secondary blast. This plot clearly differs in scale and the sophistication of the explosives from those described in previous paragraphs. It also illustrates that any attempts to identify potential attacks from the purchase of ammonium nitrate may underestimate human ingenuity. This initial plot was abandoned when 15 calls to the mobile phones failed to trigger the detonation. The attackers then returned to Scotland and rigged up a third vehicle with fuel and gas canisters, petrol and knives. Rather than leaving the vehicle outside a night club, the attackers drove it into the main doors of Glasgow airport where it was wedged against a steel block. This device also failed to detonate, possibly due to the difficulty of ensuring that the mixture of fuel and oxygen was flammable.

The attack on Glasgow Airport provides a further motivation for studying the trajectory of previous attacks to inform risk-based C-IED operations. Prior to this there was a sense of complacency in Scotland. Public and politicians felt there was little risk that we would be the target for a terrorist attack. IEDs were associated with conflicts on the other side of the globe. This attitude faded as soon as the vehicle was driven into the airport. The simulation tool, illustrated in Figure 16, was explicitly developed as part of a wider programme to increase the resilience of Scots infrastructure against the future threat from these devices.

10.3.6 Assessing the Risks to Individuals

IEDs not only pose risks for major infrastructures, such as Bishopgate or the Federal buildings in Oklahoma City. Many devices are specifically intended to kill or maim individuals. An example is provided by the pipe bomb that injured Zeev Sternhell, an Israeli academic and critic of Jewish settlement in the occupied West Bank. Car bombs have also been widely used in many countries, for instance a mercury tilt switch was used to detonate the device that killed Airy Neave, a UK Conservative politician who opposed loyalist and republican paramilitaries in Northern Ireland. Similar devices were used to target individuals who had been opposed to the regime of Augusto Pinochet in Chile.

Mail bombs have a history that is almost as long as the postal service; there are 18th century accounts from both Italy and Denmark. The Unabomber provides more recent examples. His first device was found in a parking lot. The return address was that of the intended victim. The parcel was eventually passed to a security guard who received minor injuries when he attempted to open it. The Unabomber’s early IEDs were relatively crude pipe bombs with wooden end pieces and detonators that pulled a nail across match heads. Later devices replaced this approach with batteries and filament wire, including an IED that was placed in the hold of an aircraft flying within the United States. In the UK, mail bombs have recently been sent to companies involved in DNA testing. A primary school caretaker was eventually arrested and subsequently argued that the small amount of explosives was intended to increase public awareness without risking injury to the public. These arguments were largely dismissed by the court.

In the US, a series of unassembled letter bombs were sent by someone calling themselves ‘The Bishop’ to financial firms in the Midwestern United States. Subsequent investigations have suggested that the individual involved was copying elements of a Charles Bronson film in which an assassin left a note with each bomb. Considerable care had to be taken during the arrest of ‘The Bishop’. Security forces were concerned to minimize the risks to their personnel. There was a concern that the perpetrator might detonate a device during the arrest. Crowd based modeling tools, such as that shown in Figure 16, help planning and training for these police actions. In particular, they can be used to drive training scenarios in which officers must identify a suitable opportunity to make an arrest while minimizing the potential threats both to themselves and the general public.

The fatal shooting of Jean Charles de Menezes by UK police provides a further illustration of the risks that arise during the later stages of the IED development trajectory. The Brazilian electrician was mistaken for a suicide bomber with links to the 21st July attacks on London. The subsequent inquest showed that security services must revise the way in which they plan for the arrest of terrorist suspects. This can be done through the use of simulations that recreate the flow of information between intelligence services. They can recreate the problems in information exchange that characterize real-world operations rather than the ideal situation that is often portrayed in Standard Operating Procedures. The de Menenzes shooting did significant harm to counter terrorism operations across the UK. It illustrates the loss of confidence that undermines C-IED initiatives when innocent individuals become inadvertent victims of security personnel. In the aftermath of such incidents, public attention focuses on the communications problems and on the training limitations that characterizes many C-IED programmes. For senior commanders, it is important to consider the risks of false arrest or of innocent casualties from counter terrorism operations.

One of the problems for security services is that they must act with the same degree of precaution in apprehending an innocent member of the public as they do to an individual who is eventually convicted of a terrorist offence. This can be illustrated by two recent examples from the UK. The first relates to the prosecution of a man who was found not guilty of two charges of making IEDs. During the trial it emerged that army disposal experts found fireworks and ‘thunderflash devices’ in his home. They also found an infrared transmitter that was capable of triggering the detonation of an IED. However, the defense successfully argued that this was used to operate his satellite television and that the defendant had an interest in fireworks from teenage years. In contrast, the second case led to the successful prosecution of a man who was found to be in possession of a nail bomb when bailiffs came to evict him from his house. The army bomb-disposal teams again had to make the device safe before neighbors could return to their homes. In both cases, it was difficult for the security personnel involved to assess the risks that each individual posed both to themselves and to other members of the public.

10.3.7 Accounting for the Diversity of Delivery Mechanisms

The second phase of the IED trajectory in Figure 15 describes the ‘delivery’ of a device to the intended target. This might seem like an elaborate term for a relatively simple act. However, it has a critical impact upon the risks associated with an attack. The availability of delivery mechanisms affects the likelihood of any act; for instance by allowing terrorist groups the opportunity to use vehicle based weapons. The use of particular delivery techniques can also determine the consequences of an attack. Suicide bombers carry a significant threat because they can carry an IED into the center of a crowd; in locations that cannot easily be reached by other delivery mechanisms. For example, thirteen people were injured in November 2008 when an IED was thrown from a flyover into a market in Bangkok. This incident illustrates the diverse nature of the threat from these devices; the incident was not part of an ethnic or political dispute but seems to have been a response to a civil dispute between traders and the market management following a rent increase.

Previous sections have summarized delivery techniques ranging from cars, vans and trucks through to the postal systems that convey letter and parcel bombs. The diversity of delivery mechanisms challenges some of the ‘silo thinking’ that characterizes the immediate response to IEDs in many countries. There is often an immediate response that fails to consider the risks that have been realized in other attacks. Security forces are often too quick to address a narrow number of vulnerabilities without thinking more broadly, for instance about the full range of delivery mechanisms that have been used during IED attacks. For instance, many airports, railway stations and shopping malls have responded to attacks such as the one at Glasgow Airport by pouring vast quantities of concrete to prevent the use of car bombs. At the same time, these facilities are encouraging the use of ‘greener forms’ of transport including bikes. They have increased access to individuals in wheelchairs and to families using child buggies. All of these different forms of ‘transport’ have recently been used to deliver IEDs. For instance, one person was killed and four people were wounded by a device hidden inside a bike that was detonated in India during February 2008. Key benefits of this form of IED is that steel tubing both hides the explosive and provides fragmentation materials.

One of the biggest factors in determining the risk to the public is whether an attacker is prepared to commit suicide in order to deliver their device. Figure 15 shows this in the trajectory model through several different stages at which perpetrators might attempt to escape detention. For example, the IRA developed a range of remote triggering techniques to support their use of vehicle bombs against the City of London.

Many groups have developed suicide attack methods to avoid the risk of jamming by security personnel. These methods also avoid the technical complexity involved in manufacturing reliable remote detonation devices, given the risks of disclosing information about the bomber's identity when devices fail to explode. For instance, three recent blasts were attributed to the Islamist insurgency in Algeria. A car containing an IED was driven into a police college in Issers, killing almost 50 recruits waiting for an exam. Within twenty-four hours another two car bombs were detonated near a barracks in Bouira. Suicide bombs have been used in countries as diverse as Turkey, where 6 people were killed and 90 injured in Ankara in May 2007, and Pakistan, where the 2008 attack on the Marriott Hotel in Islamabad killed more than 50 people and injured more than 200. In Vladikavkaz, the capital of the North Ossetia region between Russia and Georgia, eight people were killed in November 2008 by a female suicide bomber outside a busy market. The device was detonated as a minibus arrived at a bus stop. Numerous other examples can be cited from the conflict with Chechnya.

