Safety-I and Safety-II: The Past and Future of Safety Management

Erik Hollnagel

Professor
University of Southern Denmark

Chief Consultant
Center for Quality, RSD (DK)

hollnagel.erik@gmail.com

Understanding and predicting failures

FRAM

'My dear friend Copperfield,' said Mr. Micawber,
...
'Accidents will occur in the best-regulated families;
...
they may be expected with confidence,
and must be borne with philosophy.'

Systems; organisations

The probability that a specified event will occur.

The degree of certainty by which accidents can be expected.

The principles (models and theories) for describing and analysing accidents

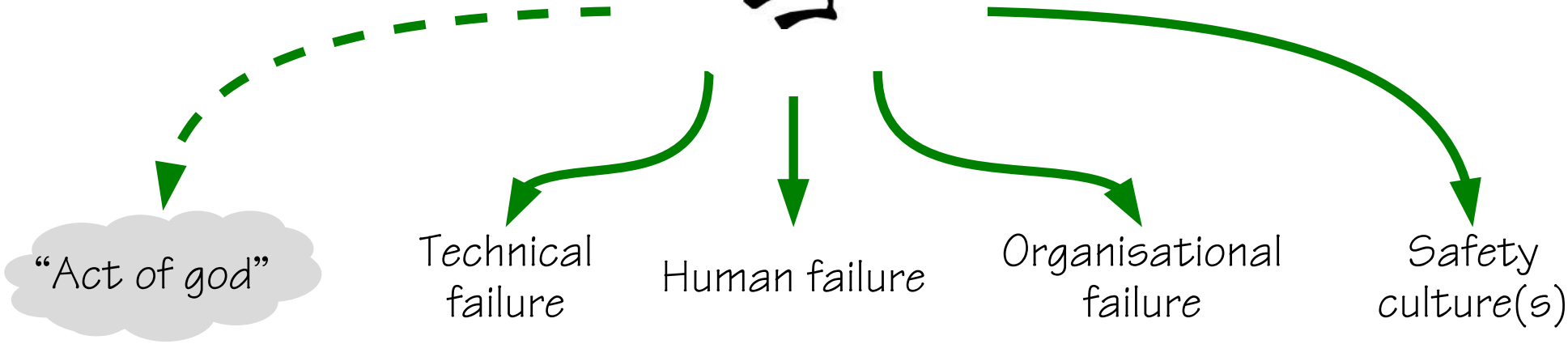
The lessons learned and the approaches to system design (prevention, protection).

Charles Dickens
David Copperfield (1850)
Chapter 28

Understanding a complicated world

FRAM

Accidents, incidents, breakdowns, disruptions,

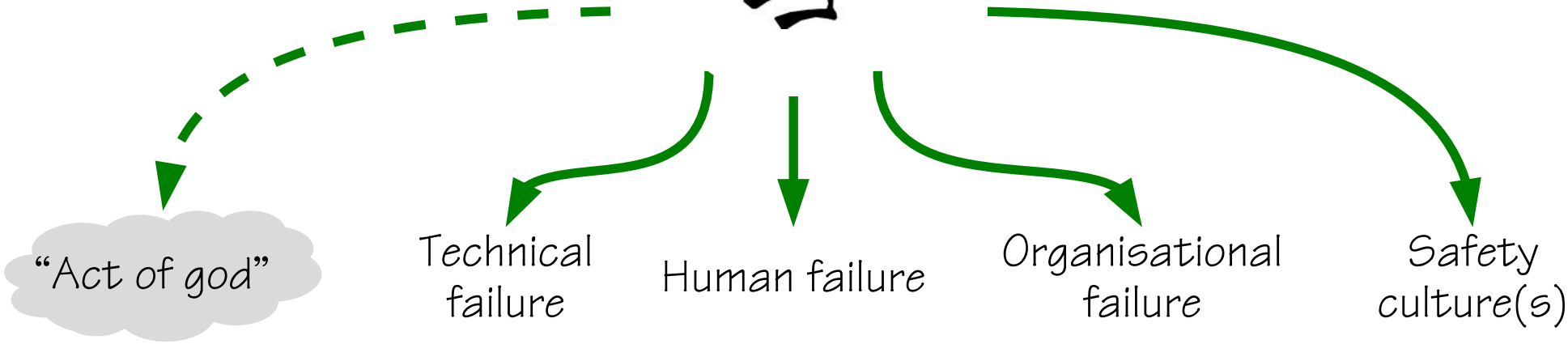


The types of causes may change over time, but we still believe in causality

Understanding a complicated world

FRAM

Accidents, incidents, breakdowns, disruptions,



The types of causes may change over time, but we still believe in causality

American National Standards Institute

Safety: Freedom from unacceptable risk.

Risk. An estimate of the probability of a hazard-related incident or exposure occurring and the severity of harm or damage that could result.

Acceptable Risk. That risk for which the probability of an incident or exposure occurring and the severity of harm or damage that may result are as low as reasonably practicable (ALARP) in the setting being considered.

Hazard. The potential for harm.

As Low As Reasonably Practicable (ALARP). That level of risk which can be further lowered only by an increase in resource expenditure that is disproportionate in relation to the resulting decrease in risk.

Safety: Freedom from unaffordable harm.

The meaning of safety

FRAM

How can it be done?

How much risk is acceptable?

How much risk is affordable?

What can go wrong?

SAFETY = FREEDOM FROM UNACCEPTABLE RISK

Prevention of unwanted events

Protection against unwanted outcomes

Normal functioning

Unexpected event

Unwanted outcome

LIFE
PROPERTY
MONEY

Accidents, incidents, …

The Boeing 787 Dreamliner

Region Syddanmark

During December 2012 and January 2013, a number of aircraft had problems with their batteries, which either were damaged or caught fire. On January 16, the FAA issued an emergency airworthiness directive grounding U.S.-based Boeing 787s.

More than 500 engineers and experts spent more than 200,000 hours to find a root cause, but failed to do so.

Improved batteries that work at a lower temperature, in stainless steel boxes with ventilation directly to the outside.

Prevention of unwanted events

Protection against unwanted outcomes

Normal performance

Unexpected event

Unwanted outcome

LIFE
PROPERTY
MONEY

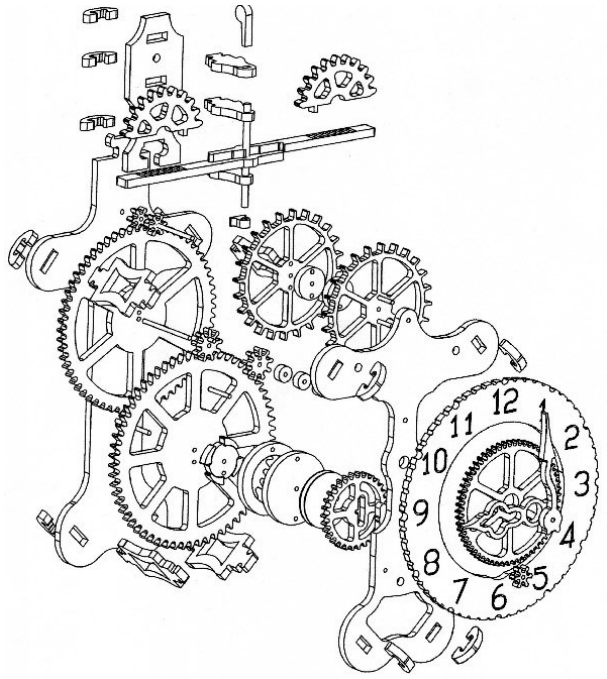
Boeing's estimate p(battery failure) = 10^{-7}, but there were two failures in the first 52,000 flight hours ($p = 3.8\ 10^{-5}$).

Accidents, incidents, …

Deconstruction

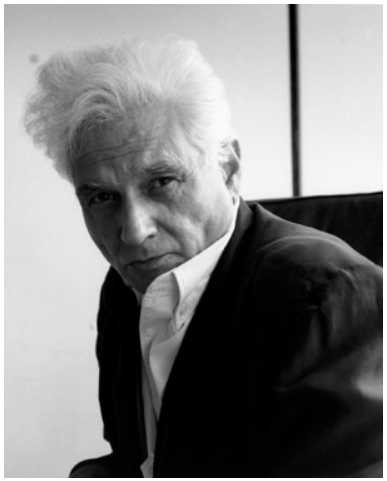
Deconstruction: a theory of criticism (usually of literature or film) that seeks to expose deep-seated contradictions in a work by delving below its surface meaning.

In the context of physical construction, deconstruction is the selective dismantlement of building components.



To disassemble something, in order to understand what it is "made of" and how it works.

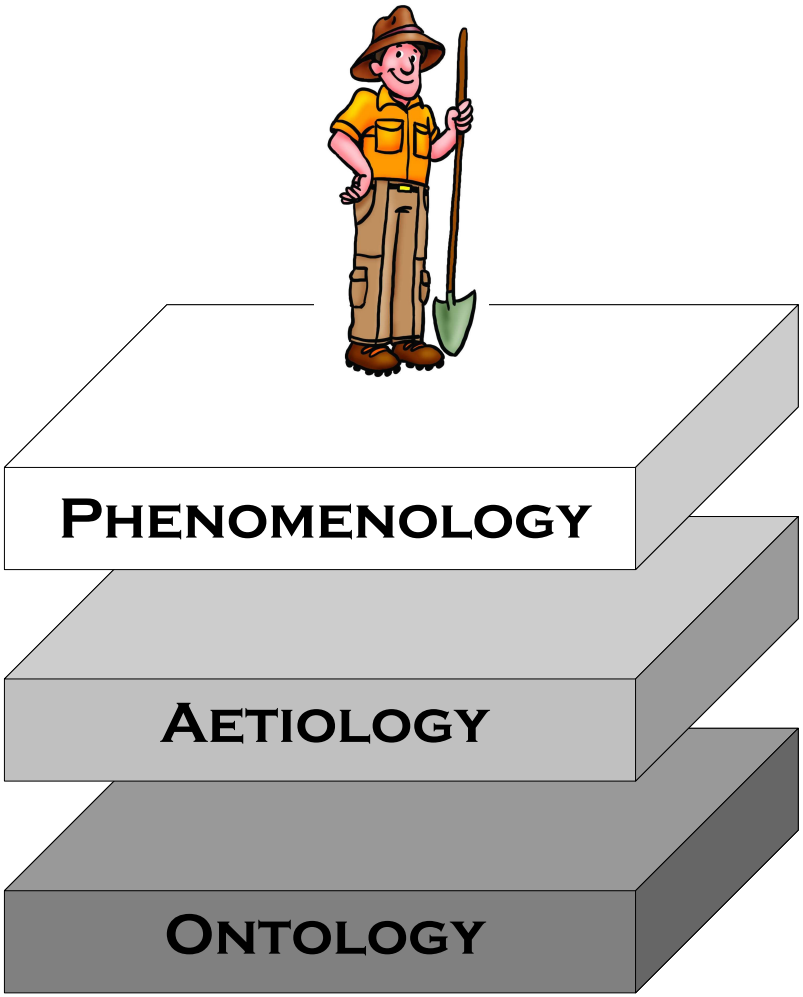
Issue: What are the assumptions behind safety?



