Incident Learning Systems: From Safety to Security

Finn Olav Sveen[1], Jose Maria Sarriegi[1] and Jose J. Gonzalez[2]
[1]Tecnun, University of Navarra
Paseo de Manuel Lardizábal, 13, 20.018 Donostia-San Sebastián, Gipuzkoa, Spain
fosveen@tecnun.es, jmsarriegi@tecnun.es
[2] Agder University College
Faculty of engineering and science, Research Cell "Security and Quality and Organizations," Serviceboks 509, 4898 Grimstad, Norway (and Gjøvik University College, Norwegian Information Security laboratory, 2802 Gjøvik, Norway)
jose.j.gonzalez@hia.no

Abstract

The complexity of modern networked systems has negative consequences in the form of intended and unintended security incidents. Information security is not the first field to grapple with such challenges. In safety, incident learning systems (ILS) have been used to control high risk environments. Many of these systems, such as NASA's Aviation Safety Reporting System, have demonstrated considerable success while others have failed. Prior to implementing ILS in information security, it is prudent to learn from experiences gained in safety. We use System Dynamics to investigate how factors such as management commitment, incentives, recriminations and resources affect a safety incident learning system. We find that the rate of incidents is not a suitable indicator of the state of the system. An increasing or decreasing incident rate may both be caused by either increased or decreased security. Other indicators, such as the severity of incidents, should be used.

1. Introduction

Modern computer networks are highly complex and interact in ways which the designers never intended. These unforeseen interactions may cause errors or unintended consequences in the system (Schneier 2000). The complexity makes it difficult if not impossible to implement satisfactory security with a purely preventive approach. It is likely that there will always be cracks in the defensive wall.

Other high complexity environments, such as those facing considerable safety challenges, have for many years utilized incident learning systems to counter high complexity. "Although accidents may be "*normal*," disaster is not an inevitable consequence of complex socio-technical systems. Since incidents of varying severity *are* normal, a system must be put in place to control the severity of these incidents. Without such a system the incident rate and severity will not be controlled and only *then* is a disaster predictable." (Cooke 2003) Incident learning systems can be thought of as a form of quality improvement systems (Gonzalez 2005). These systems aim to improve quality by continuously eliminating deviations from the quality standard.

The most well known incident learning system is probably NASA's Aviation Safety Reporting System (ASRS). Such systems allow organizations to capture and document breaches of safety, their causes and possible solutions. ASRS and similar systems have demonstrated considerable success (Lee and Weitzel 2005). Incident learning systems have been widely adopted in chemical processing industry, health care and aviation.

The complexity of modern networked systems and the considerable success that many incident learning systems have had, prompts us to call for their application to information security. In addition there is a convergence between the safety and security realms. A previously purely mechanically operated pump is today controlled by embedded microprocessors running Linux or other similar standard systems, making them vulnerable to many of the same threats that are faced in a traditional desktop environment. Security breaches in equipment such as pumps may lead to potentially severe accidents. Since identifying all possible security vulnerabilities in a networked system prior to start up is incredibly difficult, if not impossible, it becomes necessary to anticipate that incidents will happen and to have an organization and routines in place to mitigate and learn from incidents.

The preceding factors motivated us to undertake a study on safety incident learning systems to see what the field of information security may learn from these systems. Although many safety incident learning systems have been successful, there are also many that have been partial or even complete failures.

In this paper we present a System Dynamics model that is based on safety literature[1]. The model is not based upon a single case, but is a synthesis of different cases and general safety theory. Our goal is to transfer experience from the safety to the information security realm. As such, the model has not yet been adapted to include security issues such as exponentially growing attack rates or automated reporting tools (Wiik, Gonzalez, and Kossakowski 2004). We believe it is necessary to look at the fundamental lessons of safety incident reporting systems before moving on to include specific security issues.

We chose System Dynamics since it has previously been successfully applied to investigate other aspects of the dynamics of incident learning systems (Cooke 2003, 2003; Cooke and Rohleder 2006). System Dynamics is particularly well suited to complex, feedback-driven socio-technical systems.

In section 2, Models of Incident Reporting Systems, we describe briefly some of the theoretical basis of our System Dynamics model. In section 3, System Dynamics Incident Learning System Model, we first show an overview of the model in causal loop form before we move on to explain the stock and flow structure. Section 4, Model Runs, contains our analysis of the model's behavior. Finally in section 5, Conclusions and Future Work, we revisit information security.

2. Models of Incident Reporting Systems

Although there are many theoretical models for safety incident learning systems, we have chosen three to base our simulation model on. The three models are presented below.

Nyssen et al.(2004) presents a generic structure for an incident reporting system in healthcare. The main points are summed up below.

1. Reporting
2. Analysis and classification
3. Identification and proposal of remedial actions

[1] The model was created using Vensim DSS (http://www.vensim.com/)

4. Assessment

Reporting is achieved by the means of an *interface*, either by questionnaire, an interview or automatic data collection. A questionnaire is the currently most used method. The reporting system should include a method to *analyse data* and they have a *classification scheme*. In many reporting systems, the classification scheme is built empirically on the basis of the reported data and is domain specific. In other systems the classification is derived from psychological models. There is now consensus among experts to define accidents as a system failure; however, analysis illustrating the multi-causal aspect of an accident is still rare. The next step is to identify and propose *remedial and preventive actions* and then implement and follow up. An incident reporting system should also include some sort of *assessment* of how they are working. Up until now, reporting systems which include an assessment phase have been rare.

Phimister et al. (2003) present an alternative seven stage framework:

1. Identification: An incident is recognized to have occurred.
2. Reporting: An individual or group reports the incident.
3. Prioritization and Distribution: The incident is appraised and information pertaining to the incident is transferred to those who will assess follow-up action.
4. Causal Analysis: Based on the near-miss, the causal and underlying factors are identified.
5. Solution Identification: Solutions to mitigate accident likelihood or limit impact on the potential accident are identified and corrective actions are determined.
6. Dissemination: Follow-up corrective actions are relayed to relevant parties. Information is broadcast to a wider audience to increase awareness.
7. Resolution: Corrective actions are implemented and evaluated, and other necessary follow-up action is completed.

The seven stages have a "conjunctive" effect on each other. Near-misses[2] that are not identified can not be used to reduce risk exposure. Identified near-misses that have been reported but are not acted upon further will, at best, have a modest impact on reducing site-risk exposure.

Kjellén (2000) presents a six stage model of incident learning.

1. Reporting and collection of data
2. Storing of data in a memory and retrieval of data from it
3. Information processing
4. Distribution
5. Decisions
6. Production System

The first step involves the collection of data on accidents and near-misses. This is achieved by investigations, workplace inspections, audits and risk analyses. Data collection methods include observation, interviews, self-reporting, group discussions, etc. In the second step data is stored in a memory and also retrieved for later use. The memory is typically a database. The third step is the analysis and compilation of the

[2] A near-miss is an incident that narrowly avoided becoming an accident.

retrieved data into meaningful information as well as the development of remedial actions. The fourth step is the dissemination of information to decision-makers within the organization. Kjellén also includes the decisions made and the industrial production system. He notes that these six steps form a loop and that it must be closed for the incident learning system to work.

These three models form the main basis of our SD model of a safety incident learning system. Where appropriate we have also drawn on other sources.

3. System Dynamics Incident Learning System Model

Terminology

As previously explained, the model is based mostly upon safety literature. However, since we want to transfer experience from safety to information security we have chosen to use information security terminology. Henceforth we shall use the term event to describe a potential breach of safety instead of the equivalent safety term near-miss. The term incident will denote an actual breach of safety which in safety terminology would be an accident.