In contrast to these relatively primitive delivery mechanisms, it is likely that the transfer of IED design techniques will continue to influence future delivery mechanisms – for instance, through the development of rocket based devices similar to those being fired into Israel. Hezbollah have used Katyushas from former Soviet and Chinese stockpiles, such as the Soviet BM-21 Grad missile as well as 'derivatives' from the Iranian Fajr missiles. These delivery systems are not considered in detail here because they are closer to standard military munitions than the majority of 'improvised' explosive devices.

10.3.8 Assessing the Risks of Innovation in IED Technology

A further challenge in assessing the risks of IED attack scenarios is that the technologies used by terrorists and insurgents change over time. In other words, we should never underestimate the role of improvisation in the development of these devices. This can be illustrated by recent blasts in which IEDs were hidden inside ATMs, or cash machines, although these were not programmed to recognize individual PIN numbers. The evolution of new techniques emphasizes the need both to learn and extrapolate from previous attacks around the globe.

The increasing risk associated with the rising sophistication of IEDs can be seen in the innovations that occurred during the Northern Ireland 'Troubles'. Molotov cocktails led to clockwork timers with five to ten minutes delay and then remotely controlled devices with anti-handling features, such as tilt switches, that would detonate if attempts were made to defuse or move the IED. For example, the Brighton Hotel Bomb was planted more than twenty days prior to its detonation. This device was constructed using the timer components from VHS video recorders. Other 'innovative' devices were constructed using transceivers and servo motors from model aircraft.

Technical innovation did not cease with the Mitchell peace process in Ireland. Previous generations of pressure pad detonators have been replaced by infrared triggers. IED's have also been developed to exploit GSM and other forms of radio signals, including pulsed transmissions that offer greater resilience to jamming. Security forces have responded by installing electronic counter measures such as the 'Element B' systems. However, these innovations seldom offer complete protection. They can be difficult to install and maintain. It is particularly difficult to ensure that jamming devices are deployed to offer equal protection across all potential targets. This leads to further tension when, for example, allied troops are protected while the same counter measures are not available to local coalition forces. The IED 'arms race' continues not only in the iterative improvement of remote detonation but also in the use of Explosively Formed Penetrators (EFPs) to counteract changes in vehicle protection. The systems approach, advocated in this chapter, stresses that IED risks cannot be considered in isolation from the many tactical changes that have profoundly changed the ways in which these weapons have been deployed in recent months.

10.3.9 Anticipating the Dynamic Refinement of IED Tactics

Many of the tactics used in recent IED attacks were first developed by Hezbollah following Israel's invasion into Lebanon. In the mid 1980s, suicide bombers were used to drive vehicles against their intended targets. However, security forces changed their tactics to reduce the risk from this form of attack. Physical barriers segregated civil traffic from potential targets. In consequence, greater emphasis was placed on the use of roadside bombs planted well in advance of their detonation. This tactic was used in the remotely detonated bomb that killed Israeli Brigadier General Erez Gerstein in February 1999. Since then, Israel has continued to pioneer IED countermeasures. However, they recognize that there can never be complete protection from this form of attack. The building of the Gaza wall illustrates the difficulty of preventing IEDs.

Western security forces have copied many of the counter measures adopted by the Israelis, for instance in segregating potential bombers from their targets. However, the likelihood and consequence of particular modes of attack has also been strongly influenced by information exchange between potential attackers. There are strong suspicions that members of Hezbollah, assisted by Iranian Revolutionary Guards, helped to transfer expertise in the use of IEDs to the local militias that attacked British forces around Basra. These suspicions are supported by the transfer of specific techniques between these conflicts. For instance, Hezbollah developed the use of stacked mines to increase the blast that was needed to destroy Israeli vehicles. The same approach has been used against UK and US forces in Western and Southern Iraq during 2005. There are other parallels in the tactics used to conceal roadside bombs, in particular the use of false rocks and road-kill in both Lebanon and in Afghanistan. Explosively Formed Penetrators or 'shaped charges' have also been used in all three conflicts.

One of the catalysts for the exchange of IED tactics has been the exchange of video footage of previous attacks. Hezbollah quickly recognized the propaganda impact of filming their work. This raised awareness of their operations and may also have helped recruit additional support. However, the videos had further uses; they were included in training manuals and were studied to improve subsequent tactics. These developments reiterate the importance of systems, risk-based approaches to C-IED operations. Not only must security agencies focus on countermeasures and the detection of present threats, they must also consider the impact that such documentation and video footage can have upon the shape of future threats. In particular, a detailed analysis of Internet video footage might provide scenarios for simulations, such as that shown in Figure 16. Films of previous attacks can inform the training of security personnel just as it presently informs the training of future bombers.

10.3.10 Assessing the Risks of Multiple Coordinated Attacks

Previous sections have described the increasing threat posed by the use of coordinated IEDs. Terrorist and insurgent groups have learned that multiple simultaneous attacks carry a greater risk than either one large device or a series of isolated detonations. One of the early examples of this was provided by the coordinated attack on US Embassies perpetrated by Al Qaida during August 1998. 224 people were killed by bombings in Nairobi, Kenya, and Dar es Salaam, Tanzania. These attacks illustrate the importance of being able to extrapolate from previous attacks – they are widely recognized as precursors not just of the London and Madrid bombings but also of the 9/11 attacks.

In retrospect, it is difficult to argue that security personnel could have used the embassy explosions to predict subsequent attacks in Europe and North America. However, official reports into all these incidents have made the point that it is precisely this ‘leap of imagination’ that we should encourage in homeland security. It is possible to identify other emerging patterns that might provide precursors to future attacks. For example, the opening sections of this chapter explained that the simulation tools in Figure 16 were based on the coordinated use of IEDs in Iraq. In several previous incidents, a primary car bomb was detonated before suicide bombers used secondary blasts to target the crowds that gathered after an initial explosion. This pattern can also be seen in the 2002 Bali bombings; a suicide bomber first triggered a backpack device in a bar. The crowds that then fled from the scene of this first blast were caught by a secondary fertilizer-based IED hidden in a van.

Further variations on the coordinated use of IEDs have emerged from the Mumbai attacks in December 2008. Ten gunmen fired at a number of points in India’s largest city over a 60 hour period. IEDs were not the primary weapons used; however, they did play an important role. Two devices were found in the wreckage of the Taj Mahal Palace hotel – Police have not disclosed the details but they did comment on the relative sophistication of their construction, especially of the timing devices.

Following the Mumbai attacks, security agencies conducted a sweep of Chhatrapati Shivaji train station and declared it to be safe. However, several days later IEDs were found amongst lost luggage. The public again had to be cleared from the building. It is, therefore, possible to identify several different patterns in the coordinated use of IEDs – these include:

- the near simultaneous attacks in different countries;
- simultaneous attacks across the transportation or other infrastructures in the same country;
- the coordinated use of suicide bombers and vehicle based devices to draw crowds into secondary explosions;
- the use of armed attacks in conjunction with IEDs that may then be used to target security forces etc.

It is clear that most local security agencies have only begun to consider a very limited subset of the scenarios that have already been witnessed in other areas of the globe. This has significant and pressing implications for future risks to the general public.