Jacques Derrida (1930-2004)

Excavating through the layers of safety

FRAM

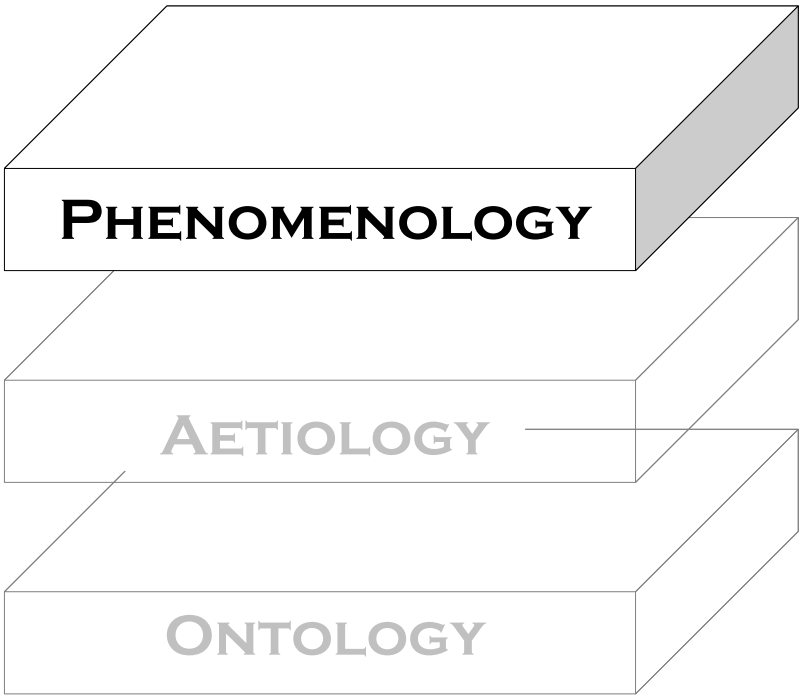


The observable characteristics (of safety).
The safety phenotype.

The origin or causes of the observable phenomena.
The safety genotype.

The nature and essential characteristics of safety.
What really goes on.

The phenomenology of Safety-I



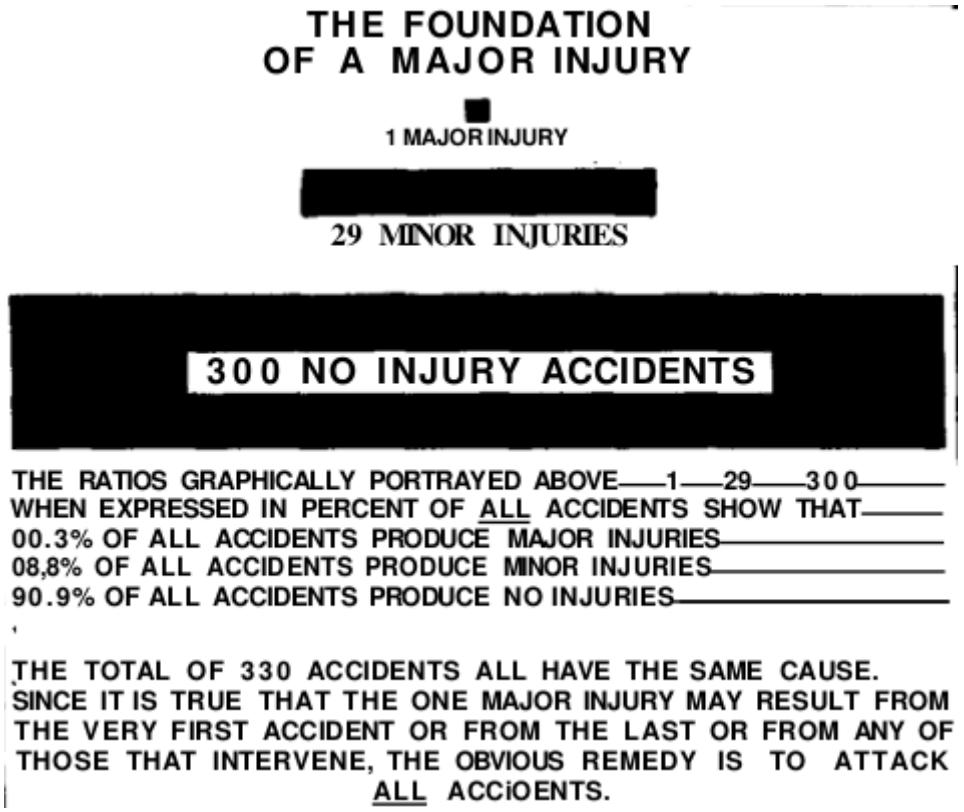
ICAO (International Civil Aviation Organization)
'the state in which the risk of harm to persons or of property damage is reduced to, and maintained at or below, an acceptable level through a continuing process of hazard identification and risk management.'

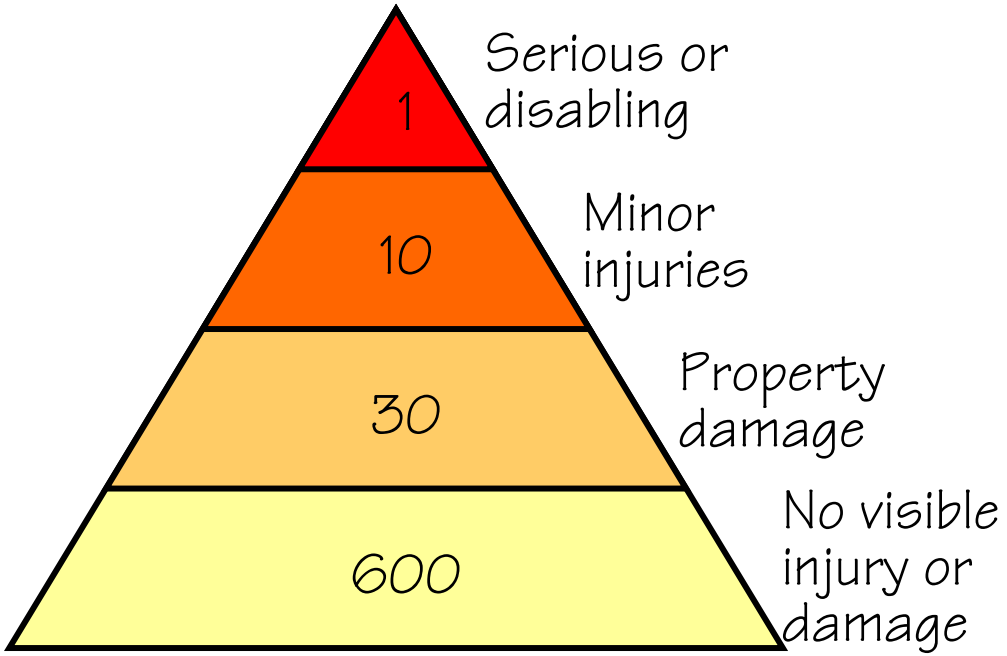
US AHQR (Agency for Healthcare Research and Quality)
'freedom from accidental injury'
'avoiding injuries or harm to patients from care that is intended to help them.'

Although safety is defined as the absence of adverse outcomes (and risks), the phenomenology is the presence of these – hence the absence of safety.

A higher level of safety is measured by a smaller number of adverse outcomes.

The accident pyramid





Source: Bird, F. (1974). Management guide to loss control. Atlanta, GA: Institute Press. Analysis of 1 753 498 reported accidents, representing 21 different industrial groups.

Risk Matrix (IATA) as a trade-off

IATA = International Aviation Transport Association

Likelihood or probability	Severity / Scope of damage				
	Insignificant	Minor	Moderate	Critical	Catastrophic
	No or minor injury or negligible damage	Minor injury or minor property damage	Serious but non-permanent injuries or significant property damage	Permanent disability or occupational illness or major property damage	May cause death or loss of property
Often	Medium	High	Substantial	Substantial	Substantial
Occasionally	Medium	High	High	Substantial	Substantial
Possible	Small	Medium	High	High	Substantial
Unlikely	Small	Medium	Medium	High	High
Practically impossible	Small	Small	Small	Medium	Medium

Small	Safety is largely guaranteed.
Medium	Safety is partially guaranteed, normal protective measures are required.
High	Safety is not ensured, protective measures are urgently required.
Substantial	Safety is not ensured, enhanced protective measures are urgently required.

Counting and understanding

If the numerator is 1 adverse outcome (accident) ...

$$\frac{\text{Numerator}}{\text{Denominator}}$$

... then the denominator is the average number of events without adverse outcomes.

$$\frac{1}{7{,}000{,}000}$$

Likelihood of being in a fatal accident on a commercial flight.

$$\frac{1}{20{,}000}$$

Core Damage Frequency for a nuclear reactor (per reactor year).

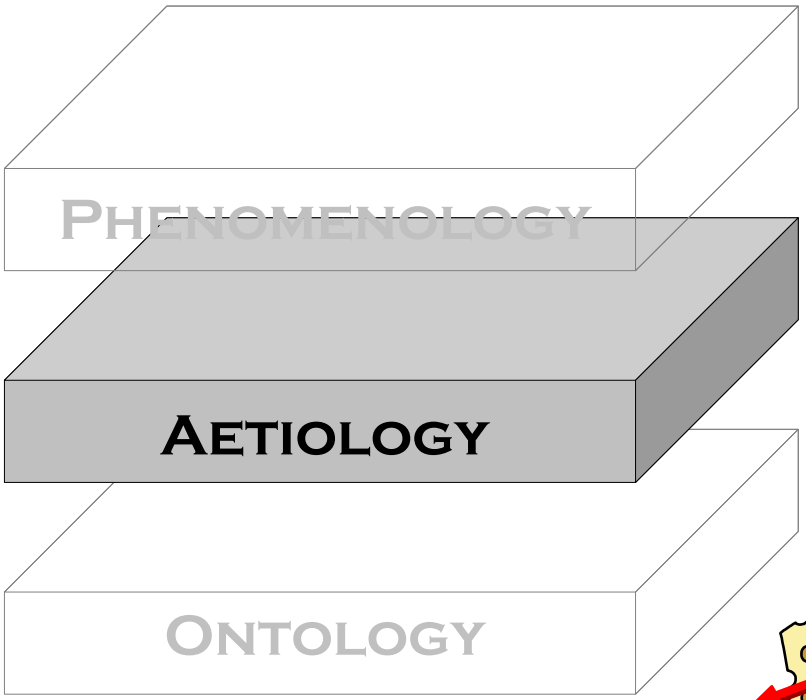
$$\frac{1}{10}$$

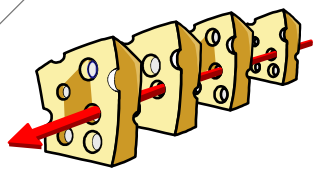
Likelihood of iatrogenic harm when admitted to a hospital.

The aetiology of Safety-I

FRAM

What is the origin of what we can observe?
How do accidents happen?





Accidents are the (natural) culmination of a series of events or circumstances, which occur in a specific and recognisable order.

Accidents are prevented by finding and eliminating possible causes.

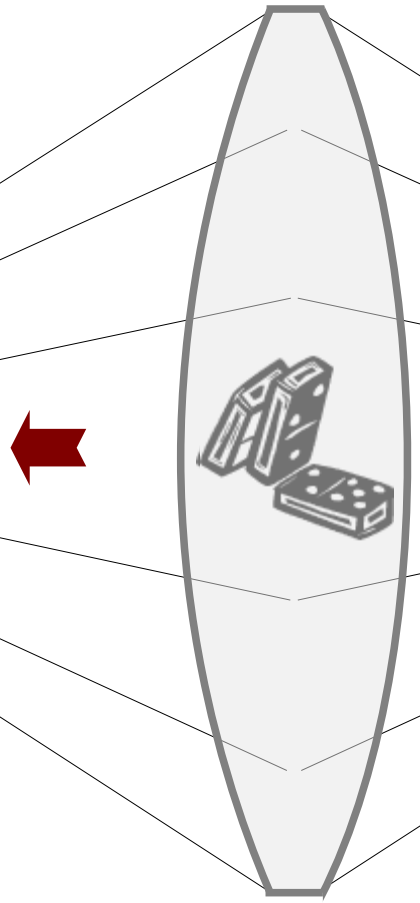
Accidents result from a combination of active failures (unsafe acts) and latent conditions (hazards).