High Level Overview

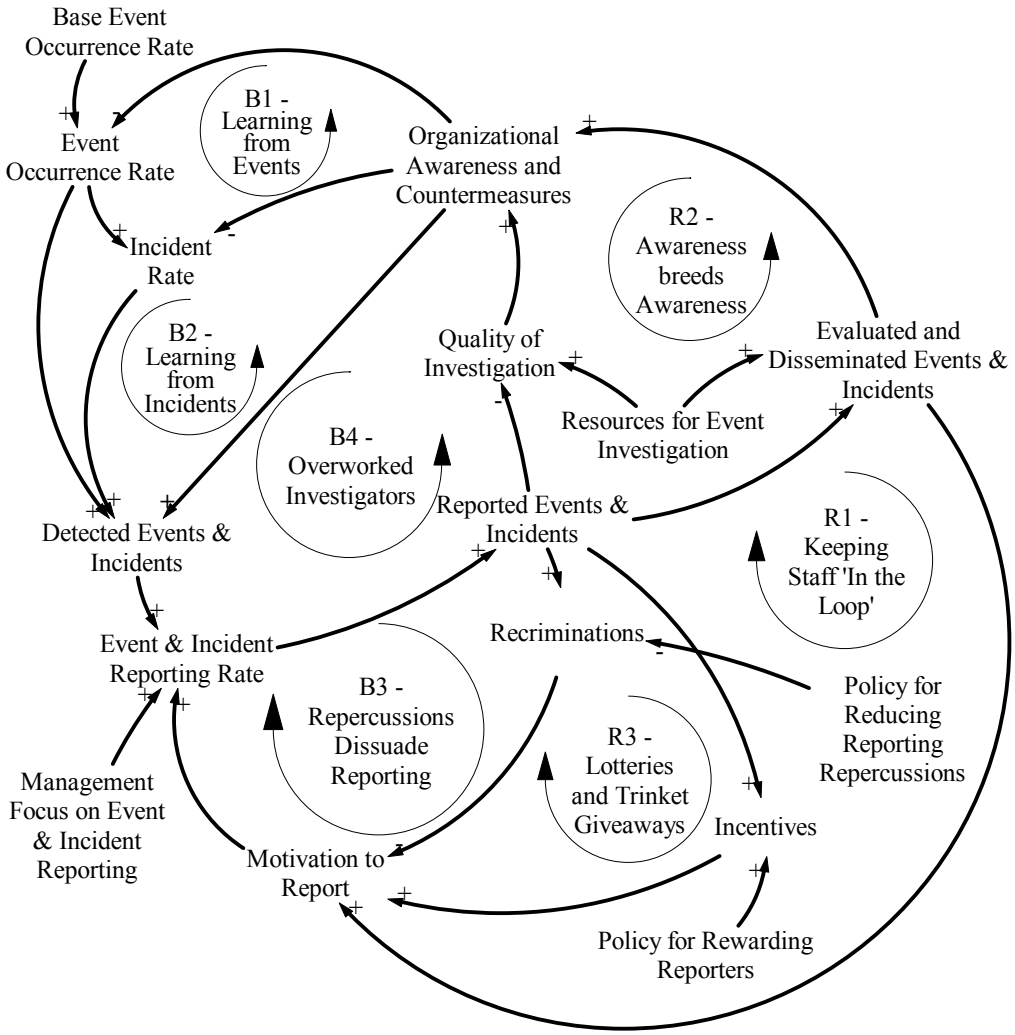


Figure 1. High level overview of the Incident Reporting System model

Before we go into the details of the model's stock and flow structure we will give a brief overview of the model's main structure and the issues that it covers. Such an overview is shown in causal loop form in Figure 1. High level overview of the Incident Reporting System model An incident reporting system aims to reduce future cost (monetary, injuries, fatalities) by controlling the incident and event rates through learning from incidents as seen in loops B1 and B2. This constitutes a form of negative feedback. Negative feedback loops or balancing feedback loops describe controlling actions that seek to lead the system to a specific state.

When reports of investigated incidents are spread to relevant personnel in the organization and countermeasures implemented, the organization as a whole should become more aware of safety issues. This increased awareness should lead to a better ability to detect incidents and events as depicted in loop R2. As shown in loop R3, organizations may utilize incentives to speed up the process of learning from incidents and events. R2 and R3 are reinforcing, or positive feedback loops. These loops reinforce underlying effects.

However, present in the system there may be recriminations that are detrimental to motivation of personnel to report incidents, depicted in loop B3. Furthermore, loop R1 shows the influence of feedback to reporting personnel on the motivation to report. If this feedback is lacking, reporters may be dissuaded from reporting in the future. A crucial part of the system is also the resources assigned to investigate reported incidents and events. As shown in B4, insufficient resources may lead to overworked investigators leading to reduced quality of investigation. In addition to reduced quality, insufficient resources would lead to reduced throughput which will impact loops B1, B2, R1 and R2 negatively.

We will now turn to a more detailed description of the stock and flow structure of the model. Many of the concepts shown in the causal loop diagrams are disaggregated in the stock and flow simulation model.

Incidents and Events

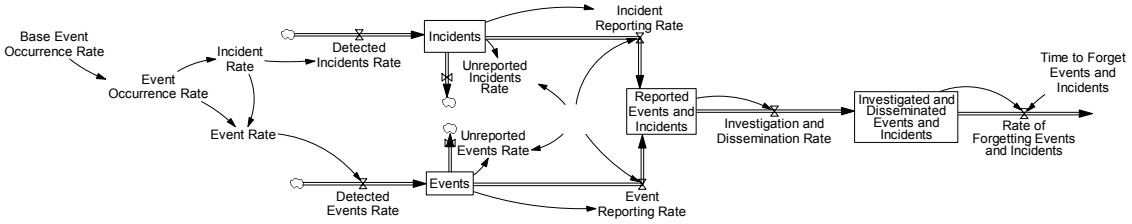


Figure 2. Flow of reported incidents and events.

The purpose of this modeling work is to investigate how events and incidents can be captured and learned from. The issue of how events and incidents occur is therefore considered outside the scope of this model. Thus we have modeled the source of incidents and events as an exogenous constant.

Base Event Occurrence Rate = 400 events / month

This variable represents the amount of events that would occur in the system if learning did not take place. The effect of learning is depicted by the influence of general

awareness about safety or security issues, as well as specific countermeasures. Countermeasures may be technical, such as firewalls, or organizational such as access control.

*Event Occurrence Rate = Base Event Occurrence Rate*Effect of Awareness and Countermeasures on Event Occurrence Rate*

The events that occur in the system may be mitigated and kept from escalating. In this case it stays an event (near-miss). If not mitigated the event becomes an actual breach of security: an incident. The safety community debates whether incidents and events have the same causes. For our model we assume that they do. Only a small fraction of events actually become incidents, in line with the iceberg model, i.e. only the tip of the iceberg of problems is seen, but there are many more near-misses that might have been incidents. The timeframe for escalation of an event to incident is relatively small, ranging from seconds to hours. The time frame of the model is five years. It is therefore not necessary to include the escalation process itself in the model. Therefore, we instead change the probability of an event being an incident.

*Undetected Incidents Rate = Event Occurrence Rate*Fraction of Incidents*

Undetected Events Rate = Event Occurrence Rate- Undetected Incidents

After occurrence, events and incidents must first be detected before they can be reported. We will return to the factors affecting detection later.

*Detected Incidents Rate = Incident Rate*Fraction of Detected Incidents*

*Detected Events Rate = Event Rate*Fraction of Detected Events*

The stocks *Incidents* and *Events* represent the incidents and events that have been detected. The detector must now decide whether or not to report them. This process is usually undertaken by line personnel such as operators or nurses.

*Incident Reporting Rate = (Incidents*Fraction of Reported Incidents)/Time to Report Events*

Unreported Incidents Rate = (Incidents(1-Fraction of Reported Incidents))/Time to Report Events*

*Event Reporting Rate = (Events*Fraction of Reported Events)/Time to Report Events*

Unreported Events Rate = (Events(1-Fraction of Reported Events))/Time to Report Events*

If not reported the event or incident is lost forever unless it reoccurs. Reported incidents and events flow into the *Reported Events and Incidents* stock. We assume that all incidents and events have the same potential for learning. This is not true in reality since incidents and events have differing severity, but because the model works on averages over time it is a reasonable simplification. Furthermore we assume that there is a learning curve where investigation of similar incidents in the future will give incrementally diminishing returns. See below in section Learning Effects for more details.

Reported events and incidents must be investigated to contribute to learning. In an ideal environment an investigative team takes over and attempts to find the root cause(s). Lessons learned must subsequently be distributed to all relevant parties. We have

chosen to aggregate the investigation and dissemination steps into a single variable. Investigating without dissemination does not make much sense. It would break the chain and learning would stop.

Investigation and Dissemination Rate = min(Investigation Capacity, Reported Events and Incidents/Time to Evaluate Event)

Investigation Quality

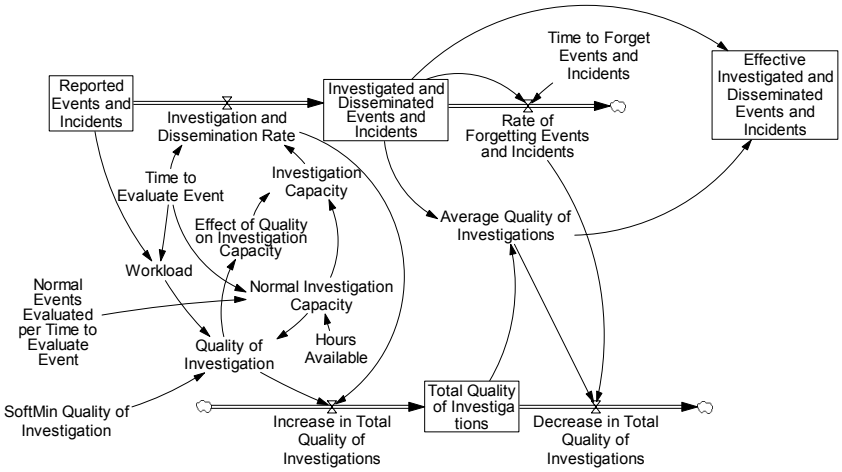


Figure 3. Quality of Investigation

Learning from incidents depends on the thoroughness of the investigation step. This is contingent upon the investigative team's skill, their resources and the time available, as well as the cooperation of those involved in the incident. Failing to find the underlying systemic antecedents to incidents may dissuade reporting as its perceived usefulness falls. In the words of Johnson, "Incident reporting systems can provide important reminders about potential hazards. However, in extreme cases these reminders can seem more like glib repetitions of training procedures rather than pro-active safety recommendations. Over time the continued repetition of these reminder statements from incident reporting systems is symptomatic of deeper problems in the systems that users must operate." (Johnson 2003, p.27)

There is little point in reporting incidents if they are not properly investigated. Lack of resources may actually lead to more recriminations within the system. Investigators tend to blame human error since it is the least labor intensive for them (Kjellén 2000).

In the model the quality of an investigation has been simplified to a function of the workload and available resources.

Workload = Reported Events and Incidents/Time to Evaluate Event

*Quality of Investigation = (Normal Investigation Capacity/Workload)*SoftMin Quality of Investigation(1/(Normal Investigation Capacity/Workload))*

If the investigative team has more work than they have resources to handle, they increase their capacity by lowering the quality of investigations. This model does not take into account the possibility of triage.

Effect of Quality on Investigation Capacity is a lookup table where lower quality is translated into higher capacity.

*Investigation Capacity = Normal Investigation Capacity*Effect of Quality on Investigation Capacity*

It is not only the quality of the events and incidents currently being investigated that determine long term learning effects, but also previously investigated incidents will have an effect on e.g. willingness to report. The total quality of investigations is therefore captured in a co-flow. The average quality of investigations determines how strong the effects of learning are on future incidents and events.

Total Quality of Investigations (stock) = +Increase in Total Quality of Investigations-Decrease in Total Quality of Investigations

Average Quality of Investigations = Total Quality of Investigations/Investigated and Disseminated Events and Incidents

*Effective Investigated and Disseminated Events and Incidents = Investigated and Disseminated Events and Incidents*Average Quality of Investigations*

Motivation to Report

The decision of whether to report an incident depends on the amount and strength of management focus, reporting incentives and recriminations.