10.3.11 Predicting the Impact of Warnings and Hoaxes

Previous sections have described a series of challenges that complicate the use of risk assessment techniques within C-IED programmes to help identify future attack scenarios. A key theme in this work has been to use a systems model covering diverse phases in the preparation of an IED through to deployment, execution and dissemination for different patterns of attack [15]. Technical innovation continues to increase our ability to counteract the masking techniques used to disguise IEDs prior to detonation. However, sensing systems are still limited in their range and by the costs both of installing and maintaining them. They also create significant overheads when security personnel are forced to respond to a large number of ‘false hits’. These insights are illustrated by the five million security alerts that were logged during the 16 days of the Turin Winter Games, a figure that was exceeded in Beijing [16]. The limited precision and recall of automated sensors make it likely, therefore, that most C-IED operations will continue to depend upon intelligence provided by the public.

It is important not to overlook the opportunities for risk reduction that are offered by the warnings, which are often issued by the perpetrators of an attack. The intention behind these warnings is either to reduce public casualties or increase injuries sustained by the emergency services. However, a study of previous attacks reveals how security personnel often miss the opportunity to mitigate the consequences of IED attacks. For example, the 911 operator who received the warning about the pipe bomb in Centennial park during the Atlanta Olympic Games could not dispatch a response team because she could not enter ‘Centennial’ into her computer system. This had not been updated with the new names given to major venues as part of the preparations for the Games [17]. The operator was eventually put on hold for two minutes while the Command Center began asking for the street address of the Park. In the meantime, members of the public had reported a suspicious bag. Officers on the scene were reluctant to broadcast a warning in case panic ensued. Police teams reached the Park just as the device exploded.

Just as important as learning the lessons from previous incidents, is the need to inform our future response by studying previous hoax calls. These can increase the risks associated with IEDs by wasting security resources. More than 100 reports of suspicious packages were made in the 24 hours following the explosion in Centennial Park. All proved to be harmless. These incidents placed immense stress on the police and other security agencies. The paradoxical effect of increasing public awareness was that the sheer number of false alarms may have created opportunities for subsequent malicious acts. It is important not to underestimate the impact of these calls. For instance, one report led to the closure of the 'Underground Atlanta' shopping mall. Thousands of people had to be evacuated during the evening following the bombing. Although the subsequent search lasted less than an hour, the evacuation caused considerable traffic problems. The mall was adjacent to the Five Points interconnection for Atlanta's MARTA rapid transit system. Thousands more people were affected when this main north-south and east-west transfer point was closed. The package turned out to be a clothes iron.

It is important also to consider the impact of warnings because they can misdirect emergency personnel and, thereby increase the consequences of IED attacks. For example, a warning was issued some forty minutes before the Omagh bomb exploded. This was ambiguous and Police began clearing the wrong area. Instead members of the public, including women and children, were directed towards the bomb.

10.4 A 'Systems' Perspective on C-IED Risk Assessment

This chapter has argued that a 'systems' approach can address the threat to public safety from Improvised Explosive Devices (IEDs). Rather than focusing narrowly on electronic counter-measures or on the detection of disaffected groups before an incident, we have argued that security agencies should mitigate risks across all stages of the IED trajectory. Figure 15, therefore, enumerated different phases from the preparation of a device through deployment, execution and the dissemination of propaganda and operational insights following an attack. These phases were then used to structure an analysis of previous incidents, borrowing a 'lessons learned' approach from safety engineering, similar to that described in Chapter Three.

This 'systems' approach helps to identify patterns of attack. These, in turn, can be used to assess the risks of future incidents. However, it is clear that this approach cannot provide a panacea for C-IED operations. Numerous logistic problems frustrate the transfer of intelligence about previous attacks. It is difficult to gather information about tactics pioneered in remote regions or in areas that are not under the control of friendly forces. This enables terrorist groups to refine their attacks in a way that cannot easily be monitored by intelligence and security forces. Further problems arise from the speed with which terrorist groups can exchange information about previous tactics. There is also evidence to suggest that many terrorist groups also exploit 'lessons learned' techniques in which video footage supplements debriefing reports in order to assess the effectiveness of an operation.

More fundamental objections can be made against risk-based C-IED operations. There is a considerable danger of preparing for previous terrorist tactics and not for those that will be used in the future. Too often, security forces have focused on the threat from isolated car bombs or of individual suicide bombers. This has created considerable vulnerabilities to the coordinated and mixed-weapon attacks seen in Bali and more recently in Mumbai.

The task of assessing the likelihood and consequences of IED attacks is further complicated by the rapid pace of tactical and technical innovation. Initial assessments of the threat posed by particular strategies must be continually revised as terrorist groups refine the techniques that they use. For instance, many of the tactics pioneered by Hezbollah and by Iranian forces have now been exploited in other regions. This analysis of dynamic civil threats builds on previous chapters of this book, which have argued that military risk assessments must continually be revised in order to reflect technical and tactical innovation by enemy forces.

One of the problems in preparing for future IED attacks is that it can take weeks or months to prepare drills and exercises. Hence, they cannot easily respond to rapid innovations in the methods being refined through IED attacks around the globe. In this chapter, we have argued that software simulation tools provide a flexible means of anticipating future threats. These systems enable security and intelligence personnel to alter the tactics and technologies deployed in particular attacks. For example, the size and composition of a device can be configured in the software to model different consequences in terms of the fragmentation and blast that would be produced by any detonation. These models can also be configured with the layout of local buildings, for instance using the 3D models that are increasingly being developed by architects during the construction and alteration of major buildings. Crowd simulations and models of emergency personnel can also be used to explore other aspects of the IED trajectory. A key point here is that these systems are not, typically, intended to provide accurate predictions of future high-risk IED attacks. In contrast, they are intended to help emergency personnel and security services train for a wide range of attack scenarios with a level of flexibility that is difficult to reproduce in conventional drills. There is a strong parallel here with the use of simulation to prepare pilots and co-pilots for a range of adverse events within the field of aviation. There is no guarantee that they will meet exactly the same scenarios that they face during their training. However, the simulations are intended to help aircrew rehearse their response to unexpected and challenging situations.

It is important to stress that the use of software simulations cannot be separated from the use of risk assessment techniques across the IED development trajectory. Unless we can assess the likelihood, vulnerability and consequences of particular attack methods then it will be difficult to identify the scenarios that are to be modeled and then permuted in the software models. It is also important to repeat the argument made throughout this book that risk assessment is not an end in itself. It forms one small part of a wider strategy for C-IED operations. Unless we integrate these techniques into appropriate simulation tools and intelligence operations that are sensitive to the development of new tactics and techniques then future thinking will be dominated by past events. There is a danger that we will again be victims to the 'failure of imagination' that was criticized by the 9/11 Commission and subsequent investigations into the London bombings.

10.5 References for Chapter Ten

[1] Thomas Dempsey, Counterterrorism In African Failed States: Challenges And Potential Solutions, Strategic Studies Institute, U.S. Army War College, Pennsylvania, USA, April 2006.

[2] R.J. Erickson, Legitimate Use of Military Force Against State-Sponsored International Terrorism, Technical Report, US Air Force, The Air University, Maxwell Air Force Base, Alabama, USA, July, 1989.

[3] T. Masse, S. O'Neil, J. Rollins, The Department of Homeland Security's Risk Assessment Methodology: Evolution, Issues, and Options for Congress, Congressional research Service, Washington DC, USA, February 2007.

[4] H. H. Willis, A.R. Morral, T.K. Kelly, J.J. Medby, Estimating Terrorism Risk, The Rand Corporation, Center For Terrorism Risk Management Policy, 2005.