Accidents are prevented by strengthening barriers and defences.

Risks as causal extension of failures

Region Syddanmark

Decomposable, simple linear models



If accidents are the culmination of a chain of events ...

... then risks can be found as the probability of component failures

Find the component that failed by reasoning backwards from the final consequence.

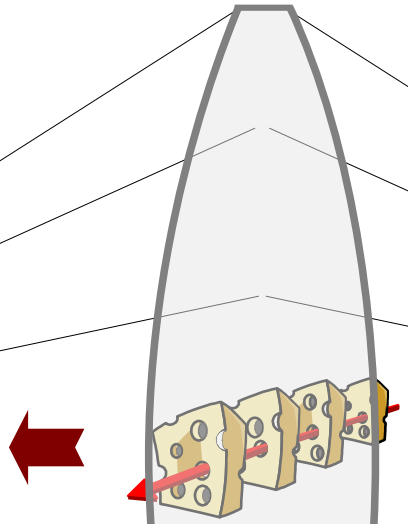
Find the probability that something "breaks", either alone or by simple, logical and fixed combinations.

Combinatorial (complex) linear model

Region Syddanmark

Decomposable, composite linear models

If accidents happen as a combination of active failures and latent conditions ...



... then risks are the likelihood of weakened defences in combination with active failures

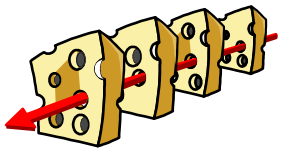
Look for how degraded barriers or defences combined with an active (human) failure.

Combinations of single failures and latent conditions, leading to degradation of barriers and defences.

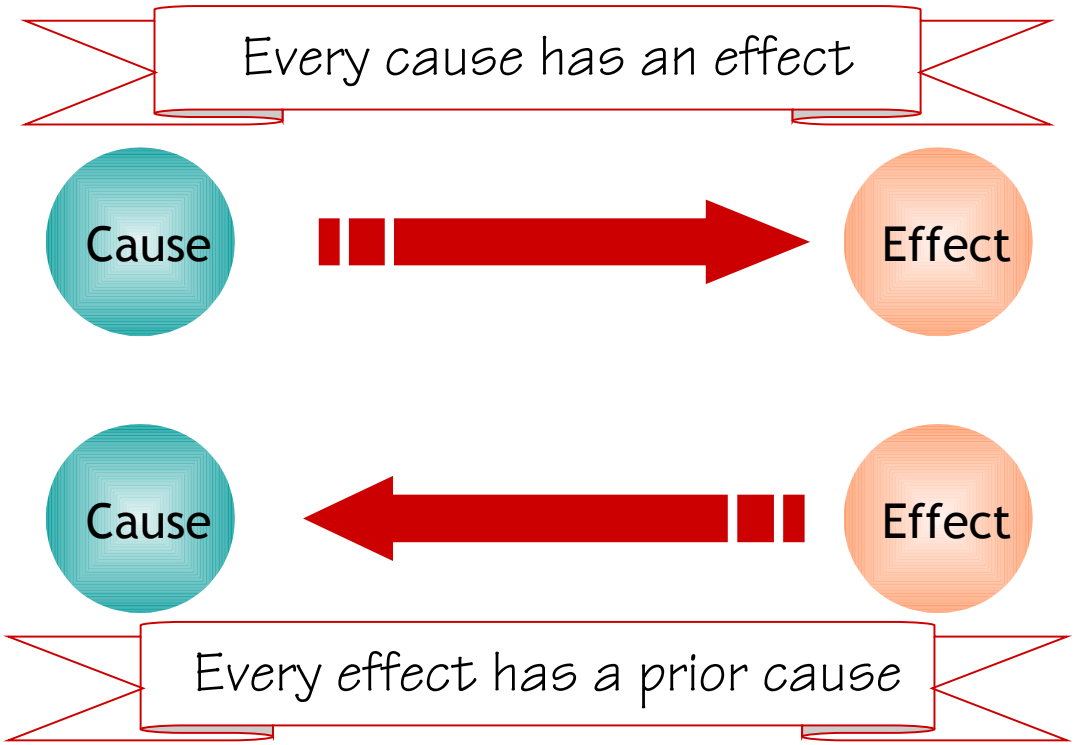
The causality credo

FRAM

(1) Adverse outcomes happen because something has gone wrong (causes).
(2) Causes can be found and treated.
(3) All accidents are preventable (zero harm).

Accident investigation		Risk analysis
Find the component that failed by reasoning backwards from the final consequence.		Find the probability that components "break", either alone or in simple combinations.
Accidents result from a combination of active failures (unsafe acts) and latent conditions (hazards).		Look for combinations of failures and latent conditions that may constitute a risk.

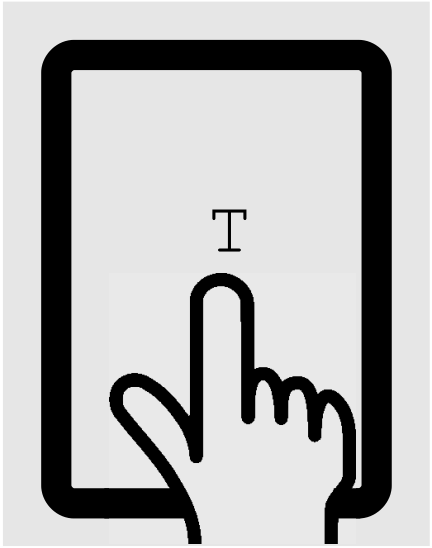
Principle of causation



David Hume 1711-1776

When we look about us towards external objects, and consider the operation of causes, we are never able, in a single instance, to discover any power or necessary connexion; any quality, which binds the effect to the cause, and renders the one an infallible consequence of the other. We only find, that the one does actually, in fact, follow the other.

The letter T



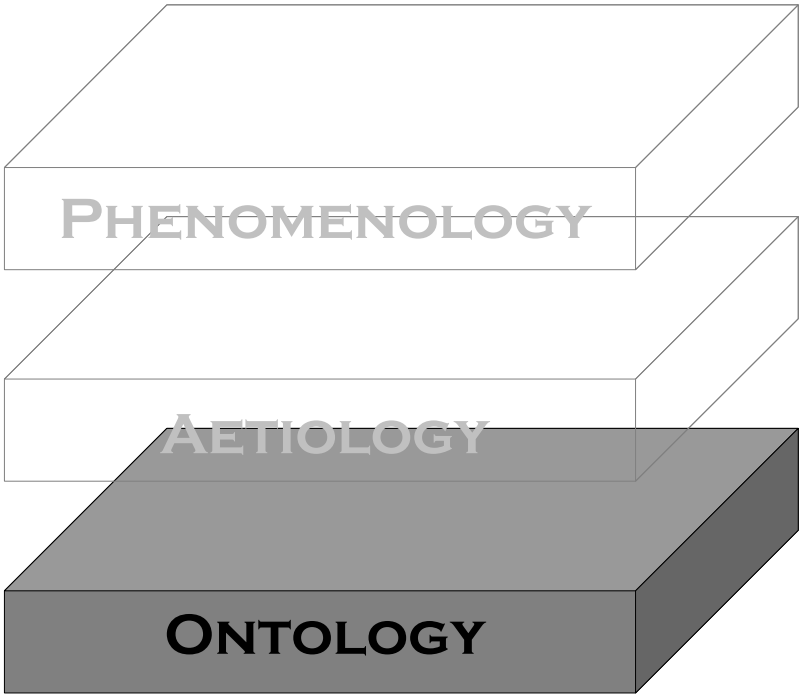
Can we understand and explain how the letter 'T' is produced?

Can we understand and explain why it may sometimes go wrong?

The ontology of Safety-I

FRAM

When tracing adverse outcomes back to their underlying causes, it is assumed that the "components" either have functioned correctly or have failed.

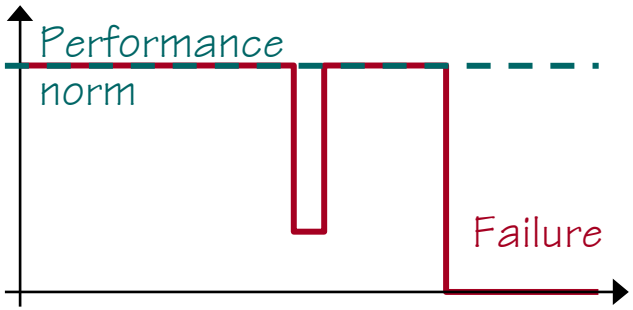


In the technical world, things usually function until they fail. When simple systems, such as a light bulb, fail, they are discarded and replaced by a new (and identical) one.

$$e \in E, e = \begin{cases} 1: \text{component or system functions} \\ 0: \text{component or system fails} \end{cases}$$

Human actions that fail are called "human errors".

Performance is bimodal: things either work correctly (as designed) or they fail.



Human error

FRAM

"More than seventy percent of all crashes of scheduled commercial aircraft are caused directly by 'controlled flight into terrain.'"

Federal Aviation Administration — 2001.

Sudden Stock Crashes Usually Caused by Human Error, SEC Says

International News - April 2011

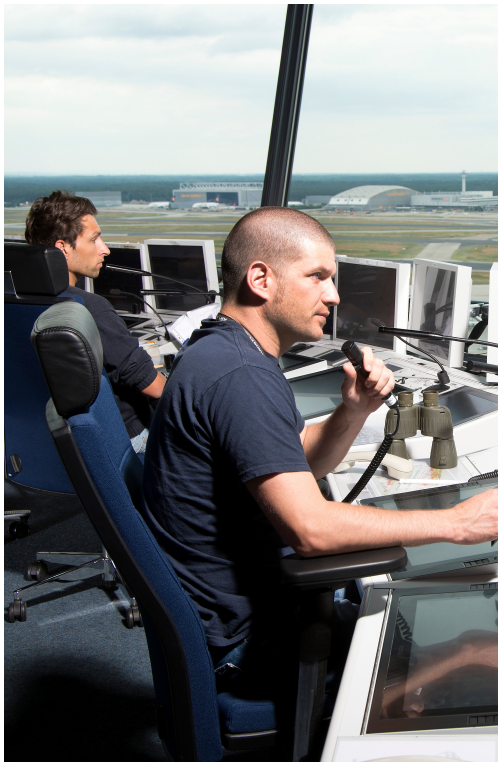
Human error accounts for 90% of road accidents

90.3%1 of crashes involved human error, such as risky driving behavior, inadvertent errors, and impaired states.