More than 100% of detected incidents or events can not be reported. It is also likely that staff will not go out of their way to report the last few incidents and events, as these may be the more insignificant ones. Soft minimum functions[3] are therefore used to keep 'Fraction of Reported Incidents' and 'Fraction of Reported Events' between unity and zero.

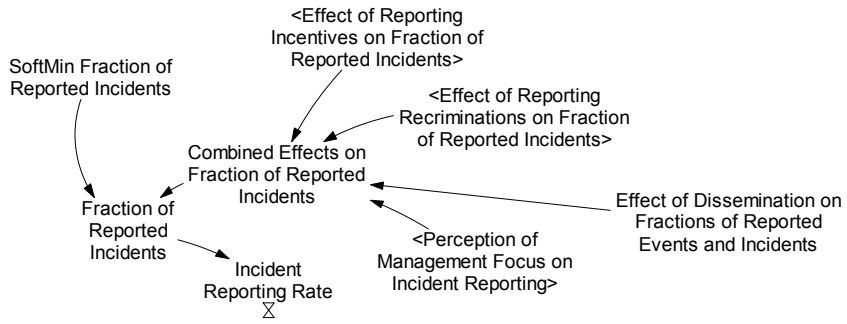


Figure 4. Factors affecting reporting of incidents.

*Fraction of Reported Incidents = Combined Effects on Fraction of Reported Incidents*SoftMin Fraction of Reported Incidents(1/Combined Effects on Fraction of Reported Incidents)*

*Combined Effects on Fraction of Reported Incidents = Effect of Dissemination on Fractions of Reported Events and Incidents*Effect of Reporting Incentives on Fraction of Reported Incidents*
**Effect of Reporting Recriminations on Fraction of Reported Incidents*Perception of Management Focus on Incidents*

[3] See Sterman (2000) for a definition and explanation of soft minimum and maximum functions.

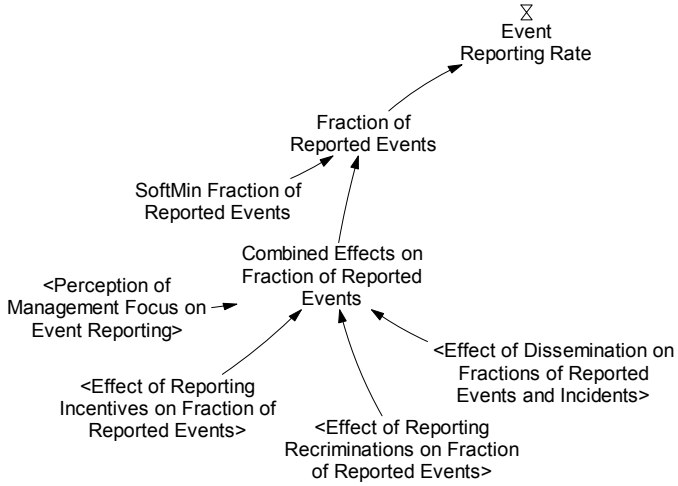


Figure 5. Factors affecting reporting of events.

Fraction of Reported Events = Combined Effects on Fraction of Reported Events
**SoftMin Fraction of Reported Events(1/Combined Effects on Fraction of Reported Events)*

*Combined Effects on Fraction of Reported Events = Effect of Dissemination on Fractions of Reported Events and Incidents*Effect of Reporting Incentives on Fraction of Reported Events*
**Effect of Reporting Recriminations on Fraction of Reported Events*
**Perception of Management Focus on Events*

Management Focus

Fear of liability and sporadic emphasis by management may hinder the functioning of an incident reporting system (Phimister et al. 2003). When management commits to something they set the agenda for what is important and should be focused on. If top-management disregards safety, middle-managers and staff will do so too.

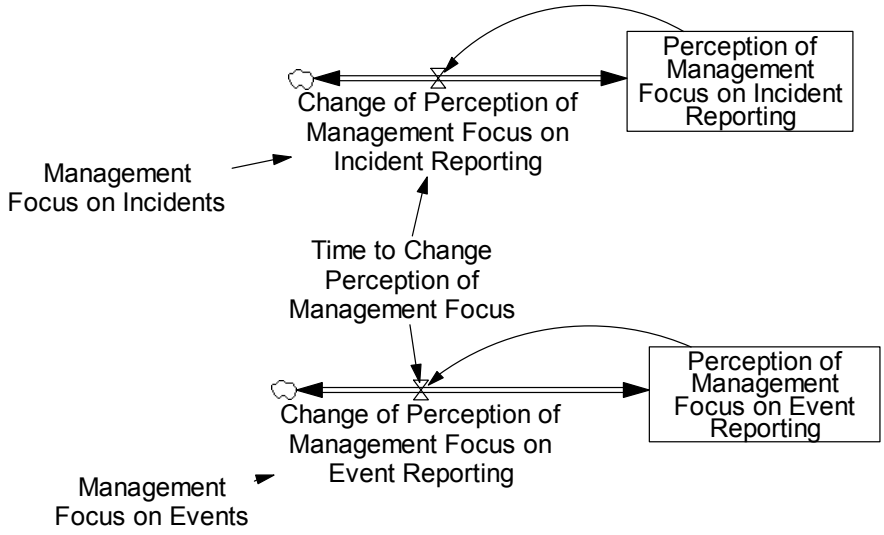


Figure 6. Management focus and the organization's perception of it.

'Management Focus on Incidents' and 'Management Focus on Events' represents how important management thinks incident and event reporting is. We assume that it takes time for management to communicate and change staff perception of focus. In the model it takes three months for the change in management focus to penetrate the organization.

Perception of Management Focus on Incident Reporting = Integ(Change of Perception of Management Focus on Incidents)

Change in Perception of Management Focus on Incident Reporting = (Management Focus on Incidents-Perception of Management Focus on Incident Reporting)/Time to Change Perception of Management Focus

Perception of Management Focus on Event Reporting = Integ(Change of Perception of Management Focus on Event Reporting)

Change of Perception of Management Focus on Event Reporting = (Management Focus on Events-Perception of Management Focus on Event Reporting)/Time to Change Perception of Management Focus

In addition to setting focus, management also decides on incentive programs and to a large extent influence the reporting culture. Thus it is likely that management plays a pivotal role in reducing reporting recriminations in the workplace. Cooke's case study of Nova Chemicals' Decateur plant reveals that strong management involvement was crucial to turn it from a low to a top safety performer (Cooke 2004).

Reporting Recriminations

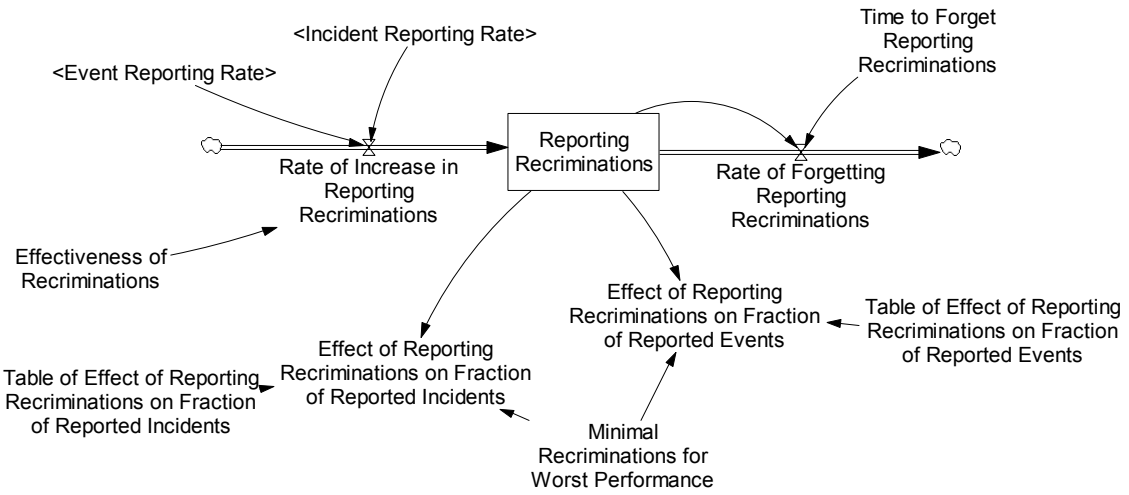


Figure 7. Recriminations and their effect on reporting.

A working environment has many factors that may potentially work against reporting. Staff may fear punishment for breaking rules or making mistakes. Punishment has detrimental effects on reporting. To avoid this, NASA's Aviation Safety Reporting System (ASRS) is completely anonymous (Johnson 2003). "The ASAP [American Airlines Aviation Safety Action Program] and ASRS programs have been successful because they offer protection for the reporting individuals; hence, both programs have experienced high participations rates." (Lee and Weitzel 2005)

Reporting an incident may lead to persecution from colleagues, who may feel that they are being snitched upon and that the reporter is disloyal. Employers may punish staff for making mistakes, and in such a way encourage hiding incidents (Johnson 2003; Phimister et al. 2003). A worker may also be dissuaded from reporting an incident because of fear of being seen as incompetent by other staff (Anderson and Webster 2001).

Furthermore, confidentiality and disclosure issues may not just stem from the need to protect a worker's identity from colleagues or employers. Accident investigators often have a complex relationship with the media and public disclosure of sensitive information can jeopardize an enquiry (Johnson 2003).

Fear of persecution may also stem from cultural differences. In the Taiwanese aviation industry, as a result of Chinese culture, punishment is often seen as the only solution to a problem. Unlike the western aviation industry where punishment is often the last resort. Consequently incident reporting systems in Taiwan's aviation industry have often been used as a means to punish air crew, severely limiting participation in incident reporting schemes (Lee and Weitzel 2005).

Another example of punishment culture can be found in nursing. The nursing literature is full of examples of a person centered blame approach (Anderson and Webster 2001). Anderson and Webster (2001) describe a professional culture where the nurse is seen as the only source of drug administration error and punishment is seen as the only effective solution. Such a culture will dissuade many from participating in an incident reporting scheme.