[5] Department of Homeland Security, DHS Awards $844 Million to Secure Nation's Critical Infrastructure, May 16, 2008, Washington D.C., USA. Available from http://www.dhs.gov

[6] Department of Homeland Security, About the Office for Bombing Prevention, Available on http://www.dhs.gov/xabout/structure/gc_1184010933025.shtm, last accessed August 2010.

[7] Congressional Research Service. Improvised Explosive Devices (IEDs) in Iraq and Afghanistan: Effects and Countermeasures (Washington, D.C.: Aug. 28, 2007).

[8] D.L. Zajac, B.A. Bissonnette, J.F. Carson, The First Army IED Training Methodology, Infantry, 43-46, Training Notes, July-August 2005

[9] UK National Counter Terrorism Security Office (NaCTSO), Vulnerability Self Assessment Tool (VSAT).
Available on: http://www.nactso.gov.uk/OurServices/VSAT.aspx, last accessed August 2010.

[10] 9/11 Commission, Report of the 9/11 Commission, Washington D.C., USA. Available from http://www.9-11commission.gov/report/911Report.pdf

[11] Intelligence and Security Committee Report into the London Terrorist Attacks on 7 July 2005, Report Cm 6785, Her Majesty's Stationery Office, Norwich, UK.

[12] H.E. Dickenson, Chief of Staff, First Marine Division (Rein), Division Order P3820.2A, Standard Operating Procedures for the First Marine Division: Countermeasures Against Mines and Booby-Traps (San Francisco: Department of the Army Office, Chief of Engineers, February 1, 1969), p. 1-1, as republished in Smith, Landmines/Vietnam (1972), p. H-39.

[13] Harry N. Hambric and William C. Schneck, The Antipersonnel Mine Threat: A Historical Perspective, Symposium on Technology and the Mine Problem, Naval Postgraduate School, Monterey, CA, November 18-22, 1996, p. 15.

[14] Operation Iraqi Freedom: DOD Should Apply Lessons Learned Concerning the Need for Security over Conventional Munitions Storage Sites to Future Operations Planning, GAO-07-444 March 22, 2007, http://www.gao.gov/products/GAO-07-444.

[15] C.W. Johnson and L. Nilsen-Nygaard, Extending the Use of Evacuation Simulators to Support Counter-Terrorism: Using Models of Human Behavior to Coordinate Emergency Responses to Improvised Explosive Devices. In R.J. Simmons, D.J. Mohan and M. Mullane (eds), Proc of the 26th Int. Conf on Systems Safety, Vancouver, Canada 2008, International Systems Safety Society, Unionville, VA, USA, 0-9721385-8-7, 2008.

[16] C.W. Johnson, On the Convergence of Physical and Digital Security for Public Safety at Olympic Events. In R.J. Simmons, D.J. Mohan and M. Mullane (eds), Proc of the 26th Int. Conf on Systems Safety, Vancouver, Canada 2008, International Systems Safety Society, Unionville, VA, USA, 0-9721385-8-7, 2008.

[17] C.W. Johnson, Using Evacuation Simulations to Ensure the Safety and Security of the 2012 Olympic Venues, Safety Science, (46)2:302-322, 2008.

11 The Future of Military Risk Assessment

Military operations inevitably carry greater risk than most civilian occupations. In consequence, 'lessons learned' systems have been established to identify safety improvements from previous mishaps. However, accident and incident rates remain stubbornly high amongst many armed forces. One reason for this is that in order to provide practical benefits, 'lessons learned' systems must be integrated into decision-making and planning processes. Previous initiatives have failed to yield safety improvements because accident and incident reports simply reiterate problems that are already well understood without triggering the necessary corrective actions. Short term operational demands, political pressures, internal rivalries all combine to ensure that hazards are not adequately addressed in many armed forces.

This book has argued that risk assessment techniques help to improve the safety of military operations. They guide operational, tactical and strategic decision making. They provide a framework for the allocation of finite resources in proportion to the perceived likelihood, vulnerability and consequences of a range of future hazards. Accident and incident reporting systems help to provide objective data to support these assessments. Simulations and subjective expert judgment can also be used to inform forward looking risk assessments for hazards that have not yet resulted in adverse events.

A number of armed forces have already integrated risk assessment tools into their military doctrine. The UK armed forces have responded to a number of accidents by introducing mandatory risk assessments for certain classes of operation. For instance, eleven UK Puma helicopters involved in major accidents between 2001 and 2007, with three aircraft being lost in 2007 alone. This prompted a strategic review into the operation of these aircraft [1]. One consequence of this was that Operational Risk Management was introduced to ensure that senior commanders formally considered the potential hazards of Puma missions before providing their approval. These techniques have also been used more widely, for example as a result of UK Army publication MMP 201: A Commander's Guide to Health, Safety and Environmental Risk Management/Headquarters Land Command. The US Army has arguably taken this approach further than any other military organization. Previous chapters have described how their Composite Risk Management programme encourages all soldiers to identify the likelihood and consequence of potential hazards both on and off duty.

The operational and tactical introduction of risk assessment has been mirrored by the strategic use of these techniques to support resource allocation and procurement. For instance, The US Department of Defence's Business Transformation Agency has developed the Enterprise Risk Assessment Model (ERAM) to mitigate risks during acquisitions [2]. A 'risk assessment team' spends two weeks reviewing existing project documentation. This analysis then informs a series of more focused interviews with program stakeholders that last from 2-3 days. These meetings determine the future funding and direction of procurement programs. Similarly, the

UK Ministry of Defence has established a joint risk management policy between the Chief of Defence Materiel and the UK's Chief Scientific Advisor. This is intended to ensure that risk management techniques are used across all phases of military procurements from conception through to decommissioning. The Australian Defence Risk Management Framework fulfills a similar role across their armed forces [3]

This book has identified a number of concerns that limit the utility of civilian risk assessment techniques. Military procurement is very different from many other safety-critical applications. There is a constant need for rapid technological innovation in the face of new threats or vulnerabilities. Urgent operational requirements, for example to introduce night vision technology or Unmanned Airborne Systems (UAS), can undermine the most rigorous procurement processes. Political pressures also play a stronger role than might be expected in most other industries. These factors combine to increase the likelihood that some procurement projects will fail to deliver their intended benefits.

Even when risk assessments have guided an initial procurement, it can be difficult to update a hazard analyses in response to changes in the operational context of a system [4]. Military systems are, typically, deployed into complex and unpredictable environments. In such circumstances, operational demands often reveal new threats that had never occurred to the teams who first conducted a hazard analysis. This book has also identified a range of problems that stem from changes in the nature of 21st Century warfare. For example, it is particularly difficult to characterize the risks that arise in counter insurgency operations. Commanders must offset the potential hazards to their troops against the need to support the local population. Providing the level of protection and security that is required during peacekeeping operations will often expose troops to a higher risk of insurgent attacks, for example using Improvised Explosive Devices (IEDs).

Rather than develop a series of high-level theoretical arguments, previous chapters have used a mixture of operational expertise, after action reviews and mishap reports to identify the limitations of risk assessment in military operations. Our analysis raises significant questions about the utility of hazard analysis. For example, US Army Field Manual 3-04.513 states that: "Risk management is a commonsense tool that leaders can use to make smart risk decisions in tactical and everyday operations. It is a method of getting the job done by identifying the areas that present the highest risk and taking action to eliminate, reduce, or control the risk. It is not complex, technical, or difficult" [5]. In contrast, previous chapters have shown that risk management is often complex and technical and difficult. Military decisions are, typically, based on uncertain information. They are taken in the face of fatigue, with time pressures and limited resources.