Foundation for Traffic Safety, 2006

Different process ➡ different outcome

FRAM



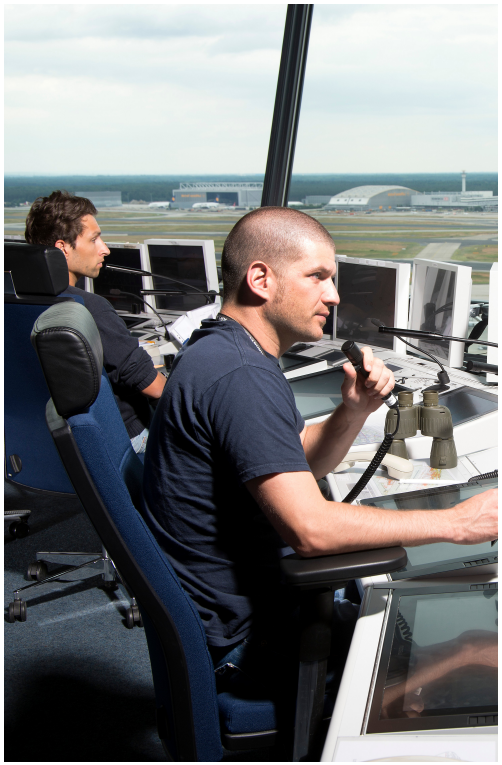
Function (work as imagined) ⟶ Success (no adverse events) — Acceptable outcomes

Hypothesis of different causes: Things that go right and things that go wrong happen in different ways and have different causes

Malfunction, non-compliance, error ⟶ Failure (accidents, incidents) — Unacceptable outcomes

Increasing safety by reducing failures

FRAM



Function (work as imagined) → Success (no adverse events) → Acceptable outcomes

"Identification and measurement of adverse events is central to safety."

Malfunction, non-compliance, error

Failure (accidents, incidents)

Unacceptable outcomes

"Find-and-fix"

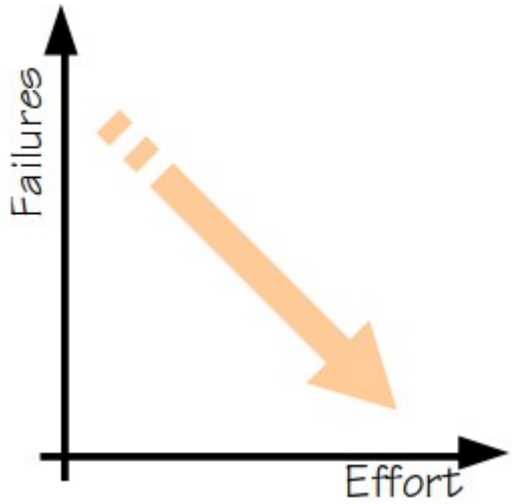
Safety-I – when nothing goes wrong

FRAM

Safety-I: Safety is the condition where the number of adverse outcomes (accidents / incidents / near misses) is as low as possible.

Safety is therefore defined by its opposite – by the lack of safety.

The lack of safety means that something goes wrong or can go wrong.



Safety-I requires the ability to prevent that something goes wrong.
Safety-I is reactive, and assumes that safety can be achieved by first finding and then eliminating or weakening the causes of adverse events.
Example: Root Cause Analysis (RCA).

Why only look at what goes wrong?

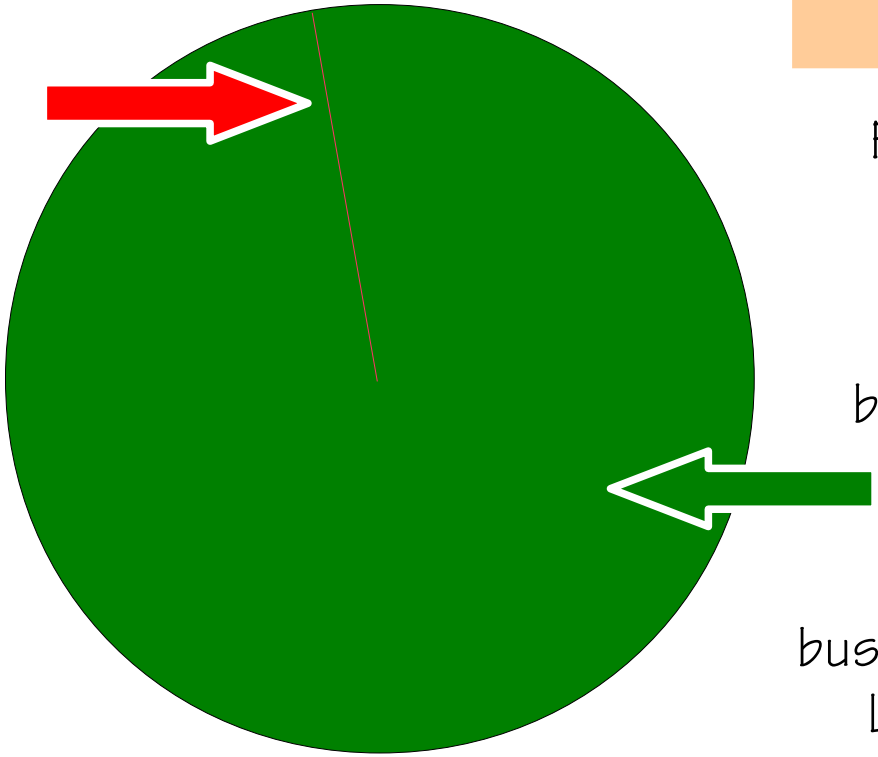
FRAM

Safety-I = Reduced number of adverse events.

Focus is on what goes wrong. Look for failures and malfunctions. Try to eliminate causes and improve barriers.

Safety and core business compete for resources. Learning only uses a fraction of the data available

10^{-4} := 1 failure in 10.000 events



1 - 10^{-4} := 9.999 non-failures in 10.000 events

Safety-II = Ability to succeed under varying conditions.

Focus is on what goes right. Use that to understand everyday performance, to do better and to be safer.

Safety and core business help each other. Learning uses most of the data available

Failures or successes?

FRAM

When something goes wrong, e.g., 1 event out of 10.000 (10E-4), humans are assumed to be responsible in 80-90% of the cases.

Who or what are responsible for the remaining 10-20%?

Investigation of failures is accepted as important.

When something goes right, e.g., 9.999 events out of 10.000, are humans also responsible in 80-90% of the cases?

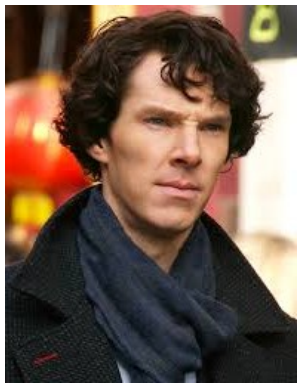
Who or what are responsible for the remaining 10-20%?

Investigation of successes is rarely undertaken.

Noticing the unnoticeable

"Is there any point to which you would wish to draw my attention"?

"To the curious incident of the dog in the night-time."



"The dog did nothing in the night-time."

"That was the curious incident," remarked Sherlock Holmes.

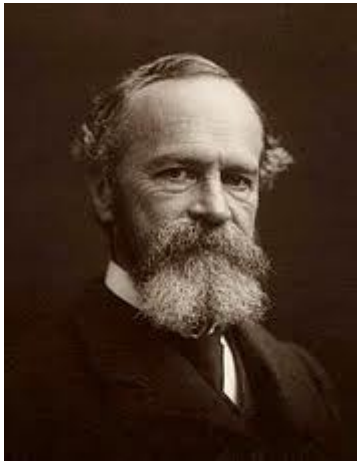
It is necessary to know what is 'normal' – what usually happens or should happen – in order to notice and/or understand what is unusual.

Notice the unnoticeable

FRAM

Learn from situations where nothing out of the ordinary seemed to happen
Try to understand what actually takes place.
Recognise the adjustments that people make and try to learn from them.

"Habit diminishes the conscious attention with which our acts are performed"



William James
(1842-1910)

"Habitual actions are certain, and being in no danger of going astray from their end, need no extraneous help"

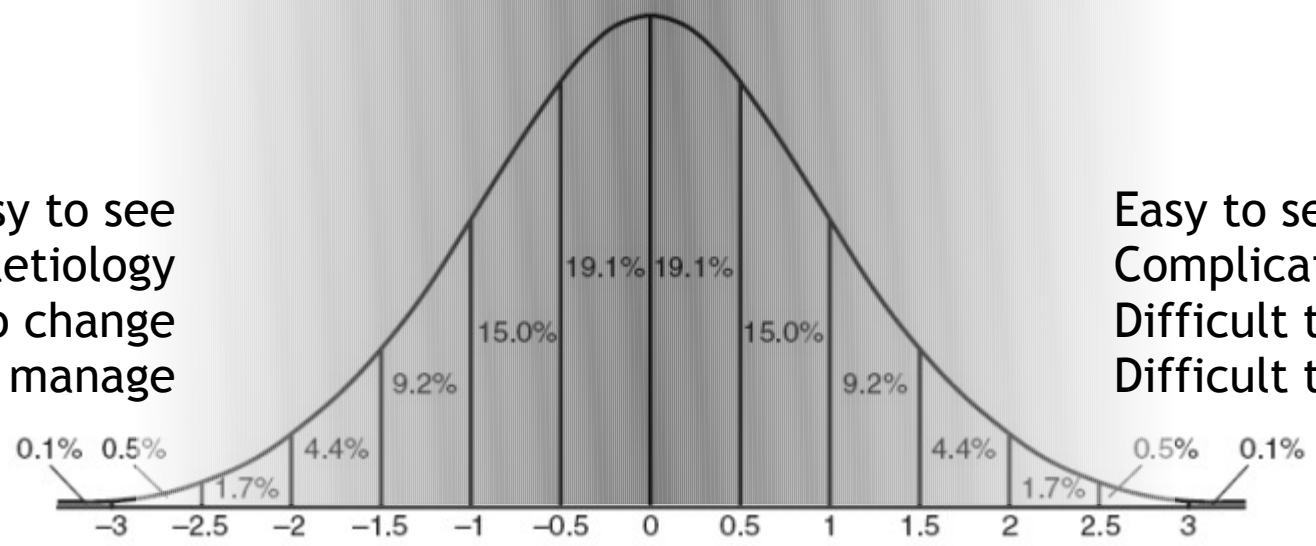
We stop paying attention to something as soon as when we get used to doing it.
After some time we neither notice it, nor do we think it is necessary to do so.
This applies both to what we do ourselves and what others do.

What should we be looking for?

FRAM

‘Difficult’ to see
Uncomplicated aetiology
Easy to change
Easy to manage

Easy to see
Complicated aetiology
Difficult to change
Difficult to manage



Easy to see
Complicated aetiology
Difficult to change
Difficult to manage

Safety focus:
Accidents & Disasters

Generally ignored or unknown

Gratefully accepted

Counting and understanding

The numerator is how many there are of a type of event (accidents, incidents, etc.) This number is known (with some uncertainty)

$$\frac{\text{Numerator}}{\text{Denominator}}$$

The denominator is how many cases something could have happened but did not. This number is usually disregarded and is mostly unknown.

In 2011 there were a total of 490,007 movements in Frankfurt Airport, but only 10 infringements of separation and 11 runway incursions. The ratio was 2.04 10-5 and 2.25 10-5, respectively.

In 2012 trains stopped at a red signal ca. 13.000.000 times in Belgium. In 130 cases a train passed through a signal (SPAD), a third of these were serious, but only one accident. The probability of a SPAD is 10-5, and of an accident 7.7 10-8.

We count things that go wrong and try to understand them.
But we do not count things that succeed, nor do we try to understand them

Stopping at a red light

People drive in different ways, depending on multiple factors (age, gender, nationality, weather, vehicle, traffic environment, etc.)

Most drivers stop at a red traffic light, but very few do it in the same way.