Phimister et al. (2003) classifies recriminations into four groups:

1. Peer pressure
2. Investigation style
3. Direct disciplinary action
4. Unintended disciplinary action

In our model it is unnecessary to operate with four different types of recriminations. What are of interest are how strong the recriminations are and their effect on reporting. We therefore simply use the word recrimination for all four.

In the model we track recriminations as a co-flow to incident and event reports. Each report is accompanied by a recrimination whose strength is determined by 'Effectiveness of Recriminations'.

*Rate of Increase in Reporting Recriminations = (Event Reporting Rate+Incident Reporting Rate)*Effectiveness of Recriminations*

The recriminations flow into the 'Recriminations' stock. Over time the bad experiences following from recriminations may be forgotten by the organization as staff and management is changed or new management principles gain prominence. Safety experts we have spoken to have told us that bad experiences with reporting linger for a considerable time. Sometimes people remember for many years. 'Time to Forget Recriminations' has therefore been set to 24 months.

Rate of Forgetting Reporting Recriminations = Reporting Recriminations/Time to Forget Reporting Recriminations

The recriminations and their strength partially determines how many detected incidents and events that are reported.

Effect of Reporting Recriminations on Fraction of Reported Incidents = Table of Effect of Reporting Recriminations on Fraction of Reported Incidents(Reporting Recriminations/Minimal Recriminations for Worst Performance)

Effect of Reporting Recriminations on Fraction of Reported Events = Table of Effect of Reporting Recriminations on Fraction of Reported Events(Reporting Recriminations/Minimal Recriminations for Worst Performance)

Reporting Incentives

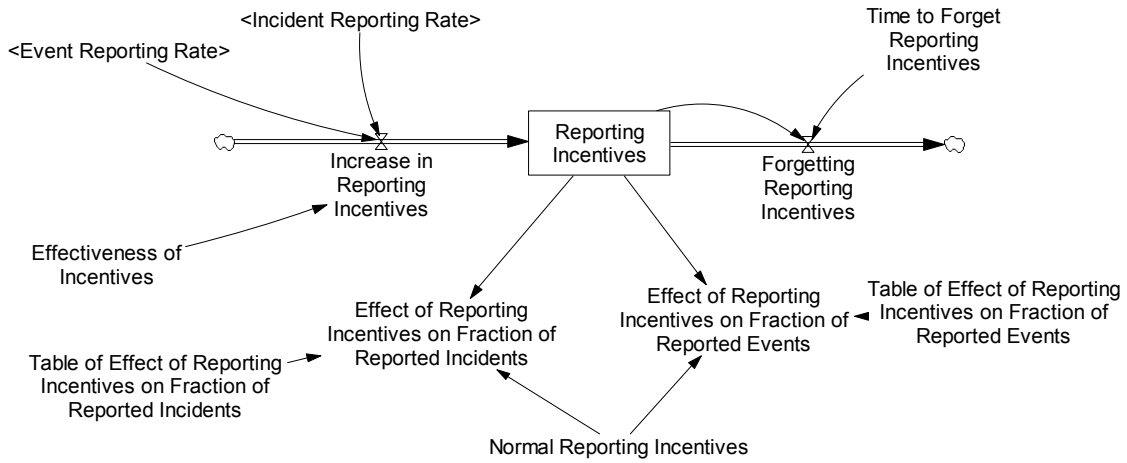


Figure 8. Incentives and their effect on reporting.

Some companies find it useful to reward reporting through incentive schemes. In their study of safety in the chemical processing industry Phimister et al. (2003) identified two different types of incentives: giveaways and lotteries.

Incentives have been modeled with a similar structure as reporting recriminations. The effect of incentives increases the likelihood of reporting as opposed to decreasing it.

Given the relatively light value of incentives such as giveaways and lotteries, we assume that incentives are quickly forgotten. In the model it takes three months to forget an incentive.

*Rate of Increase in Reporting Incentives = (Event Reporting Rate+Incident Reporting Rate)*Effectiveness of Incentives*

Rate of Forgetting Reporting Incentives = Reporting Incentives/Time to Forget Reporting Incentives

Effect of Reporting Incentives on Fraction of Reported Incidents = Table of Effect of Reporting Incentives on Fraction of Reported Incidents(Reporting Incentives/Normal Reporting Incentives)

Effect of Reporting Incentives on Fraction of Reported Events = Table of Effect of Reporting Incentives on Fraction of Reported Events(Reporting Incentives/Normal Reporting Incentives)

Keeping staff 'in the loop'

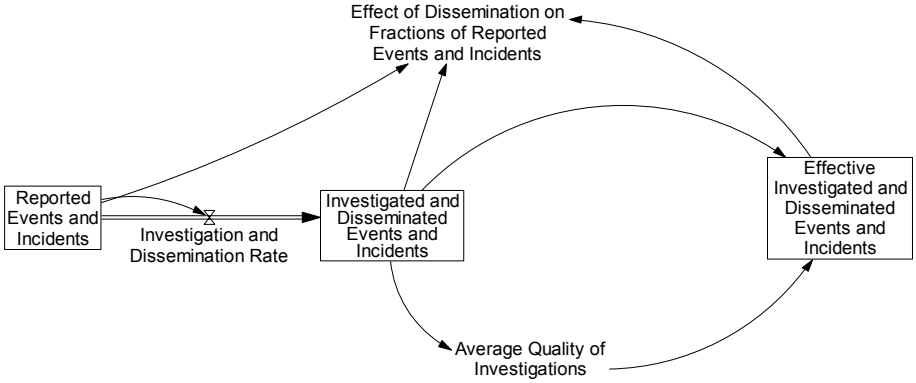


Figure 9. The keeping staff 'in the loop' effect.

Feedback to the reporter of an incident or an event is crucial to motivate for reporting in the future. If reports are perceived to lead to improvements, motivation to report increases. Similarly, if their reporting is not perceived to lead to improvements, motivation decreases. Johnson calls this effect "*keeping staff 'in the loop'*" (Johnson 2003).

Effect of Dissemination on Fractions of Reported Events and Incidents = Effective Investigated and Disseminated Events and Incidents/(Investigated and Disseminated Events and Incidents+Reported Events and Incidents)

This effect may not apply only to feedback within organizations but also between organizations. An example is Taiwan's use of mandatory aviation incident reporting to the Taiwanese Civil Aviation Administration (CAA). According to Lee and Weitzel (2005) the CAA's aviation incident database contains considerable amounts of incident data, but due to lack of funding, the data has not been used for trend analysis. Furthermore, the data has been inaccessible in nature and thus have not been used by Taiwanese air carriers or Taiwan's Aviation Safety Council (a Taiwanese aviation incident investigation group).

Learning Effects

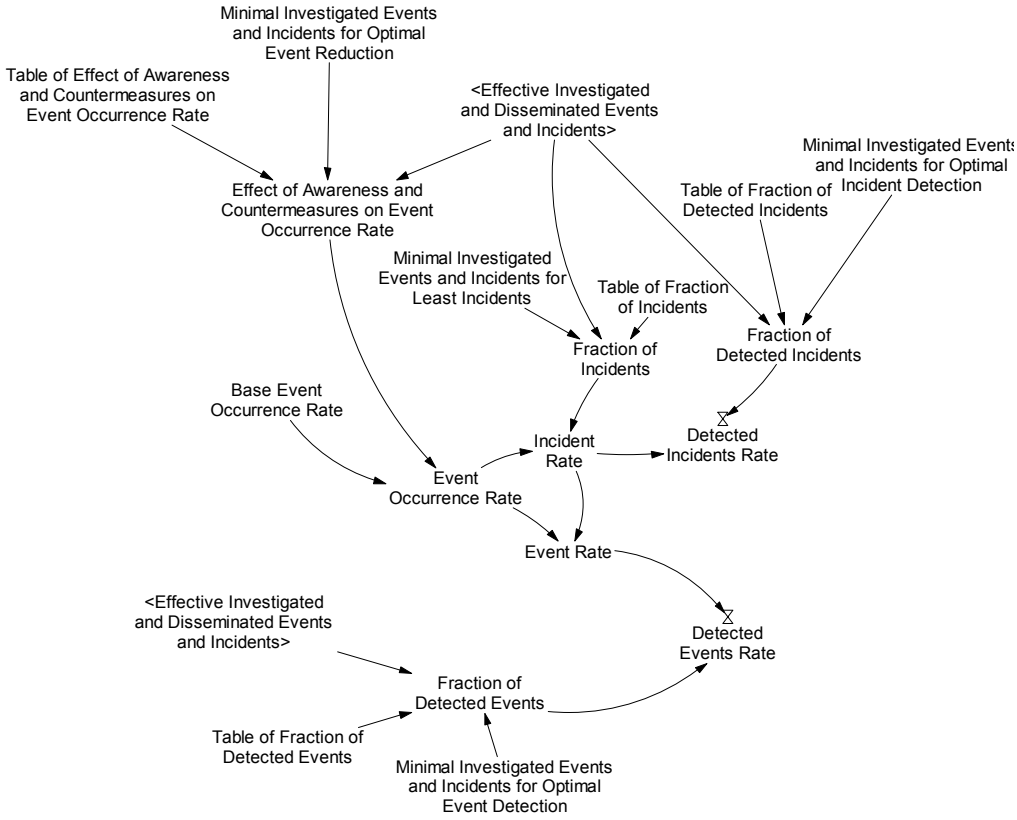


Figure 10. Factors affecting learning from incidents and events.

Investigated incident and event reports allow decision makers to implement countermeasures such as physical barriers or changed routines. Disseminating information about investigated events and incidents to staff, in general raises awareness about safety issues. Increased awareness and countermeasures reduces the amount of events occurring.

Effect of Awareness and Countermeasures on Event Occurrence Rate = Table of Effect of Awareness and Countermeasures on Event Occurrence Rate(Effective Investigated and Disseminated Events and Incidents/Minimal Investigated Events and Incidents for Optimal Performance)

Increased awareness may also serve to reduce the number of events that become incidents. Increased knowledge about security may allow staff to take action to mitigate events, keeping them from becoming incidents.