Many of the limitations stem from the use of techniques that were developed to support decision making in the financial services industry, in industrial process control or in environmental engineering. These industries have introduced regulatory frameworks to ensure that risk assessments are applied by management and technicians with adequate resources and training. Regulators also act to sustain the necessary safety culture within civil applications. Military systems are very different. They lack the external regulatory supervision that safeguards most other industries. They must also adjust to a level of external threat that simply does not exist in civil society. It is seldom possible for military personnel to exhaustively identify all potential hazards or threats. Similarly, there may be vulnerabilities in an armed force that only emerge after an operation has begun. For example, a lack of training, inadequate standard operating procedures or substandard equipment may only be exposed by new mission objectives or by changes in the enemy's tactics and technology. Finally, military forces must respond to political direction in a manner that would only otherwise be acceptable in a command economy.

11.1 The Boundaries of Military Risk Assessment

This book urges caution when using civilian techniques that are largely unproven in a military context. A number of case studies illustrate the mishaps that have occurred when existing hazard analysis methods fail to mitigate the threats that arise in combat and peacekeeping operations. Later sections will argue that these limitations can be eased by the development of techniques that are specifically intended to support military decision making. This vision requires a new attitude to operational research. It depends upon the close integration of decision science with direct military expertise in strategic, tactical and operational decision making. The insights derived from previous chapters help to establish requirements for future generations of military risk management techniques:

- *Military risk management must not be bounded by anecdote.*
  One of the biggest limitations to military risk assessments is the lack of information about previous adverse events. Although, the opening chapters of this book have described a number of formal systems that support the exchange of information about military mishaps, most decision makers continue to rely on anecdote and word of mouth as the principle means of finding out about previous hazards in many operations. This leads to a lack of consistency in risk assessment as different people then make very different estimates about the likelihood and consequences of adverse events.

- *Military risk management must not be bounded by mishap reports.*
  Risk assessments cannot rely upon the information in military incident and accident reporting systems. Underreporting remains a significant problem. This is especially difficult when forces are operating in many different and remote regions of the world. There are also more fundamental issues when risk assessments focus too much on previous hazards and not enough on potential future changes in tactics and technologies that may not be captured for many months within a conventional 'lessons learned' system.

- *Military risk management must not be bounded by subjective bias.* There is no guarantee that characterizing threats/hazards in terms of likelihood, consequence or vulnerabilities will lead to greater consistency. Chapter three considered a host of subjective influences that affect risk-based decision making. These can be exacerbated by the introduction of risk assessment tools. For instance, risk adverse commanders have exaggerated potential hazards in order to secure additional support during military operations.

- *Military risk management must not be bounded by static analyses.* Many of the case studies in this book have illustrated the problems of revising initial risk assessments as more information is obtained about potential threats and hazards. Many armed forces are particularly slow to notice these changes, even when they are within their own control. For instance, most units suffer a higher frequency of mishaps during the early phases of a rotation. New personnel must learn to cope with new levels of operational demand. It is a continuing surprise that this lesson is often not explicitly represented in formal risk assessments. Too often remedial actions are not triggered until after a significant number of mishaps have already occurred. For instance, it can take many months before 'exiting' rotations are routinely involved in training their replacements.

- *Military risk management must be sensitive to human factors* Existing military doctrine ignores key findings in the human factors literature that has identifies constraints on risk assessment. In particular, Chapter Five has summarized the impact that fatigue can have upon military decision making. These effects cannot be addressed through the extension of civil risk assessment techniques. For example, a key finding is that as personnel become more and more fatigued, they are less able to recognize the effects of that fatigue. In consequence, increasing levels of tiredness will lead to a loss of situation awareness that is compounded by the difficulty of acknowledging the effects of fatigue on an individual's decisions;

- *Military risk management must not only look at the positive effects of new technologies.* Several of the chapters in this book have identified the short-sighted way in which risk assessments have been conducted to support the introduction of novel technologies. For example, the proponents of night vision technology overlooked many of the risks that were created when semi-trained individuals were issued with new equipment immediately before combat operations. In consequence, significant numbers of personnel have been involved in vehicle turn-overs. The accident rate involving night-vision aviation continues to be appalling. Similarly, the proponents of Unmanned Airborne Systems (UAS) seldom considered the risks to personnel who must recover vehicles that are lost in enemy areas. In such circumstances, ground forces have to go into locations that were considered too dangerous for conventional aircrews. From this it follows, that risk assessments should consider the negative knock-on effects of innovative technologies as well as the beneficial effects of risk mitigation.

- *Military risk management must not be unduly focused on previous operational environments.* Chapter seven argued that many recent military operations had been hindered by planning that was based on Cold War threats. This legacy should not be surprising giving the lengthy procurement processes and risk assessment procedures noted in previous paragraphs. However, one critical lesson has been that armed forces should be trained to cope with many different environments. This includes long duration exercises where existing equipment must be maintained and operated under field conditions. Unfortunately, budget cuts are closing many overseas bases. There is a danger that personnel will be denied the experiences that are most useful when they have to reconfigure equipment and SOPs to extremes of cold, of heat, of dust etc.

- *Military risk management must not be bounded by political expediency.* Political decisions place lives at risk. For example, additional troops are often delayed so that senior politicians can announce a 'surge' to the media. There is often political pressure to introduce new technologies, including UASs, before they are fully tested. These decisions are never motivated by a desire to increase operational risks. However, politicians are usually isolated from the effects that their decisions have upon military personnel. It is, therefore, important that commanders find ways of communicating potential risks to those that govern their actions. The case studies in this book have been deliberately selected to provide examples that might be shown to politicians in the future. For example, there is growing pressure on military personnel to support the deployment of autonomous systems within network 'enabled' forces. Much of this work has yet to be supported by an appropriate assessment of the potential risks should these systems fail to deliver their anticipated benefits.

- *Military tactics and operational decision making cannot be driven by risk assessments alone.* There is a danger that enemy forces could gain tactical or operational advantages based on knowledge of the risk assessment procedures that are used to guide military decision making. Previous sections have described how insurgents have used previous observations of coalition responses to Improvised Explosive Devices to rig secondary charges. This represents one example of a more general point; if all decision making is driven by risk assessment then this may create tactical vulnerabilities that can be exploited.

- *Military risk management must be supported by 'worst case' contingency planning.* Chapter nine described some of the problems that arise when an initial risk assessment is not supported by an adequate contingency plan. This leads to brittle plans that fail when things begin to go wrong. Arguably the most important element in contingency planning is to identify the nature of the hazards that might force units to abandon a mission. Without this, there is a danger that risk assessments will continually be revised 'on the fly' as new hazards are encountered. These revisions will continue until it is no longer possible to maintain either the success or the safety of a mission.

- *Military risk management must consider the moral, ethical and political impact on local populations.* Several of the chapters in this book have identified the complexity of conducting risk assessments in counter insurgency operations where the threat to military personnel must be balanced against the need to reduce risk to a local population. These issues also affect security and emergency personnel who must determine the best ways of mitigating the threats posed by Improvised Explosive Devices (IEDs). Many policing actions, such as the use of random search or checkpoints, can be used to deter attacks. However, they also have a disproportionate impact on civil liberties. These are complex issues that must inform military decision making in a growing number of peacekeeping operations. In contrast, civil risk assessments seldom consider these concerns.

These requirements have been derived from the case studies and incident reports in previous chapters of this book. They are not intended to be exhaustive. Other requirements might be identified by extending the scope of the analysis to consider a wider range of events. In contrast, the closing pages identify a road map for the future development of military risk management techniques.