The ontology of Safety-II (variability)

FRAM

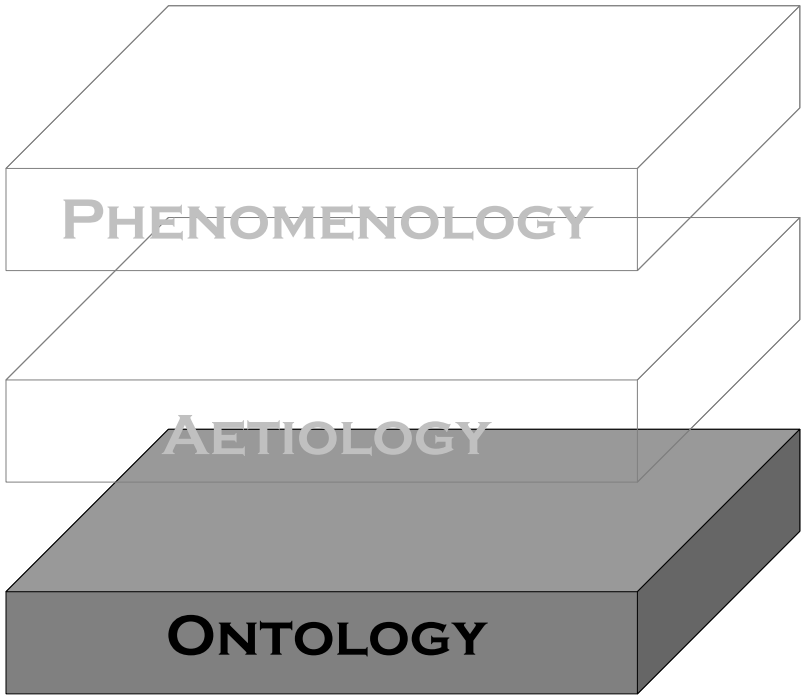
Systems are so complicated that work situations always are underspecified – hence partly unpredictable

Few things can be done unless procedures and tools are adapted to the situation. Performance variability is both normal and necessary.

Because resources (time, information, etc.) are finite, such adjustments will always be approximate rather than exact.

Individuals, groups, and organisations must adjust their performance to meet existing conditions (resources and requirements).

Most socio-technical systems are intractable; work conditions therefore differ from what has been specified or prescribed.



Performance adjustments are necessary

FRAM

Availability of resources (time, manpower, materials, information, etc.) may be limited and uncertain.

People adjust what they do to match the situation.

Performance variability is inevitable, ubiquitous, and necessary.

Because of resource limitations, performance adjustments will always be approximate.

Performance variability is the reason why everyday work is safe and effective.

Performance variability is the reason why things sometimes go wrong.

Why don't people bump into each other?

FRAM

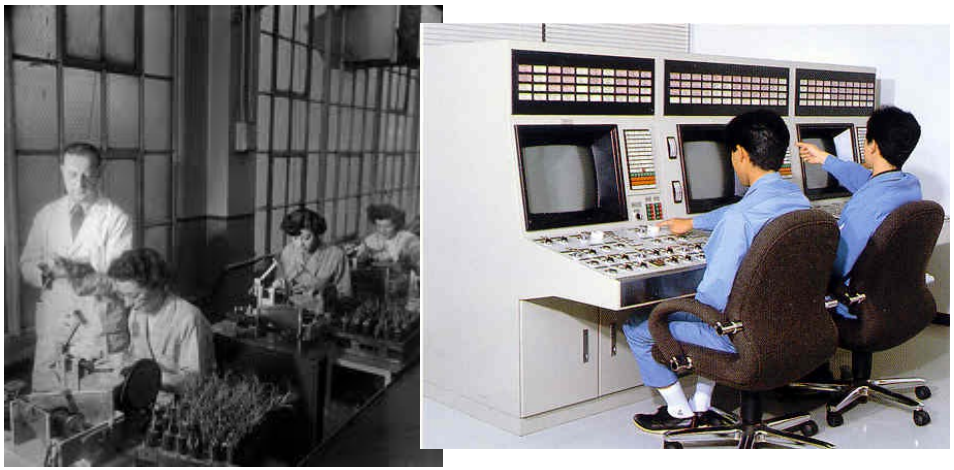
When we move in a crowd, we continuously adjust to what other people do.

Just as others continuously adjust to what we do – or may be doing.

Work as imagined – work as done

FRAM

Work-as-imagined is what designers, managers, regulators, and authorities believe happens or should happen.



Safety I: Failure is explained as a breakdown or malfunctioning of a system and/or its components (non-compliance, violations).

Work-as-done is what actually happens.

Safety II: Individuals and organisations must adjust to the current conditions in everything they do. Performance must be variable in order for things o work.

Efficiency-Thoroughness Trade-Off

FRAM

Thoroughness: Time to think

Recognising situation.

Choosing and planning.

If thoroughness dominates, there may be no time to do things.

Efficiency: Time to do

Implementing plans.

Executing actions.

If efficiency dominates, actions may be badly prepared or wrong

Time & resources needed

Time & resources available

Efficiency-thoroughness in practice

FRAM

In practice, people take the shortcuts they think are necessary to get the job done, to save time, to avoid unnecessary use of resources, etc.

When it goes well, no one takes any notice – and the shortcuts may even tacitly be encouraged.

When it goes wrong, people are blamed for 'violating' procedures and for being unsafe.

FRA Approach Phraseology

Standard phraseology

"DLH123, Langen Radar identified, cleared OSMAX 25 Transition, high speed approved"

Duration:
About 4.7 seconds

Non-standard phraseology

"Gude, DLH123, OSMAX 25 Transition, high speed"

Duration:
about 3.0 seconds

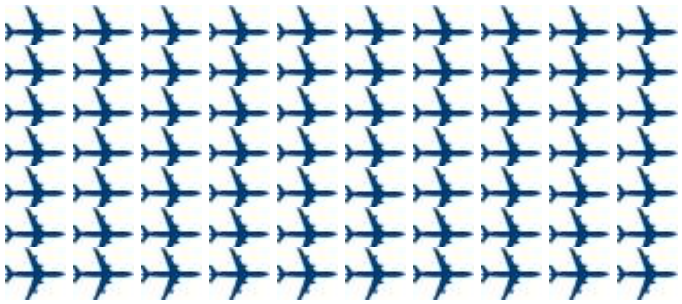
Time saved: about 1.7 seconds

Example courtesy of Jörg Leonhardt, DFS.

How much is 1.7 seconds worth?

Number of movements during peak days	> 1.500 movements/day
Number of arrivals during peak times	> 50 arrivals/hour

There are about 14 transmissions per arrival – not including the time for readbacks.



With 50 arrivals/hour this means more than 700 transmissions/hour on frequency.

Saving just 1 second per transmission corresponds to 11 minutes saved per hour.

Where can we find ETTOing?

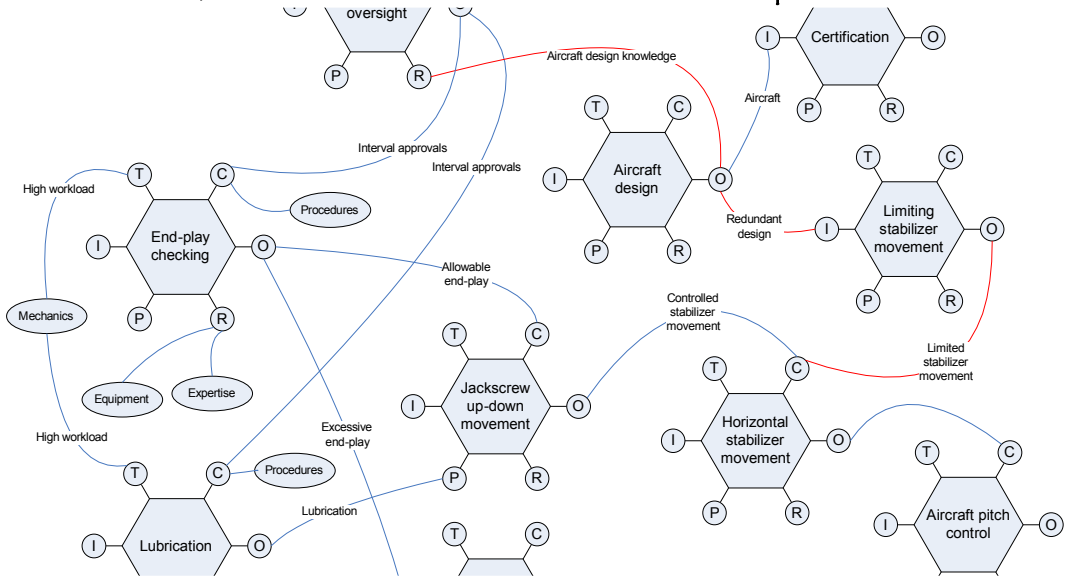
FRAM

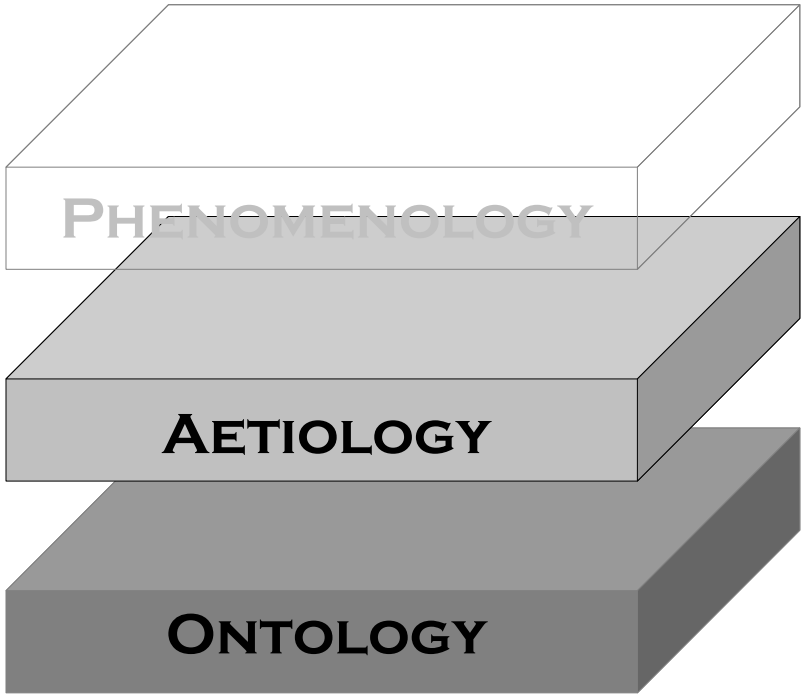
Efficiency-Thoroughness Trade-Offs are made by all professions and can be found on all levels of an organisation – from top management to daily operations.

The aetiology of Safety-II

While some adverse events can be attributed to a breakdown or malfunctioning of components and normal system functions, many cannot. These events are better understood as the result of unexpected combinations of performance variability.

Accidents result from unexpected combinations (resonance) of variability of normal performance.



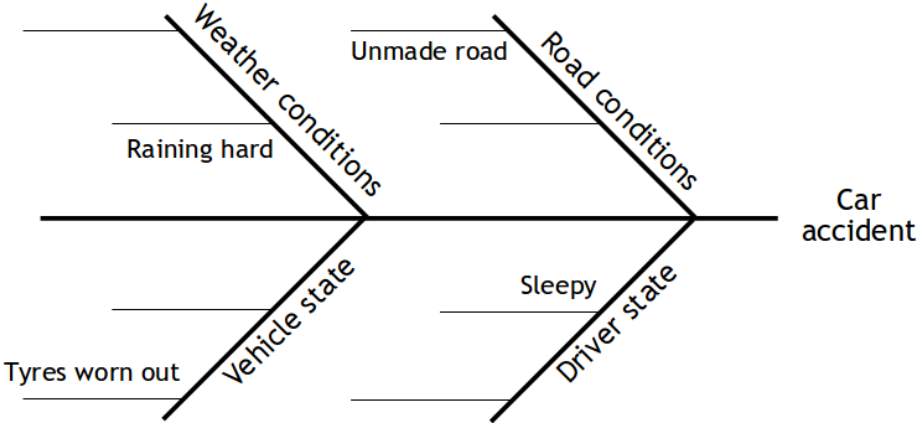


Safety is achieved by controlling variability (monitoring and damping).