Fraction of Incidents = Table of Fraction of Incidents(Effective Investigated and Disseminated Events and Incidents/Minimal Investigated Events and Incidents for Optimal Performance)

When staff becomes more knowledgeable about security matters they should also get better at detecting incidents and events.

Fraction of Detected Incidents = Table of Fraction of Detected Incidents(Effective Investigated and Disseminated Events and Incidents/Minimal Investigated Events and Incidents for Optimal Performance)

Fraction of Detected Events = Table of Fraction of Detected Events(Effective Investigated and Disseminated Events and Incidents/Minimal Investigated Events and Incidents for Optimal Event Detection and Reporting Performance)

Full Event and Incident Learning Structure

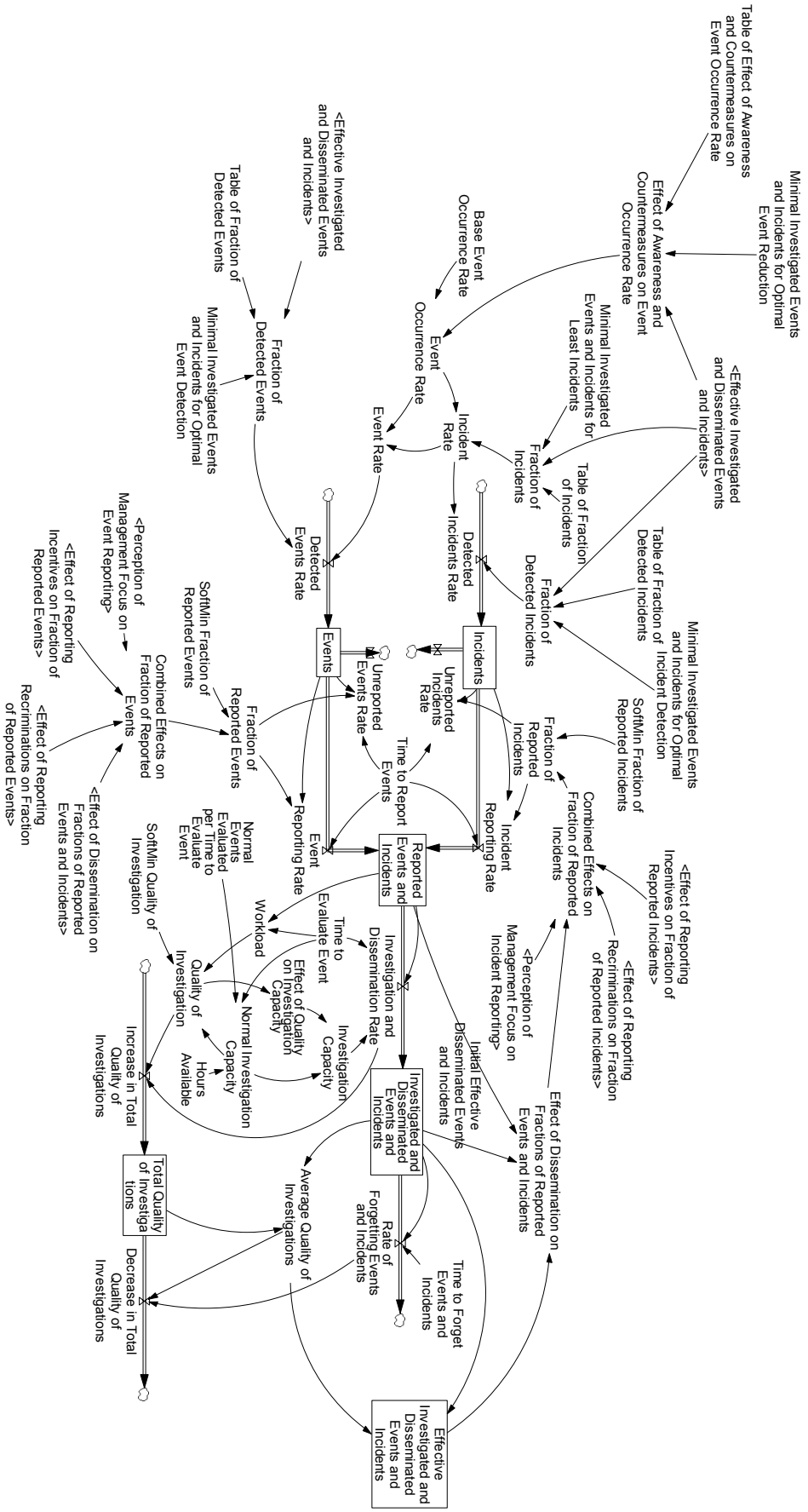


Figure 11. Full incident and event reporting structure.

4. Simulation Runs

In this simulation we assume that an incident learning system exists prior to the start of the simulation. The model is therefore initialized in equilibrium. Management initially has full focus on incident reporting, perceiving it as important. 'Management Focus on Incidents' is initially unity. Event reporting is perceived as less important. 'Management Focus on Events is set to 0.25.

Although management focuses on incident reporting, the reporting climate is not good. 'Effectiveness of Recriminations' is set to unity. An incentive scheme also exists. 'Effectiveness of Incentives' is set to unity.

Scenario No.	Scenario Name	Increased Incentives	Reduced Recriminations	Limited Resources	Management focus on Events
1	rR		X		
2	iI	X			
3	L			X	
4	MFE				X
5	MFE rR		X		X
6	MFE iI	X			X

Table 1. Simulation scenarios

Incentives and Recriminations

In the rR scenario the effectiveness of the recriminations are reduced by 75% in month 3. We assume that recriminations are not completely removed as there may several factors affecting whether or not one or more recriminations occur. For example, even if management succeeds in removing their recriminations, there may be some peer pressure left from colleagues.

'Fraction of Incidents' gradually falls throughout the simulation. 'Fraction of Reported Incidents' shows the opposite behavior, it gradually increases. The diverging behavior of 'Fraction of Incidents' and 'Fraction of Reported Incidents' have one important consequence. As we can see the 'Incident Reporting Rate' increases for twelve months, before gradually dropping to a level slightly higher than Base Run. The reduction in 'Incident Rate' is not directly visible, as managers do not have access to this rate. They can only estimate it based on what is actually reported.

Scenario iI sees an increase in the effectiveness of incentives by 75% in month three. Initially there is a gradual decrease in 'Fraction of Incidents'. However after about fifteen months the behavior stabilizes. 'Fraction of Reported Incidents' increases and reaches a top in month 9, where after it falls slightly, This behavior is caused by the buildup of recriminations. As more incidents are reported, more recriminations are accumulated, limiting improvement in 'Incident Rate'. Although there is an improvement, it can not be seen directly this time either.