11.2 A Roadmap for Military Risk Assessment

Previous chapters have identified the limitations that complicate the application of civil risk assessment techniques in military systems. However, the intention is not to completely reject the strengths that these approaches offer, for example in directing the allocation of finite resources. In contrast, the intention is to create a road map for the development of techniques that specifically support strategic, tactical and operational decision making in combat and peacekeeping operations. The ultimate contribution of this book has been to place us in a position where we can move ahead with the development of a second generation approach to risk assessment, which is specifically tailored for the changing demands of military organizations.

Figure 17 provides an overview of one way forward. As can be seen, this explicitly promotes a multi-disciplinary approach the combines both engineering and operational expertise from a range of backgrounds. Previous chapters have acknowledged the need to learn from human factors studies, in order to assess the impact of bias, of teamwork and of operational stressors on decision makers, including fatigue, on the application of hazard analysis inside military organizations. Systems engineering has also played an important role in our analysis, by identifying the importance of safety management systems in the integration of mishap reporting and prospective risk assessment.

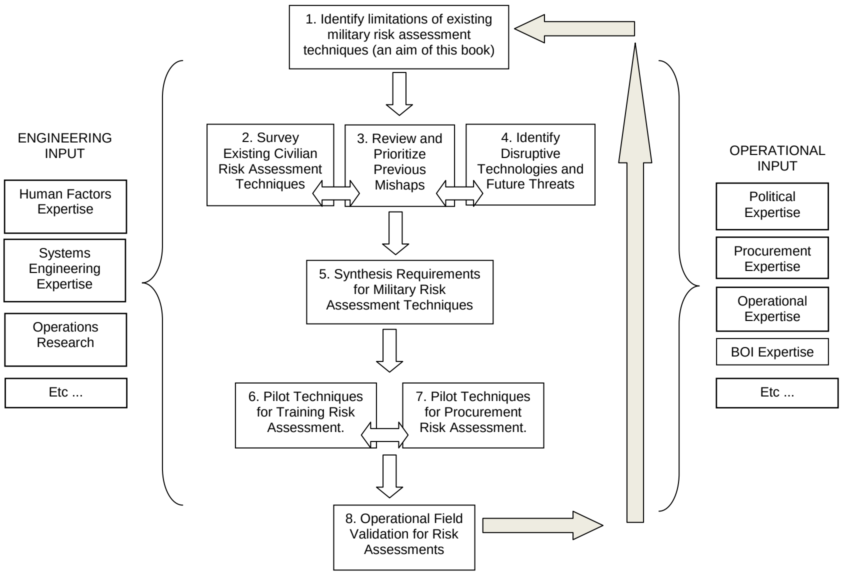

1. Identify limitations of existing military risk assessment techniques (an aim of this book)
ENGINEERING INPUT
Human Factors Expertise
Systems Engineering Expertise
Operations Research
Etc ...
2. Survey Existing Civilian Risk Assessment Techniques
3. Review and Prioritize Previous Mishaps
4. Identify Disruptive Technologies and Future Threats
5. Synthesis Requirements for Military Risk Assessment Techniques
6. Pilot Techniques for Training Risk Assessment.
7. Pilot Techniques for Procurement Risk Assessment.
8. Operational Field Validation for Risk Assessments
OPERATIONAL INPUT
Political Expertise
Procurement Expertise
Operational Expertise
BOI Expertise
Etc ...

Fig. 17. A Road Map for Next Generation Military Risk Management

Engineering input will contribute little if it is not motivated and focused by direct operational input. This is indicated on the right of Figure 17. As can be seen, the intention is to develop an integrated approach that spans procurement and operational risk management. Hence, expertise is required in these diverse areas. It is also critical to elicit input from stakeholders who have an understanding of why previous operations have failed. For this reason, any future roadmap must be informed by military incident investigators involved in previous Boards of Inquiry. Finally, this book has identified the impact of political decision making on operational, tactical and strategic risk assessments. Too often, risk assessment techniques have focused on a narrow set of engineering and technical objectives. In consequences, they often correctly identify design options from a range of hazard mitigation techniques. However, they fail to elicit the political support that is required to ensure the implementation of those options given changing operational demands and limited funding.

It is easy to sketch the diverse stakeholders that must contribute to the development of second generation, military risk management techniques. It is far harder to elicit the resources and commitments necessary to secure input from these different sources of information. However, the costs are trivial compared to the amounts wasted in previous procurement failures. They are negligible compared to the human cost in terms of the injuries and fatalities that have been documented in previous chapters.

Figure 17 also identified some of the inter-disciplinary tasks that are required in order to address the weaknesses of existing military risk assessment techniques. This book has attempted to support the first of these tasks – for example, by identifying limitations of existing risk assessment and hazard analysis techniques. However, this is only an initial step.

A second task is to survey civilian risk assessment techniques in greater detail. Most of the techniques applied by military organizations have relied on heavily simplified versions of existing approaches including HAZOPS and FMECA [6]. However, there is little evidence to suggest that any of these techniques avoid the problems of analytical bias or mitigate the problems of fatigue or operational stress that characterize military decision making. As far as we are aware, none of these approaches have been used to address the host of operational, ethical and political considerations that arise in peacekeeping operations. Many of the criticisms of military risk assessment arise because existing techniques have been applied in a relatively limited way. Often this is justified by attempts to keep the approaches as simple as possible, even when decision makers are faced with highly complex and dynamic operations. It is important to recognize, therefore, that some existing techniques could be used to address our concerns. It is not that civil techniques *could not* be used to support military risk assessment. The problem is more that they *have not* been used to assist many aspects of operational decision making.

A third task in the development of next generation risk assessment techniques is, therefore, to identify those situations where these approaches might provide the greatest support to military personnel. One means of doing this is to extend the approach developed in this book. By studying a host of previous military incidents and accidents, it is possible to identify those situations that have led to unnecessary loss of life and to mission failure. There is considerable scope for innovation here. For example, the US National Transportation Safety Board publishes a ‘top ten’ list of the most wanted safety improvements across the industries that they support. Some military organizations have established similar initiatives. These lists provide important insights into the types of hazards that should be a focus for next generation risk assessment techniques.

The fourth task identified in Figure 17 recognizes that many existing risk assessment techniques were originally developed in the 1960s and 1970s. It is unclear whether these approaches are well suited to the demands facing many military organizations. In particular, previous sections have cited concerns that it can be difficult to update initial hazard assessments as the operational demands change over time [4]. Other reviews have argued that they tend to focus on immediate engineering issues rather than the risks that emerge from the organization and management of complex operations [7]. It is for this reason that the roadmap includes a requirement to review potential disruptive technologies and future trends. Some of these are easily identified. For instance, previous chapters have considered the hazards that have been created b y the rapid introduction of UAS. Subsequent sections will consider the impact of network-centric warfare and cyber conflicts that are being considered by many military planners. It is equally clear that existing risk assessment techniques provide very limited support for the complexities that lie in front of us. A range of organizations as diverse as the US Air Force and the European Space Agency have responded to these challenges by extending generic problem solving techniques to help maximize the ‘creativity’ and ‘imagination’ that must be encouraged when trying to anticipate the hazards of future systems [7].

The next task identified in the roadmap is to synthesis the requirements for military risk assessment techniques from each of the previous stages. Previous sections have argued that it is important to derive common techniques that can be used at an operational, tactical and strategic level, covering procurement as well as long term planning. A unified approach would capture the interactions between these activities. This is critical when, as we have seen, strategic decisions have a direct impact on operational risks and vice versa. There are further pragmatic reasons for this integration, including reduced training costs and the ability to exchange risk assessments between different levels of the command structure. This synthesis then feeds into the development of techniques that can be evaluated during subsequent phases of the road map.