Stable vs. transient causes

FRAM

Causes are assumed to be stable. Causes can be 'found' by backwards tracing from the effect. Causes are 'real.'



Final effects are (relatively) stable changes to some part of the system. Effects are 'real.'

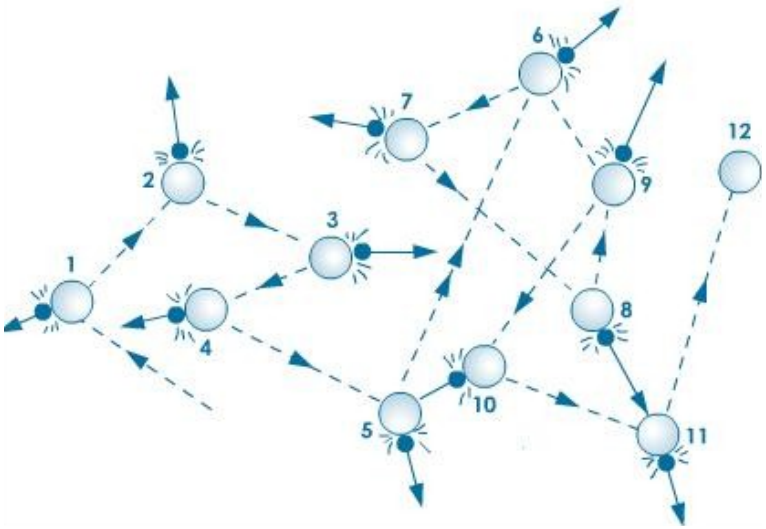
Causes can be associated with components or functions that in some way have 'failed.' The 'failure' is either visible after the fact, or can be deduced from the facts.

Stable vs. transient causes

FRAM

Outcomes 'emerge' from transient (short-lived) intersections of conditions and events.

Causes represent a pattern that existed at one point in time. But they are inferred, hence 'made' rather than 'found.'



Final outcomes are (relatively) stable changes to some part of the system. Effects are 'real.'

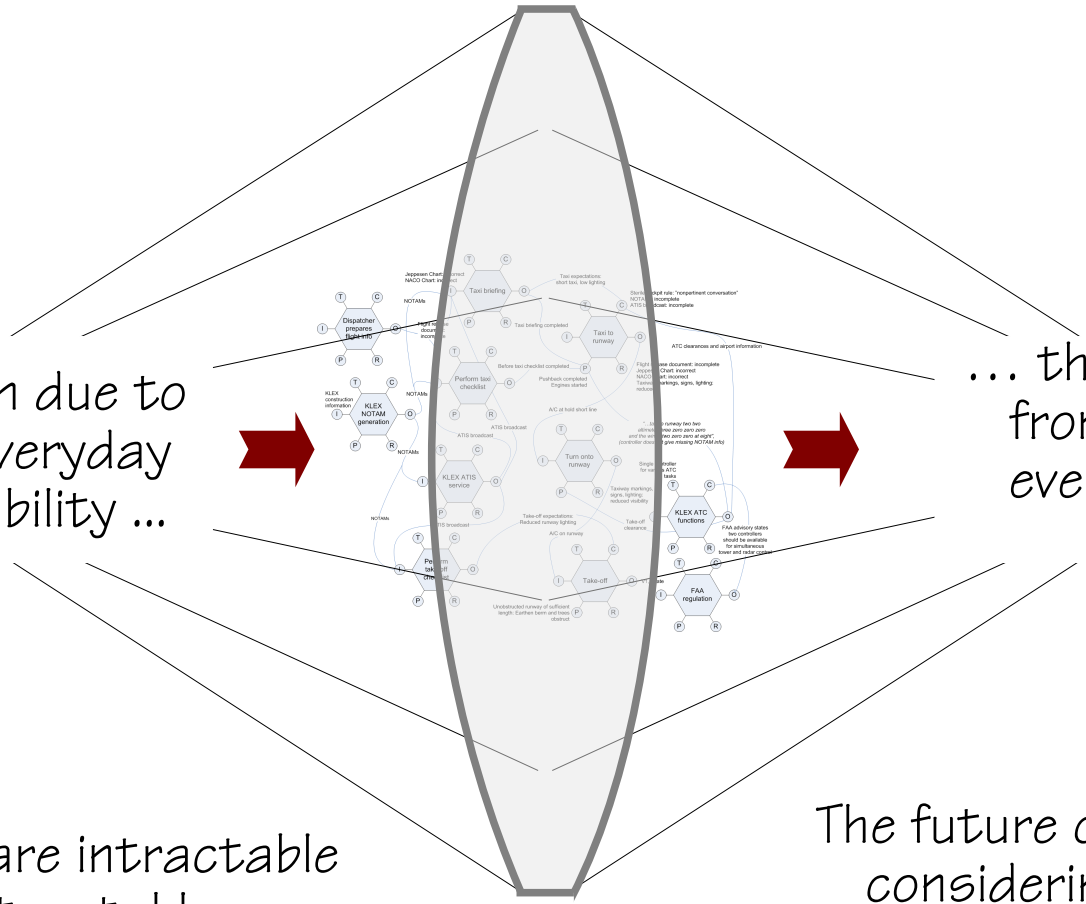
Outcomes cannot be traced back to specific components or functions, hence are not the 'effects' of known 'causes'.

Non-linear (systemic) model

Region Syddanmark

Non-decomposable, non-linear models

If accidents happen due to combinations of everyday performance variability ...



... then risks also emerge from combinations of everyday performance variability.

Systems at risk are intractable rather than tractable.

The future can be understood by considering the characteristic variability of the present.

The phenomenology of Safety-II

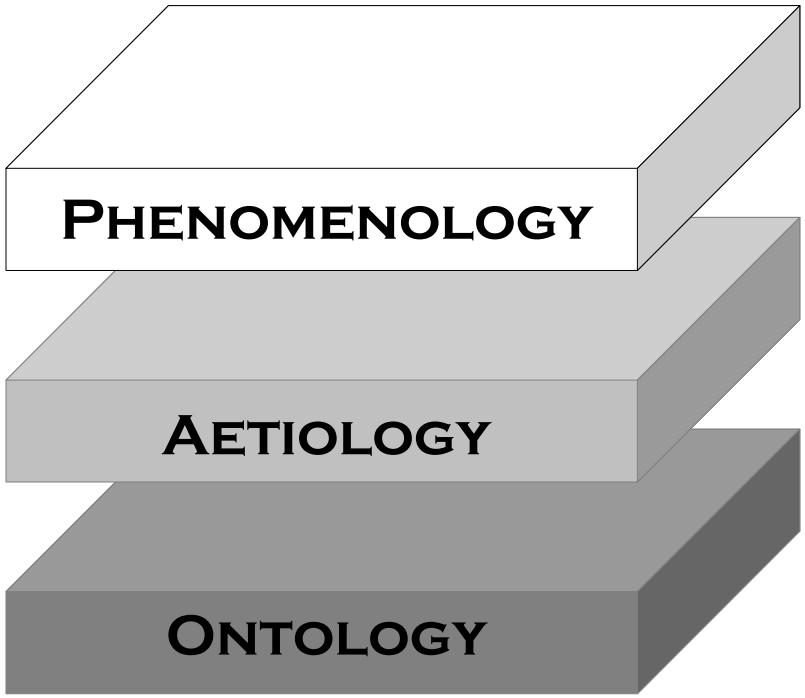
FRAM

"Safety is a dynamic non-event" (Karl Weick)

There is an absence of failures (things that go wrong), but as a result of active engagement. But if safety is a non-event, it can neither be observed, no measured

Safety is a dynamic event

There is a presence of successes (things that go right), and the more there are, the safer the system is.
If safety is something that happens, rather than something that does not happen, then it can be observed – and measured.

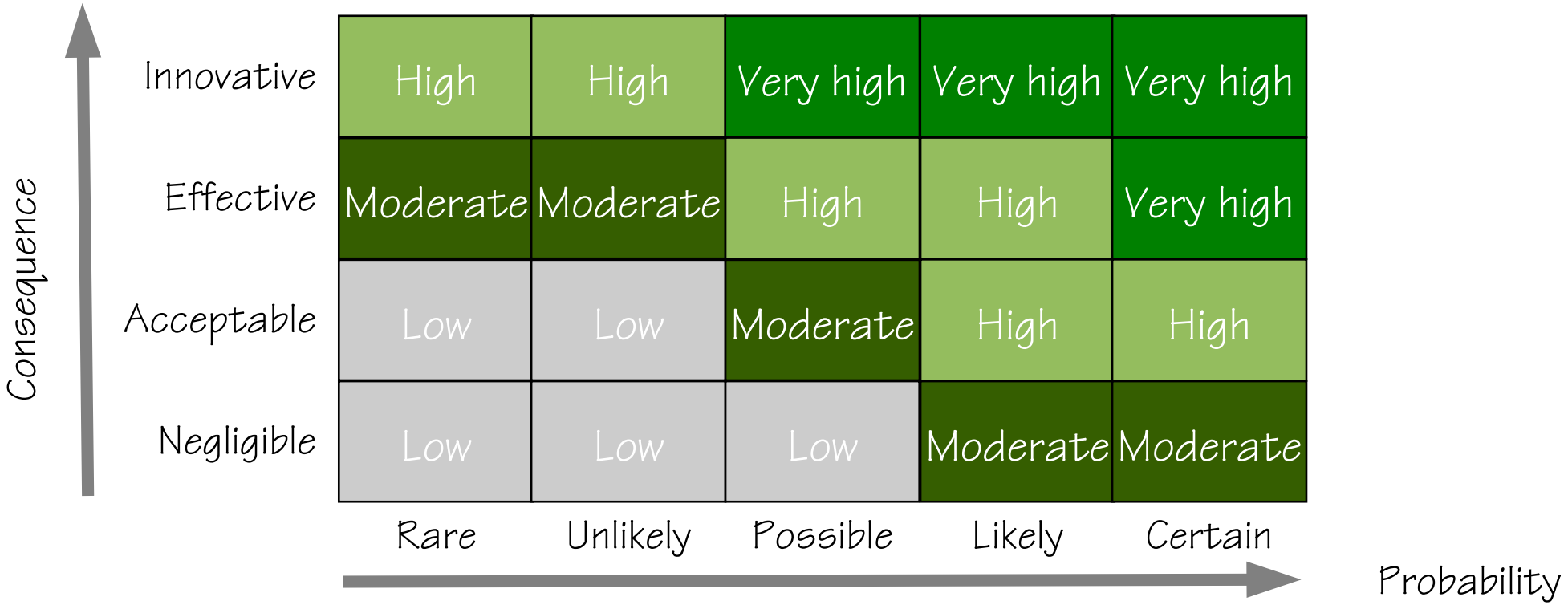


In order to ensure that a system is safe, we need to understand how it succeeds rather than how it fails.