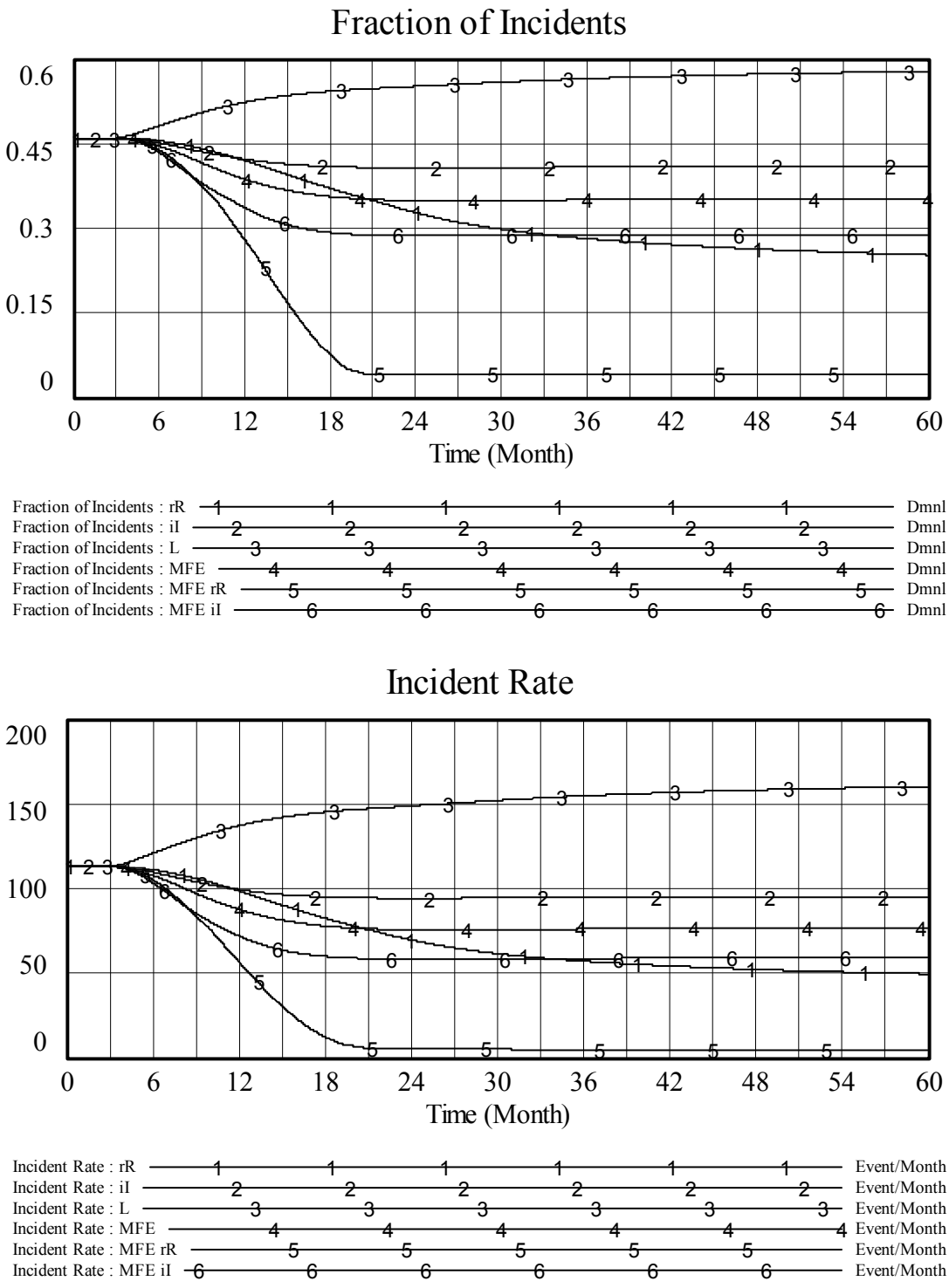


Figure 12. Fraction of Incidents and Incident Rate

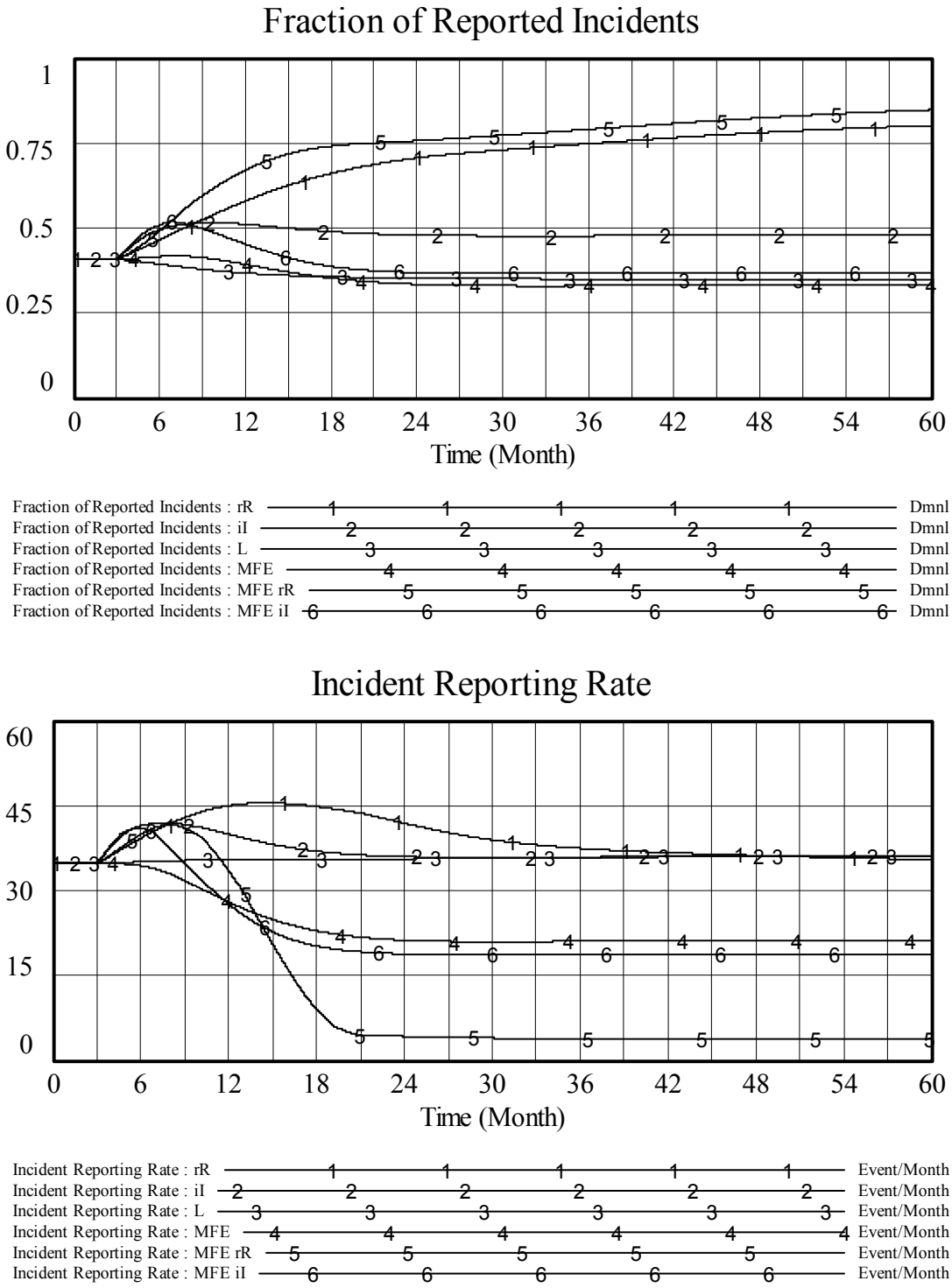


Figure 13. Fraction of Reported Incidents and Incident Reporting Rate

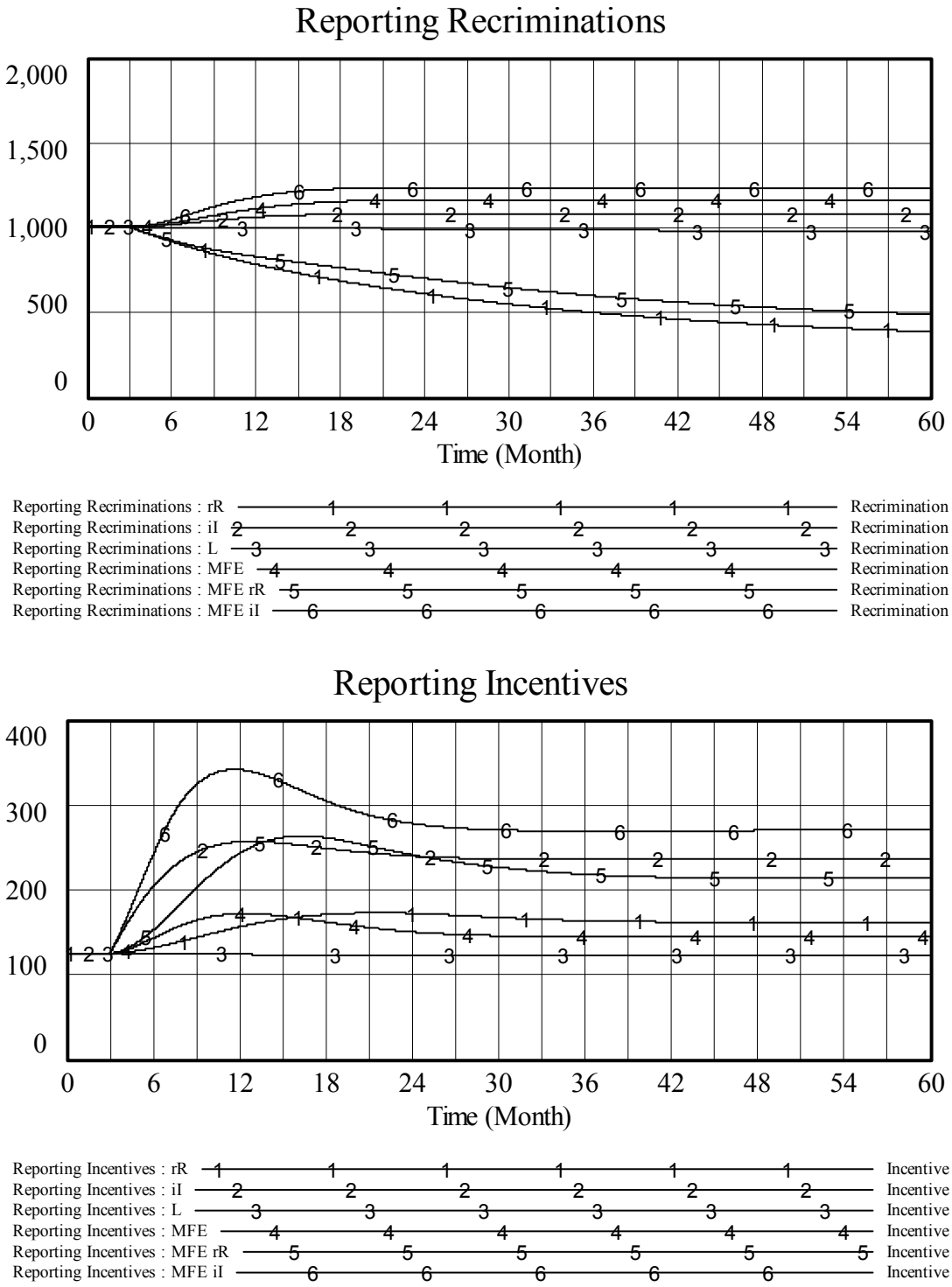
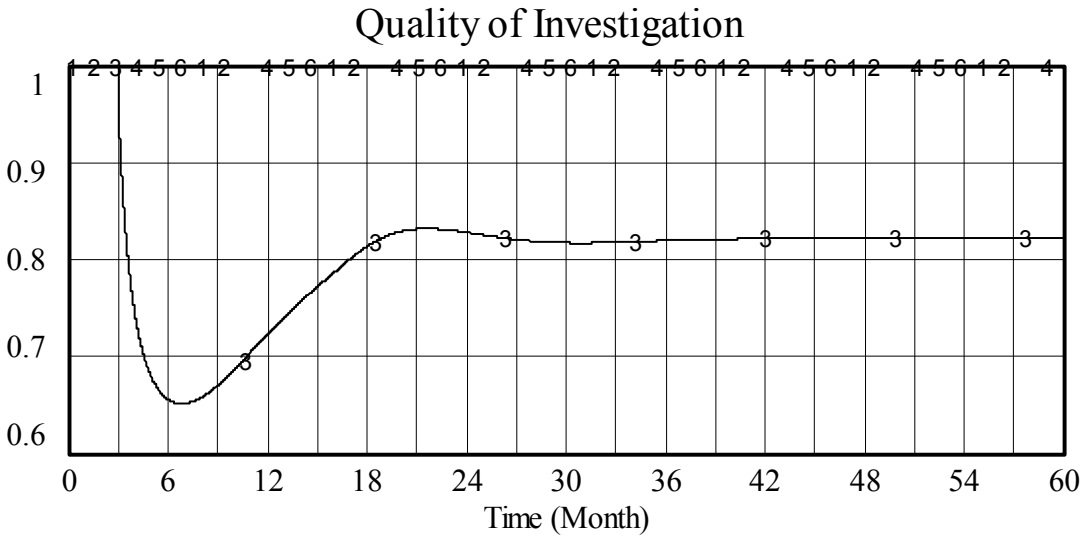
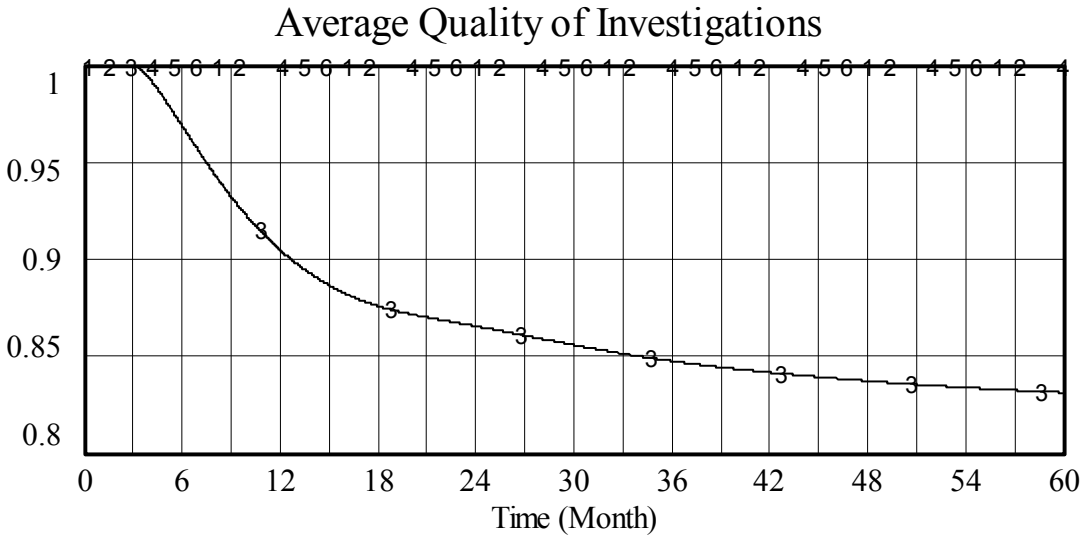