The sixth and seventh tasks help to ensure that the proposed approaches are ‘fit for purpose’. This is critical because too often we have seen risk assessment initiatives being rolled out across armed forces with little or no validation. These activities begin by applying the approaches to training and also to procurement. This is justified because previous chapters have identified a number of particular hazards associated with these two activities. A further justification is that a phased validation can delay operational validation until the value of an approach has first been demonstrated in these two areas. This final application of a proposed risk assessment technique is at the operational level; as we have seen the acute time pressures of many decisions leaves little opportunity to repair decisions that are based on inaccurate risk assessments. It is for this reason that any next generation risk assessment techniques must first demonstrate their value in other aspects of military risk assessment.

The roadmap is Figure 17 is iterative. The development of specialized approaches that are tailored for military risk assessment is not an end in itself. As we have seen, it is unrealistic to expect that such methods would entirely eliminate incidents or accidents. In consequence, we must monitor those mishaps that still occur to determine whether any subsequent refinement of the next generation approaches might help to avoid any residual incidents. Further concerns relate to the biases introduced in the opening chapter of this book; we must demonstrate that novel approaches do not unintentionally focus resources on particular classes of hazard at the expense of other potential risks.

11.3 The Future Risks of Cyber Defense

A number of further concerns motivate the introduction of an iterative feedback loop into the roadmap that is illustrated in Figure 17. In particular, we are aware of a continuing need to refine military risk assessment practices as technical innovation revolutionizes future military operations. In particular, many countries anticipate rapid and significant changes in the infrastructures that support their operations within the next 5-10 years. A key element will be the development of network centric operations [8] or network enabled capability [9]. These concepts envisage greater levels of force integration and coordination through systems of systems. These will be based on novel communications architectures and distributed computing applications that offer a range of operational benefits:

1. Networks will improve information sharing across and between joint forces;
2. Information sharing and collaboration will in turn help to sustain mutual situational awareness across heterogeneous forces;
3. Shared situational awareness, in turn, enables self-synchronization and coordination because the actions of others are transparent across the network.

Previous chapters have described how many disruptive technologies carry with them a price in terms of increased risk until Standard Operating Procedures (SOPs) are developed and technical problems are resolved. Examples have included the introduction of night vision devices and the integration of Unmanned Airborne Systems (UAS). It seems likely that many of the operational concepts that are guiding the development of network centric operations and network enabled capability will also carry additional risks. There is also a considerable concern that military organizations will again overlook the potential hazards until it is too late. Many of the key documents describing these visions do not contain any reference to the risks or vulnerabilities that are likely to characterize the initial application of these approaches.

These observations about network centric operations and network enabled capability illustrate the main themes in this book. These are summarized in Figure 18. As can be seen, there is considerable political and operational pressure to introduce innovative, disruptive technologies. The initial enthusiasm to obtain tactical and strategic benefits can obscure the risks associated with these innovations. At the same time, a lack of appropriate risk assessment methodologies also prevents many military organizations from identifying the potential hazards from innovations that include network centric operations/network enabled capability as well as UAS and night vision systems. The roadmap identified in Figure 17 is intended to ensure that we are better placed to disrupt these pressure that ultimately lead to an increased number of mishaps, similar to those that have been documented in the previous chapters of this book.

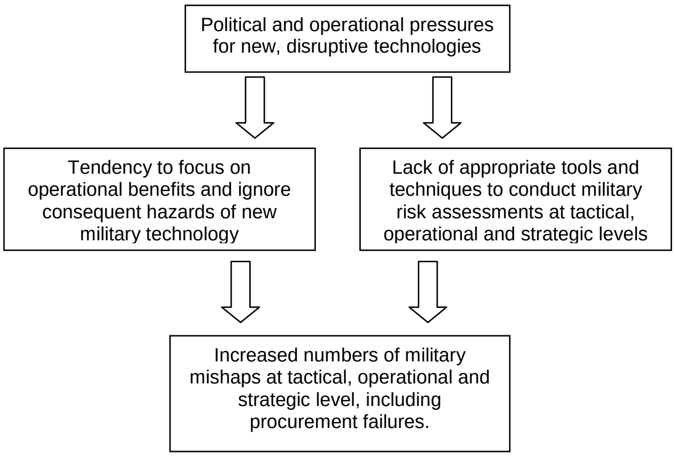

Political and operational pressures for new, disruptive technologies
Tendency to focus on operational benefits and ignore consequent hazards of new military technology
Lack of appropriate tools and techniques to conduct military risk assessments at tactical, operational and strategic levels
Increased numbers of military mishaps at tactical, operational and strategic level, including procurement failures.

Fig. 18. Innovation and Risk Assessment in Military Mishaps

The significant and varied benefits offered by network centric operations and network enabled capability make it an ideal testing ground for the development of next generation risk assessment techniques. The hazards that might arise from failures in the principles identified on the previous page suggest a range of concerns that need to be considered before lives are placed at risk. For instance, it is important to consider what might happen if system failures or cyber attacks lead to a loss of mutual situation awareness or to a loss of self synchronization. It remains to be seen whether the risk assessment approaches that are proposed within Figure 17 can meet the challenges posed by the safe and successful implementation of these new concepts.

11.4 Risk Assessment and Austerity

Technological innovation is not the only trigger for change across the armed forces. A new period of austerity is also providing a catalyst for most military organizations. For instance, the United has faced slow economic growth and relatively high levels of unemployment. The Obama administration has, therefore, argued about the increasing importance of resolving overseas conflicts and of refocusing on domestic economic concerns. Part of this debate has focused on the defence budget, which has risen to approximately $665 billion dollars. Although precise comparisons are hard to make, this exceeds the combined expenditure of the next nine largest defence spenders. Both the House and Senate have opposed elements of the Obama 2011 defense appropriations, arguing that savings can be made with increasing levels of security in Afghanistan following increased deployments in 2010.

Many other countries continue to face more severe fiscal problems. Concerns over national credit ratings and over their vulnerability to future economic crises have motivated spending reviews that have increased pressure on defense budgets. For example, the UK spends £36.9 billion per annum. Although the details have yet to be announced, it seems likely that this figure will be cut by 10 to 20% over the next two years. This has prompted speculation about delays or cancellations in procurement programmes. One approach would save some £1.7 billion by cancelling a fleet of 62 new Lynx Wildcats for the Navy and Army. It would also involve phasing out the Navy and RAF's Sea Kings. Such plans are controversial given the relative importance of rotary winged aircraft in many of the conflicts described in this book. Alternative plans focus on significant cuts in the procurement of 138 Joint Strike Fighters and a £10.5 billion pound contract for air tankers.

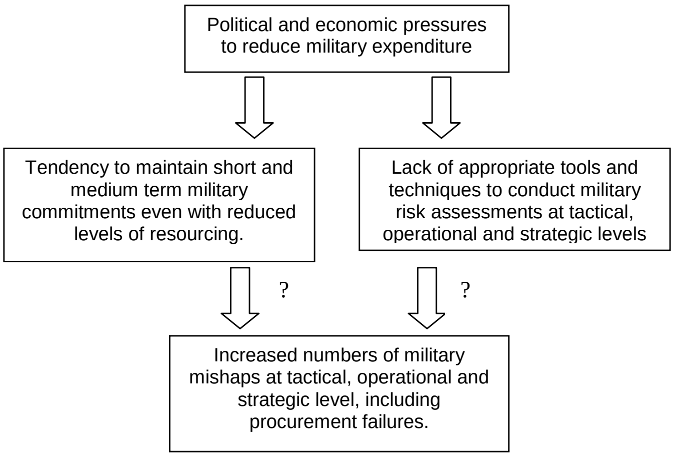

Political and economic pressures to reduce military expenditure
Tendency to maintain short and medium term military commitments even with reduced levels of resourcing.
Lack of appropriate tools and techniques to conduct military risk assessments at tactical, operational and strategic levels
?
?
Increased numbers of military mishaps at tactical, operational and strategic level, including procurement failures.