Benefit profile (= safety)

FRAM

When we look for the things that go right, we are hampered by a lack of terminology (taxonomies, models



WYLFIWYF: But what should we look for?

Why do people vary in their work?

FRAM

AVOID

anything that may have negative consequences for yourself, your group, or organisation

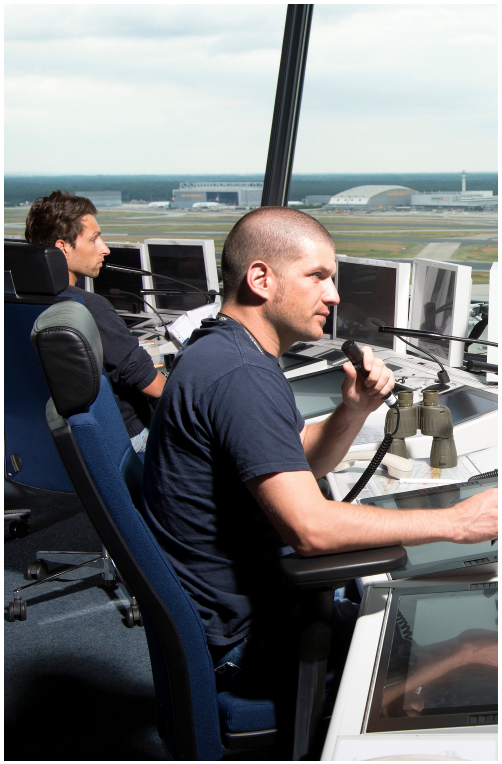
COMPENSATE FOR

unacceptable conditions so that it becomes possible to do your work.

MAINTAIN/CREATE

conditions that may be of use in case of future problems.

Same process => different outcomes



Function (work as imagined)

Everyday work (performance variability)

Malfunction, non-compliance, error

Success (no adverse events)

Failure (accidents, incidents)

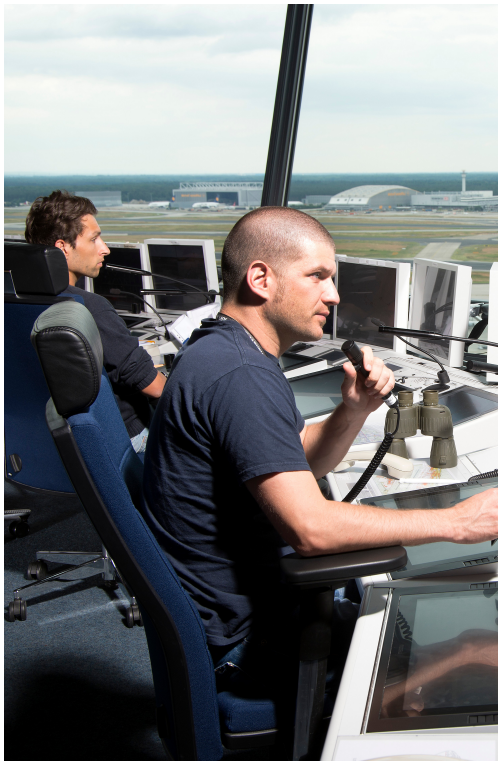
Acceptable outcomes

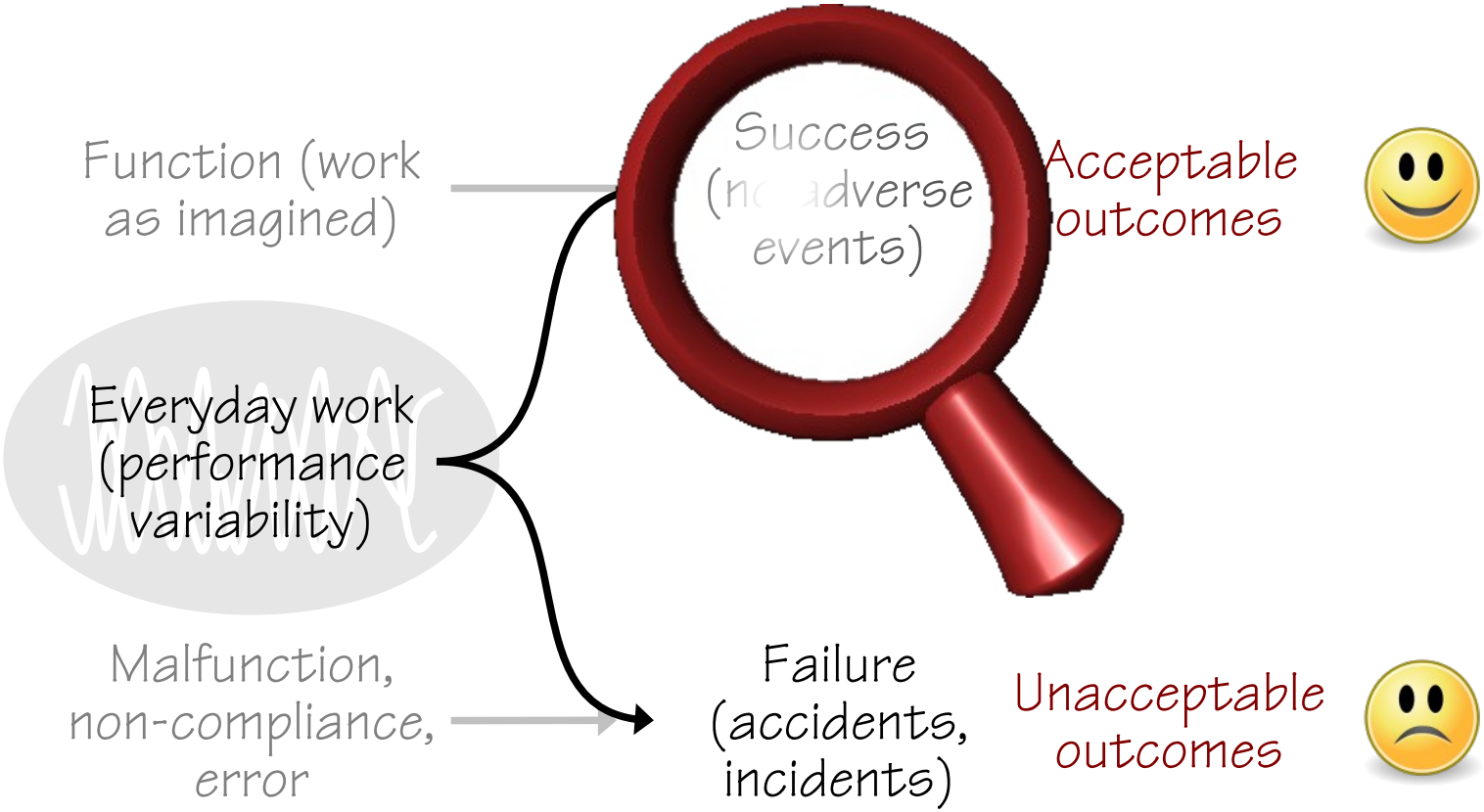
Unacceptable outcomes

Increase safety by facilitating work

FRAM

Understanding the variability of everyday performance is the basis for safety.





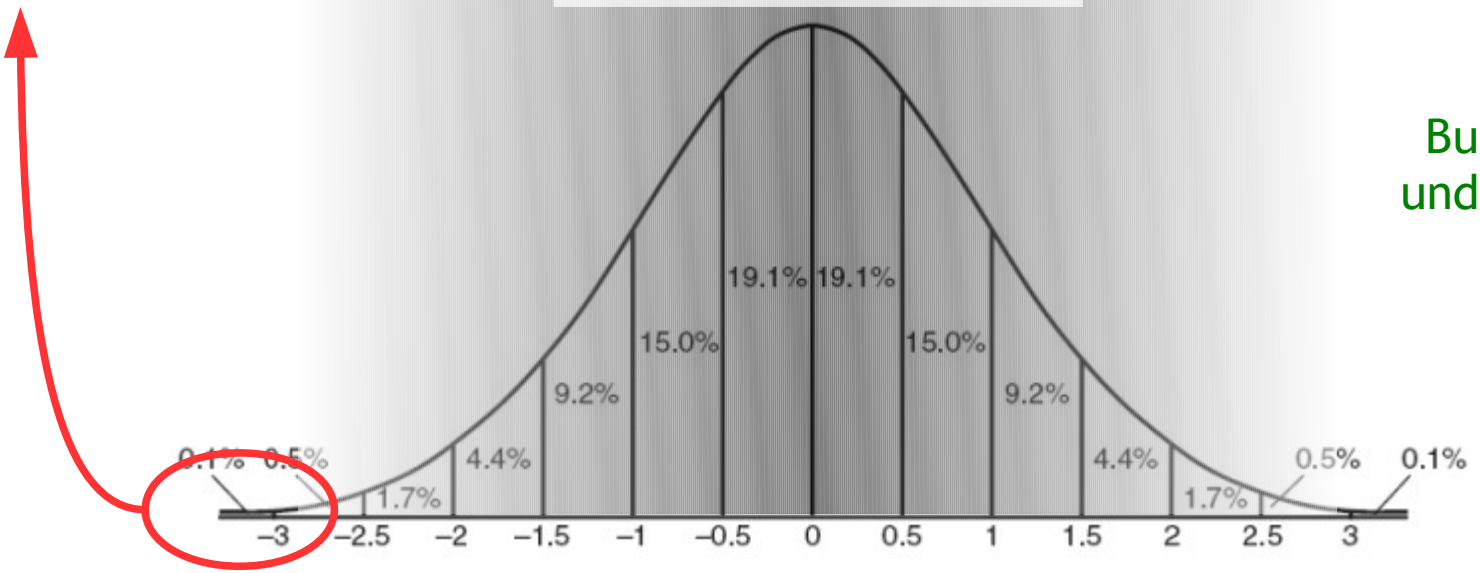
Constraining performance variability to remove failures will also remove successful everyday work.

What should we be looking for?

When we notice something that has gone wrong …

… it is a safe bet that it has gone right many times before …

… and that it will go right many times in the future.



But we need to understand HOW!

Everyday performance

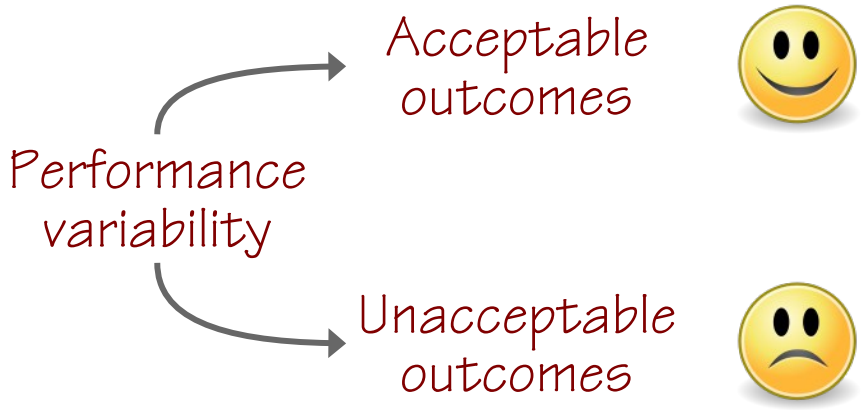
Safety II – when everything goes right

FRAM

Safety-II: Safety is a condition where the number of successful outcomes (meaning everyday work) is as high as possible. It is the ability to succeed under varying conditions.

Safety-II is achieved by trying to make sure that things go right, rather than by preventing them from going wrong.

Individuals and organisations must adjust everything they do to match the current conditions. Everyday performance must be variable in order for things to work.