Figure 14. Reporting recriminations and incentives.



Quality of Investigation : rR
Quality of Investigation : iI
Quality of Investigation : L
Quality of Investigation : MFE
Quality of Investigation : MFE rR
Quality of Investigation : MFE iI



Average Quality of Investigations : rR
Average Quality of Investigations : iI
Average Quality of Investigations : L
Average Quality of Investigations : MFE
Average Quality of Investigations : MFE rR
Average Quality of Investigations : MFE iI

Figure 15. Investigation Quality Graphs

Inadequate Resources

In the L scenario, in month three, investigative resources are reduced to initially 95% of the needed resources. A backlog of uninvestigated incidents starts to build up, causing an even higher workload. The quality of investigations is reduced to process incidents faster. The falling 'Average Quality of Investigations' is perceived through the ´keep staff in the loop effect'. Subsequently fewer reports come in, as can be seen in falling 'Fraction of Reported Incidents'.

Since fewer lessons learned are now produced the 'Fraction of Incidents' increases. Although this increase is substantial, it is offset by the decrease in 'Fraction of Reported Incidents'. Hence, only a small increase can be seen in 'Incident Reporting Rate'. A situation that has actually become much worse can be perceived as one that has not really changed much.

Management Focus on Events

The previous three runs focused on changing the basic conditions for incident reporting. We now move our focus towards event, or near-miss, reporting. The following scenarios simulate management's elevation of event reporting to the same status as incident reporting. In the MFE, MFE iI, and MFE rR scenarios 'Management Focus on Events' is increased from 0.25 to 1.0 in month three.

In the MFE scenario an increased focus on event reporting leads to an eight month increase in 'Event Reporting. The reported events represent additional lessons learned. Since the basis for learning is much greater, 'Incident Rate' is reduced. However, as in the iI scenario, recriminations start to accumulate as more event reports come in. This limits the improvement in 'Incident Rate' and it stabilizes after month 18.

The reduction in 'Incident Rate' is mirrored by 'Incident Reporting Rate'. However, we may still be deceived. Although the reporting rate is dropping it is not only due to a reduced 'Incident Rate'. The increased focus on events causes the 'Fraction of Reported Incidents' to initially slightly increase. However, the build up of recriminations soon reverses the development. Eventually the 'Fraction of Reported Incidents' stabilizes well below what it initially was. We may thus be lead to believe that the improvement is greater than it really is.

If we combine increased focus on events with reduced recriminations a favorable outcome emerges (scenario MFE rR). As in the rR scenario 'Fraction of Reported Incidents' increases, causing an initial increase in 'Incident Reporting Rate'. After about eight months the increase turns into a decrease and the fall is continued until month 21. The 'Incident Reporting' stabilizes well below what it initially was and this is reflected in the 'Incident Reporting Rate'. Absence of recriminations combined with the larger basis for learning provided by event reporting combines to create a highly effective system.

Increasing the incentives instead of reducing recriminations is followed by improvement in 'Incident Rate'. However, here too a buildup of recriminations limits the reduction. The improvement is still better than focusing solely on incentives without focusing on event reporting.