Fig. 19. Austerity and Risk Assessment in Military Mishaps

Figure 19 represents the impact that austerity measures might have upon future operational capabilities. It acknowledges that there is a lag between any reduction in the budget for military operations and a corresponding decrease in the scope of military commitments. These problems are exacerbated by the lack of appropriate risk assessment tools. Just as new procurements and technological innovations alter the hazards faced by personnel, cuts and cancellations also create significant risks. For instance, units have to ‘make do’ with equipment that might otherwise be replaced. Unless remaining resources can be channeled more effectively then there is a danger that insurgent or other opposition forces will profit from the consequences of austerity. It is difficult to ensure cost effectiveness without appropriate tools for military risk assessment.

The hazards of austerity are more complex than those created by technological innovation. Typically, the scope and extent of a cut are constrained by a host of political, financial and organizational factors that have very little to do with operational, tactical or strategic requirements. Most defense companies create contractual obligations that cushion themselves against the impact of any change in procurement decisions. In many cases, it can be more expensive to withdraw an order than to proceed with it. In such circumstance, areas of the budget are protected for contractual rather than operational reasons. Similarly, inter-service rivalries can distort the impact of austerity measures. There is very little reason to believe that most recent funding reductions have been informed by the types of risk-based decision making that has been described in the previous sections of this book. Instead, there is a pressing need for expediency to meet short-term fiscal demands. It remains to be seen whether the consequence decisions have an impact on casualty rates. There is an increasing need for military forces 'to do more with less' but this may only be achieved at greater risk to the men and women who serve around the globe.

11.5 References for Chapter Eleven

[1] C.W. Dixon and N.J.W. Moss, A Strategic review of the Puma Helicopter Force, UK Ministry of Defence, London, UK, May 2008. http://www.mod.uk/NR/rdonlyres/66D22157-514E-4B1B-87C5-C445234C9C26/0/puma_review_redacted.pdf

[2] S. Gaidow and S. Boey, Australian Defence Risk Management Framework: A Comparative Study, Land Operations Division, Systems Sciences Laboratory, Technical report DSTO-GD-0427, Australian Government Department of Defence, Defence Science and Technology Organization, 2005.

[3] US Department of Defense, FAQ: Enterprise Risk Assessment Methodology (ERAM), Technical report, Defense Business Transformation Unit, Washington DC, USA, April 2010. http://www.dod.mil/dbt/faq_eram.html

[4] C. Haddon-Cave, The Nimrod review: An independent review into the broader issues surrounding the loss of the RAF Nimrod MR2 Aircraft XV230 in Afghanistan in 2006, HC 1025 London: The Stationery Office, October 2009.

[5] US Department of the Army, FM 3-04.513: Battlefield Recovery and Evacuation of Aircraft, Headquarters, Washington, DC, 27 September 2000. http://www.army.mil/usapa/doctrine/Active_FM.html

[6] J.D. Andrews and T.R. Moss, Reliability and Risk Assessment, Longman, New York. USA, 1993.

[7] C.W. Johnson, Software Configuration Management for Safety Related Applications in Space Systems: Extending the Application of the USAF 8-Step Method. In Proceedings of the International Association for the Advancement of Space Safety, Huntsville Alabama, NASA/ESA, 2010.

[8] UK Ministry of Defence, Network Enabled Capability, Joint Service Publication, 777 Edition 1, London, UK. 2005.

[9] US Department of Defense. The Implementation of Network-Centric Warfare. Washington, D.C., 2005.

12 Acknowledgements

This book was supported by a sabbatical provided by the University of Glasgow. Thanks go to my many friends and colleagues who had to look after my teaching and administration while I was locked in my office. Particular thanks are due to Tony Licu, Andrew Kilner and Barry Kirwan who helped refine ideas on the accident pit, introduced in Chapter Nine.

Some sections are based on updated and revised versions of previous papers. I am grateful to the organizing committees, learned societies and publishers in allowing the adaptation of this material.

- Chapter Two is based on C.W. Johnson, The Paradoxes of Military Risk Assessment. In A.G. Boyer and N.J. Gauthier (eds), Proceedings of the 25th International Systems Safety Conference, Baltimore, USA, International Systems Safety Society, Unionville, VA, USA, 859-869, 2007.

- Chapter Three draws on material from an unpublished invited talk that was presented Human Factors 2002 in Sydney Australia, organized by F. Vetere, L. Johnson and R. Kushinsky.

- Chapter Four develops material that first appeared as C.W. Johnson, Degraded Modes and the 'Culture of Coping' in Military Operations: An Analysis of a Fatal Incident On-Board HMS Tireless on 20/21 March 2007. In J.M. Livingston, R. Barnes, D. Swallom and W. Pottraz (eds.) Proceedings of the US Joint Weapons Systems Safety Conference 2009, Huntsville, Alabama, 3511-3521, 2009.

- Chapter Five reworks material from C.W. Johnson, The Systemic Effects of Fatigue on Military Operations. Proceedings of the 2nd IET Systems Safety Conference, the IET, Savoy Place, London, UK, 1-6, 2007.

- Chapter Six is based on C.W. Johnson, The Role of Night Vision Equipment in Military Incidents and Accidents, pp.1-16, in Human Error, Safety and Systems Development, eds. C.W. Johnson and P. Palanque, Kluwer Academic Press, Boston, 2004. With Kind permission of Sprigner Science+Business Media, B.V..

- Chapter Seven builds on C.W. Johnson, Interactions between Brown-out Accidents and Night Vision Equipment in Military Aviation Accidents. In J.M. Livingston, R. Barnes, D. Swallom and W. Pottraz (eds.) Proceedings of the US Joint Weapons Systems Safety Conference 2009, Huntsville, Alabama, 3508-3518, 2009.

- Chapter Eight develops material from C.W. Johnson, Act in Haste, Repent at Leisure: An Overview of Operational Incidents Involving UAVs in Afghanistan (2003-2005), Proceedings of the Third IET Systems Safety Conference, NEC, Birmingham, UK, 2008, IET Conference Publications, Savoy Place, London, 2008. It also includes material from C.W. Johnson, Insights from the Nogales Predator Crash for the Integration of UAVs into the National Airspace System under FAA Interim Operational Guidance 08-01, in J.M. Livingston, R. Barnes, D. Swallom and W. Pottraz (eds.), Proceedings of the 27th International Conference on Systems Safety, Huntsville, Alabama, USA 2009, International Systems Safety Society, Unionville, VA, USA, 3066-3076, 2009.

- Chapter Nine extends C.W. Johnson, Military Risk Assessment in Counter Insurgency Operations: A Case Study in the Retrieval of a UAV Nr Sangin, Helmand Province, Afghanistan, 11th June 2006), Proceedings of the Third IET Systems Safety Conference, NEC, Birmingham, UK, 2008, IET Conference Publications, Savoy Place, London, 2008.

- Chapter Ten develops an argument first sketched in C.W. Johnson and L. Nilsen-Nygaard, A 'Systemic Approach' for Countering the Threat to Public Safety from Improvised Explosive Devices (IEDs). In J.M. Livingston, R. Barnes, D. Swallom and W. Pottraz (eds.) Proceedings of the 27th International Conference on Systems Safety, Huntsville, Alabama, USA 2009, 3048-3058, 2009.

Finally, thanks are due to my family. They have motivated, inspired and endured me through this work.

Chris Johnson, Glasgow, September 2010.