WYLFIWYF

FRAM

Accident investigation follow a What-You-Look-For-Is-What-You-Find (WYLFIWYF) principle.

This means that accident investigations usually find what they look for.
The assumptions about the nature of accidents (**causality credo**) constrain the analysis.

Human error
Technical malfunction
Organisational failure
Incorrect design

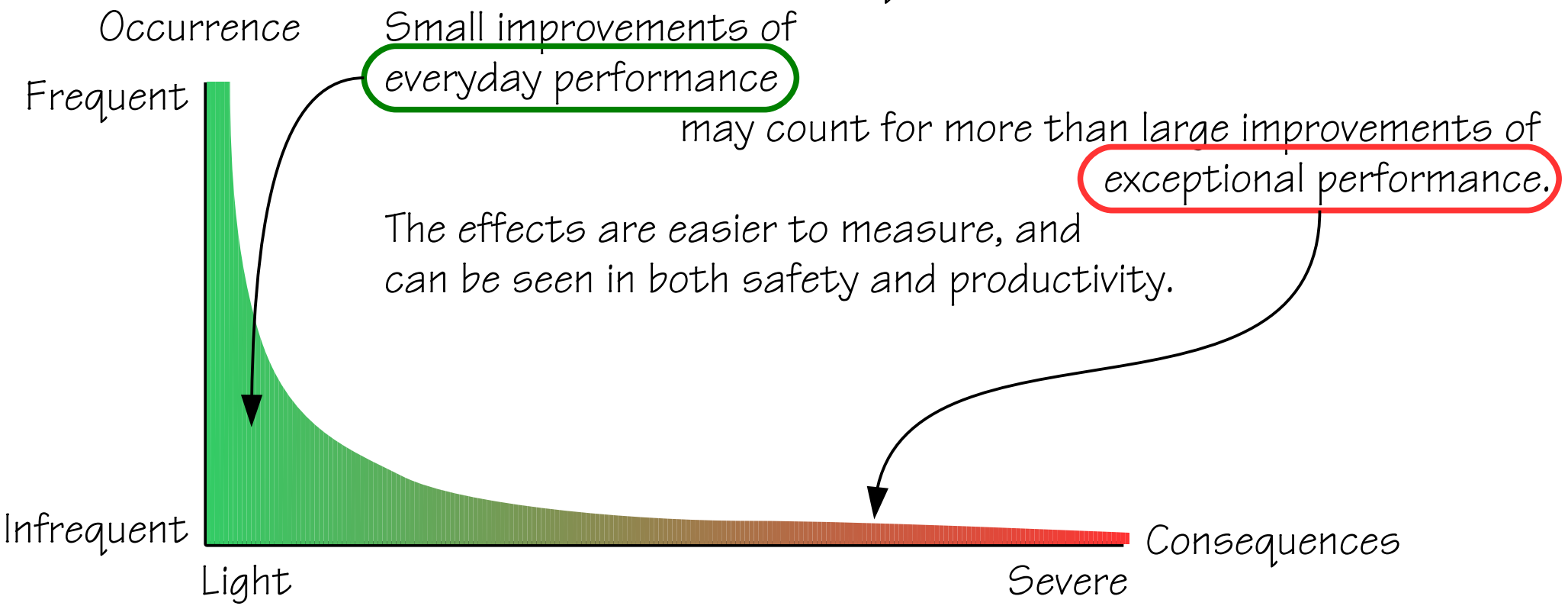
Bad maintenance
Safety culture
Latent conditions
Violation, non-compliance

To this can be added the principle of WYFIWYF:
What You Find Is What You Fix

Frequency rather than severity

FRAM

It is easier to learn from that which is frequent (and regular) than for that which is infrequent and irregular.



Adverse outcomes are more likely to be the result of usual actions under unusual conditions, than unusual actions under usual conditions.

What should we learn from?

FRAM

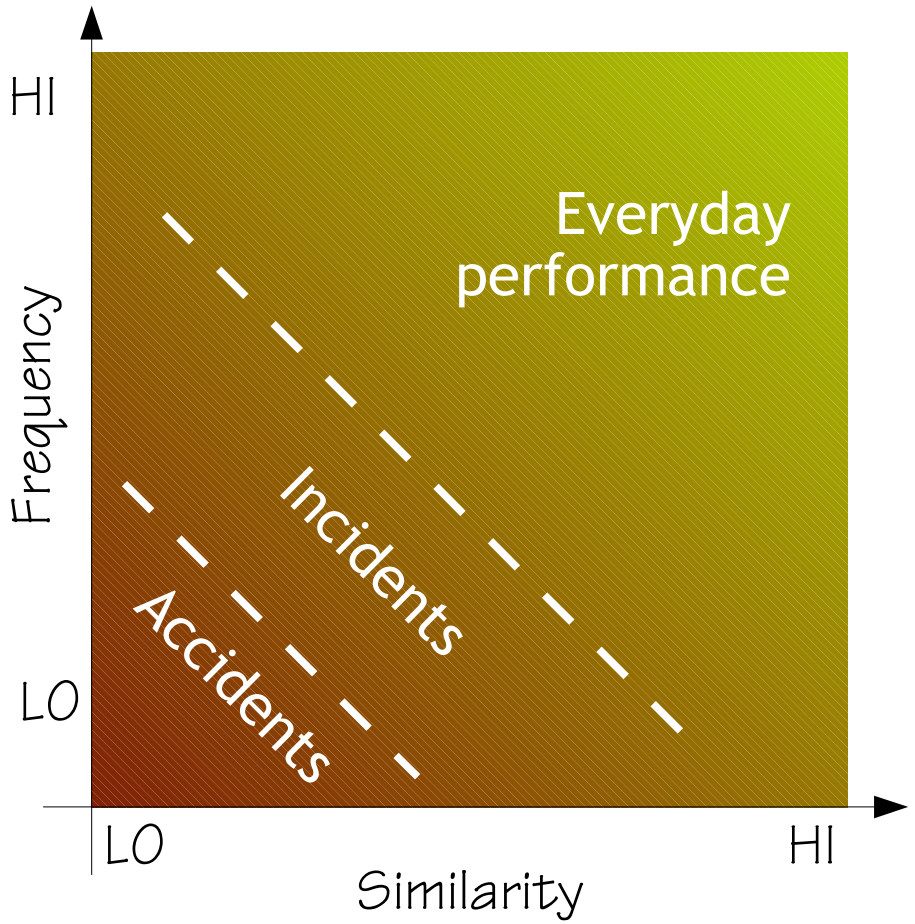
Safety-I approach

Look for what went wrong.
Reconstruct failure sequence (time-line)
Find the component or subsystem that failed.
Select events based on their severity.
Eliminate causes of failures

Safety-II approach (resilience engineering)

Look for what went right (when it could have gone wrong)
Describe the typical adjustments to performance.
Understand the role of these performance adjustments.
Select events based on their frequency
Facilitate ability to work in all situations alike

What do we need to learn?



Opportunity (to learn): Learning situations (cases) must be frequent enough for a learning practice to develop

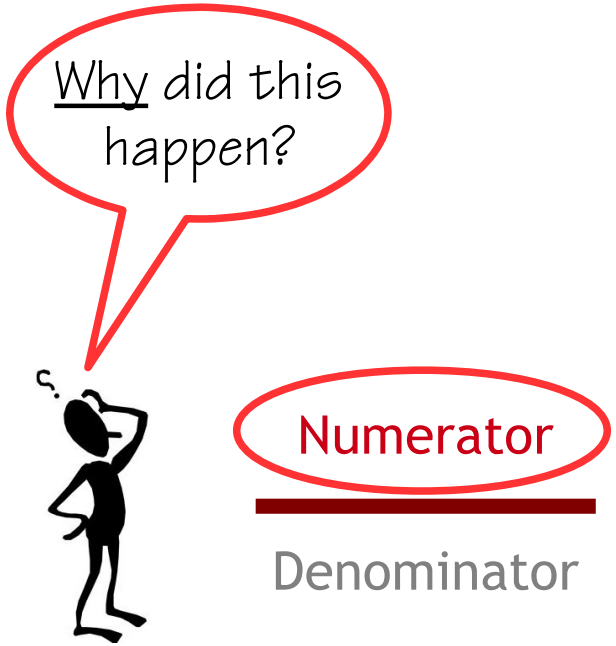
Comparable /similar: Learning situations must have enough in common to allow for generalisation.

Opportunity (to verify): It must be possible to verify that the learning was 'correct' (feedback)

The purpose of learning (from accidents, etc.) is to change behaviour so that certain outcomes become more likely and other outcomes less likely.

Absolute safety

From a Safety-I perspective, the purpose of investigations is to find what has failed or gone wrong.



All accidents are preventable (safety myth)

The numerator (number of adverse outcomes) should therefore be as small as possible.

Ideally it should be zero (zero accident ideology)

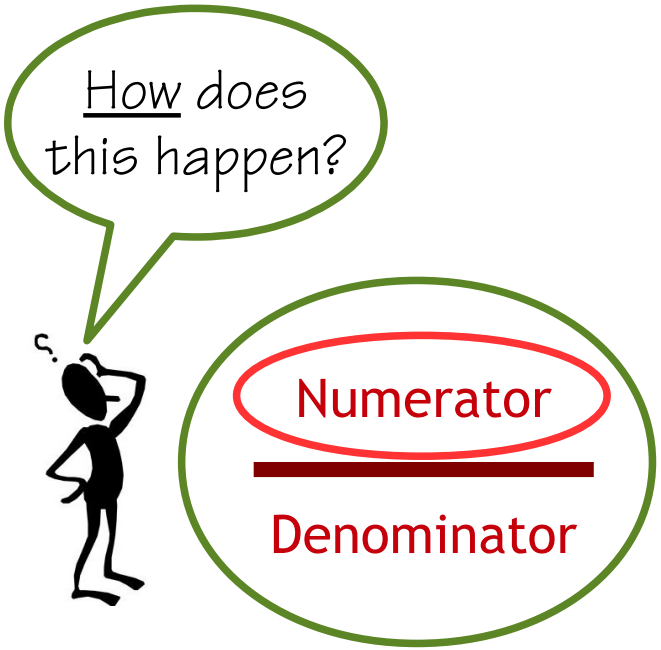
The numerator can be reduced by understanding failures and eliminating the causes ('fix and find').

Little or no concern is shown for the denominator – the complement of the adverse outcomes.

Relative safety

FRAM

From a Safety-II perspective, the purpose of investigations is to understand how things usually are done, as a basis for explaining the specific instance.



The ratio should be as small as possible

This happens if the numerator becomes smaller.

But it also happens if the denominator becomes larger.

The denominator can be increased by understanding how things do not fail, and by making sure that they happen.

Increasing the denominator will also increase productivity.

Relation between Safety-I and Safety-II

FRAM

Safety-II (resilience engineering) can reuse many methods & techniques from Safety-I, but with a different purpose and from a different perspective

Resilience

Focus of Safety-II:
Everyday actions and outcomes – risks as well as opportunities

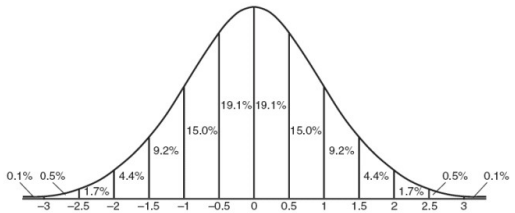
Safety

Focus of Safety-I:
Accidents & incidents

The number of adverse outcomes

$$\frac{n}{M}$$

The number of events



Thank you for your attention

Any questions?