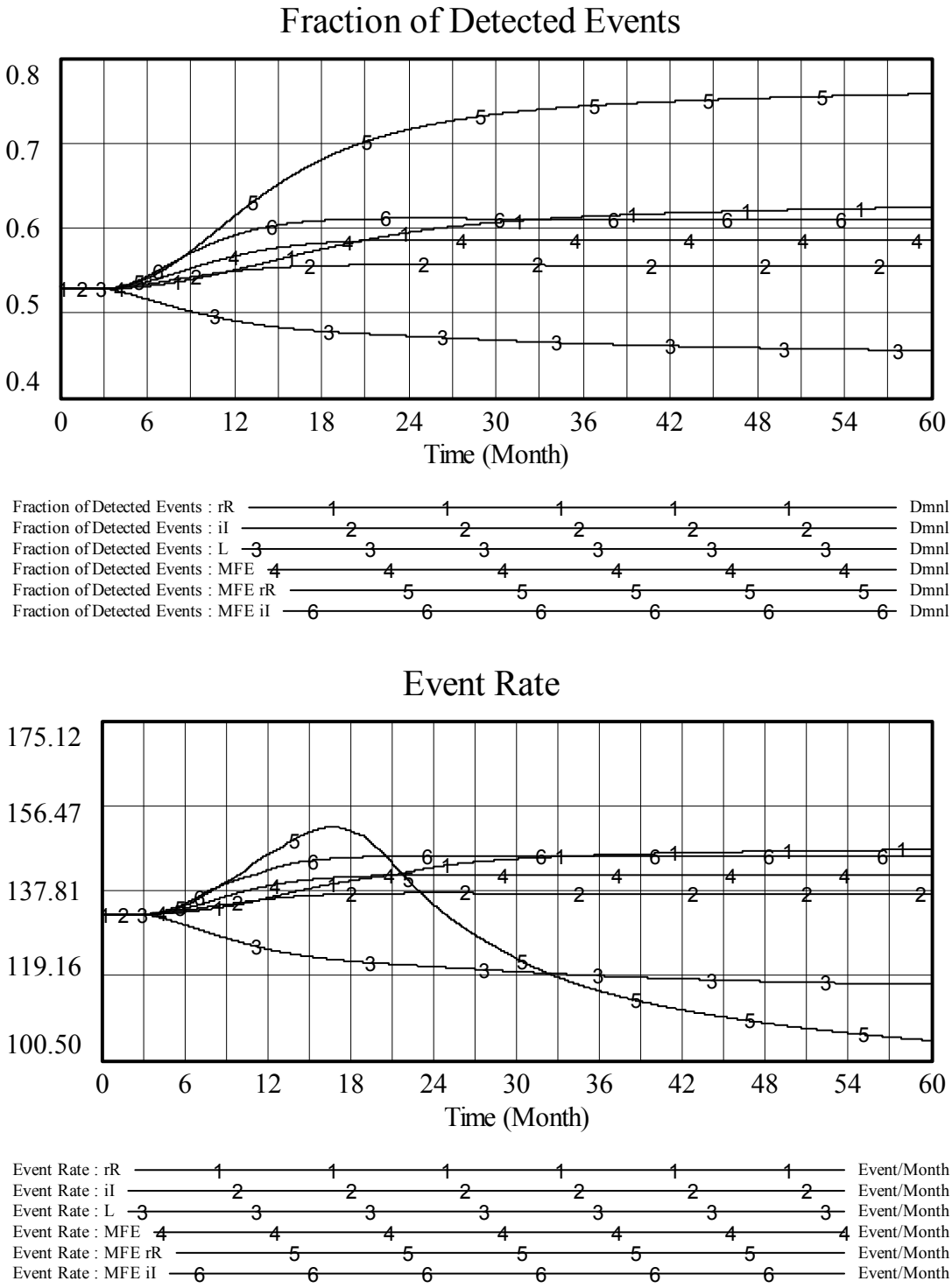


Figure 16. Fraction of Detected Events and Event Rate

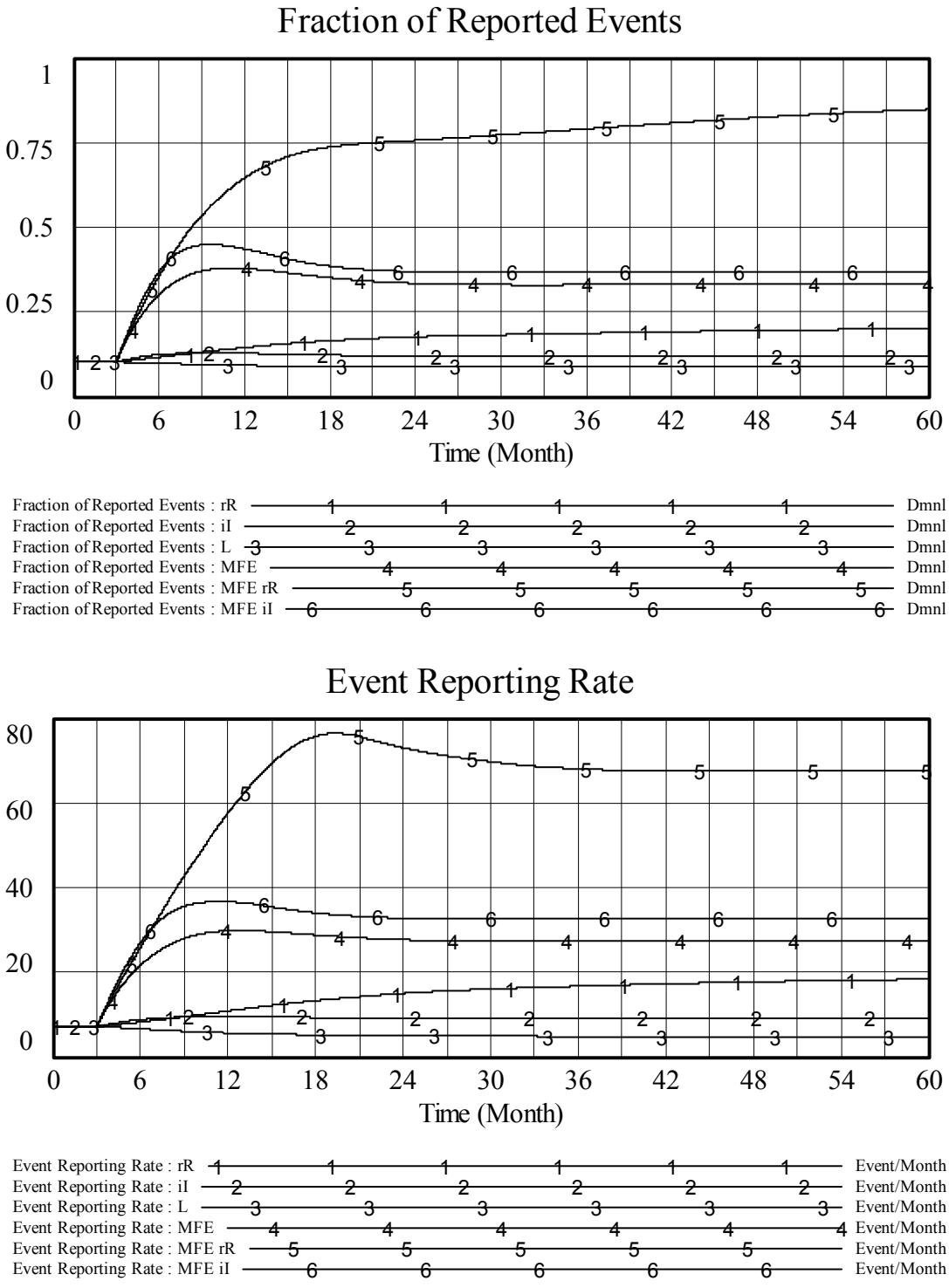


Figure 17. Fraction of Reported Events and Event Reporting Rate.

Effect of Recriminations

Although incentives may seem to be a quick and easy way to improve reporting of incidents and events, the iI and MFE iI scenarios indicate that increasing incentives without working to improve the reporting climate may be unwise. An incentive program, which may be expensive, may turn out to be ineffective.

The Relationship between Incidents and Events

Comparing 'Incident Reporting Rate' and 'Event Reporting Rate' in the preceding simulations reveals diverging behavior. If more events are reported, fewer incidents tend to be reported. This is an effect that has been shown empirically by Jones, Kirchsteiger and Bjerke (1999). The model also shows that in the case of highly effective policies that reduce underreporting, both incident and event reporting may increase for a time. However, when underreporting has been sufficiently reduced, the reduction of actual incidents becomes visible. A study of two Danish factories supports these results. The introduction of an incident reporting system at one of the factories lead to a six month increase in incident reports, followed by a decrease to a lower level than before the introduction (Nielsen, Carstensen, and Rasmussen 2006). The authors attributed the initial increase to probable reduction in underreporting.

Incident Reporting Rate as Indicator of Incidents

The preceding scenarios show that the incident reporting rate is inadequate as a single indicator of incidents. In scenarios rR and iI the rate of reported incidents eventually returns to baseline while the actual incident rate ends up lower than the baseline. In L almost no change can be seen although the incident rate is actually increasing. In the case of scenario MFE improvement in incident rate can be perceived through the incident reporting rate, but the magnitude of the improvement is masked as the fraction of reported incidents go down. These simulation results indicate that it is difficult to use the incident rate to measure whether the system has changed for worse or better. Other indicators should be used in parallel with reporting rates.

5. Conclusions and Future Work

The system dynamics model of a safety incident learning system and the literature which it is based upon show that there are many challenges one must grapple with when implementing well-functioning incident learning systems. The true state of the system may be invisible to the decision makers, as rising incident reporting rates may be both good and bad, and in many cases misleading. Thus it is not possible to rely on incident reporting rates alone. As we have seen, the relationship between event and incident reporting rates may indicate the state of the system. However, we believe that it is also necessary to measure the safety culture itself and the severity of the incidents. Falling severity should be a sign of improving safety (Cooke and Rohleder 2006).

The simulation also indicates that it may be more productive to focus on improving the reporting culture by removing recriminations rather than increasing incentives. The recriminations effectively works as a brake, limiting the growth of lessons learned.

Although the above lessons are the result of a model based on safety literature, we believe they are also important for organizations that wish to employ incident learning systems to improve their information security. The predominant technical focus in the

field of information security largely overshadows equally important human factors. Furthermore, humans are the users of the systems, and in many cases they will be the first to detect incidents, events or the symptoms of them. Well functioning incident learning systems helps users learn about security and why it is necessary. It helps them to better recognize attacks and learn how to mitigate them.

Safety hazards may be felt to be more real than information security hazards. After all, a beam that falls down may crush you, while a computer that stops, only stops. It may therefore be harder to motivate people to care about security, but it is no less important. As mentioned earlier, an increasing amount of real time computer systems are spreading throughout factories. These systems are also increasingly networked together, creating new security hazards that are also potential safety hazards. In addition, security incidents are expensive. A single incident may not necessarily cost much in itself, but when incidents accumulate they represent a large expense. Identifying all possible security vulnerabilities prior to the startup of a new networked system is incredibly hard, if not impossible. It is therefore proactive to assume that incidents will happen and to use incident learning systems to mitigate risk.

The model presented in this paper is based mostly on safety literature. As such it represents our starting hypothesis for how incident reporting systems in information security should work. Information security does have some challenges that do not exist in safety. For example, exponentially increasing attack volumes (Wiik, Gonzalez, and Kossakowski 2004). To better understand the specific challenges faced in the realm of information security we are currently undertaking case studies in three Norwegian organizations.

References

Anderson, David J., and Craig S. Webster. 2001. A System Approach to the Reduction of Medication Error on the Hospital Ward. *Journal of Advanced Nursing* 35 (1):34-41.

Cooke, David L. 2003. Learning from Incidents. In *From Modeling to Managing Security*, edited by J. J. Gonzalez. Kristiansand, Norway: Norwegian Academic Press.

———. 2003. A system dynamics analysis of the Westray mine disaster. *System Dynamics Review* 19 (2):139-166.

———. 2004. The Dynamics and Control of Operational Risk, Haskayne School of Business, University of Calgary, Alberta.

Cooke, David L., and Thomas R. Rohleder. 2006. Learning from Incidents: from Normal Accidents to High Reliability. *System Dynamics Review* 22 (3):213-239.

Gonzalez, Jose J. 2005. Towards a Cyber Security Reporting System - A Quality Improvement Process. In *Computer Safety, Reliability, and Security*, edited by B. A. G. Rune Winther and G. Dahll. Heidelberg: Springer.

Johnson, Chris. 2003. *Failure in Safety-Critical Systems: A Handbook of Incident and Accident Reporting*. Glasgow, Scotland: Glasgow University Press.

Jones, Simon, Christian Kirchsteiger, and Willy Bjerke. 1999. The Importance of Near Miss Reporting to Further Improve Safety Performance. *Journal of Loss Prevention in the Process Industries* 12 (1):59-67.

Kjellén, Urban. 2000. *Prevention of Accidents Through Experience Feedback*. London and New York: Taylor & Francis.

Lee, Ping I., and Thomas R. Weitzel. 2005. Air Carrier Safety and Culture: An Investigation of Taiwan's Adaptation to Western Incident Reporting Programs. *Journal of Air Transportation* 10 (1).
Nielsen, Kent J., Ole Carstensen, and Kurt Rasmussen. 2006. The Prevention of Occupational Injuries in Two Industrial Plants Using an Incident Reporting Scheme. *Journal of Safety Research* 37 (5):479-486.
Nyssen, A. S., S. Aunac, M. E. Faymonville, and I. Lutte. 2004. Reporting Systems in Healthcare From a Case-by-Case Experience to a General Framework: An Example in Anaesthesia. *European Journal of Anaesthesiology* 10 (21):757-765.
Phimister, James R, Ulku Oktem, Paul R. Kleindorfer, and Howard Kunreuther. 2003. Near-Miss Incident Management in the Chemical Process Industry. *Risk Analysis* 23 (3):445-459.
Schneier, Bruce. 2000. *Secrets & Lies: Digital Security in a Networked World*: Wiley.
Sterman, John D. 2000. *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Boston: Irwin McGraw-Hill.
Wiik, Johannes, Jose J. Gonzalez, and Klaus-Peter Kossakowski. 2004. Limits to Effectiveness in Computer Security Incident Response Teams. In *23rd International Conference of the System Dynamics Society*. Oxford.