Introduction to FRAM: The Functional Resonance Analysis Method

Professor Erik Hollnagel
University of Southern Denmark
Odense, Denmark
E-mail: hollnagel.erik@gmail.com
Three types of models

Principle of causation

Emergent outcomes

Single causes ("Root")

Multiple causes (Latent)

Systemic model (non-linear)

Epidemiological model (complex linear)

Sequential model (simple linear)

1920 1940 1960 1980 2000
Understanding simple systems

We can explain how things work in terms of cause-effect relations.

Antikythera mechanism, (150-100 BC)

We can therefore understand risks in the same way: as cause-effect chains starting from a component failure.
Reasoning in cause-effect relations is no longer adequate.

Difficult to imagine how events and conditions may combined.

A growing number of risks therefore remain unknown.
Working / operating environments are unstable and unpredictable. Actions / changes therefore often have unanticipated consequences.
Explaining machines / technology

When looking at a piece of equipment, such as a machine, it is natural to explain how it works by decomposing it into smaller parts.

We know how components work and how they are put together, because we have constructed the devices.
Principle of bimodal functioning

In the technological 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.

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

More intricate systems, such as engines, can be maintained and repaired, as long as it is considered worthwhile.

Complex, technological systems work according to the same principle. Failures may, however, be intermittent – especially if complex logic (software) plays a part. Performance is basically **bimodal**; either the system works correctly (as designed) or it has failed.
Common assumptions (~1970)

System can be decomposed into meaningful elements (components, events)
The function of each element is bimodal (true/false, work/fail)
The failure probability of elements can be analysed/described individually

The order or sequence of events is predetermined and fixed
When combinations occur they can be described as linear (tractable, non-interacting)
The influence from context/conditions is limited and quantifiable
Nature of socio-technical systems

All systems unique

- Must be described top-down in terms of functions and objectives.

- Decomposition does not work for socio-technical systems, because they are emergent.

- Risks and failures must therefore be described relative to functional wholes.

Complex relations between input (causes) and output (effects) give rise to unexpected and disproportionate consequences. Socio-technical systems are non-linear and event outcomes are intractable.
Humans and social systems are not bimodal. Everyday performance is variable and this – rather than failures and ‘errors’ – is why accidents happen. Since performance shortfalls are not a simple (additive or proportional) result of the variability, more powerful, non-linear models are needed.

Performance variations can be have positive as well as negative outcomes!

But human factors has tended to look for negative aspects of performance - deviations or “errors”
While some adverse events can be attributed to failures and malfunctions of everyday functions, others are best understood as the result of combinations of variability of everyday performance.

Risk and safety analyses should try to understand the nature of variability of everyday performance and use that to identify conditions that may lead to both positive and adverse outcomes.
Principles for FRAM

I The principle of equivalence of successes and failures.

II The principle of approximate adjustments.

III The principle of emergence.

IV The principle of functional resonance.
Equivalence of success and failures

Failure is normally explained as a breakdown or malfunctioning of a system and/or its components.

This view assumes that success and failure are of a fundamentally different nature.

Resilience Engineering recognises that individuals and organisations must adjust to the current conditions in everything they do. Because information, resources and time always are finite, the adjustments will always be approximate.

Success is due to the ability of organisations, groups and individuals correctly to make these adjustments, in particular correctly to anticipate risks before failures and harm occur.

Failures can be explained as the absence of that ability – either temporarily or permanently.

The aim of Resilience Engineering is to strengthen that ability, rather than just to avoid or eliminate failures.
Actions and Errors

"Knowledge and error flow from the same mental sources, only success can tell one from the other."
(Ernst Mach, 1838-1916)

Work systems and work conditions are always underspecified. Humans adjust performance to compensate for that. Human performance is inherently variable.

HUMAN ERROR! Actions with a negative outcome.
Actions with a beneficial outcome. CREATIVITY, LEARNING

Both are due to performance variability, and may therefore have the same “causes”.
Principle of approximate adjustments

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

Few – if any – tasks can successfully be carried out unless procedures and tools are adapted to the situation. Performance variability is both normal and necessary.

Many socio-technical systems are intractable. The conditions of work therefore never completely match what has been specified or prescribed.

Individuals, groups, and organisations normally adjust their performance to meet existing conditions (resources, demands, conflicts, interruptions).

Because resources (time, manpower, information, etc.) always are finite, such adjustments will invariably be approximate rather than exact.
Efficiency-Thoroughness Trade-Off

Thoroughness: Time to think
Recognising situation.
Choosing and planning.

If thoroughness dominates, there may be too little time to carry out the actions.
Neglect pending actions
Miss new events

Efficiency: Time to do
Implementing plans.
Executing actions.

If efficiency dominates, actions may be badly prepared or wrong
Miss pre-conditions
Look for expected results

Time & resources needed
Time & resources available

© Erik Hollnagel, 2011
Things go right and things go wrong for the same reasons. Efficiency-thoroughness trade-offs are both normal and necessary.

It is taken for granted that things usually go right. This is therefore rarely analysed or investigated.

When things go wrong, and particularly if the consequences are serious, the event is investigated to find the cause.

An ETTO is always approximate, because of the very reasons that make it necessary (insufficient time and information!). But making an efficiency-thoroughness trade-off is never wrong in itself!

People are expected to be both efficient and thorough at the same time – or rather to be thorough, when with hindsight it was wrong to be efficient.
Some ETTO heuristics

**Cognitive (individual)**
- Judgement under uncertainty
- Cognitive primitives (SM – FG)
- Reactions to information input overload and underload
- Cognitive style
- Confirmation bias

**Idiosyncratic (work related)**
- Looks fine
- Not really important
- Normally OK, no need to check
- I’ve done it millions of time before
- Will be checked by someone else
- Has been checked by someone else
- This way is much quicker
- No time (or resources) to do it now
- Can’t remember how to do it
- We always do it this way
- It looks like X (so it probably is X)
- We must get this done
- Must be ready in time
- Must not use too much of X

**Collective (organisation)**
- Negative reporting
- Reduce redundancy
- Meet “production” targets
- Reduce unnecessary cost
- Double-bind
- Reject conflicting information

© Erik Hollnagel, 2011
“In splitting a board, a circular-saw operator suffered the loss of his thumb when, in violation of instructions, he pushed the board past the saw with his fingers, instead of using the push stick that had been provided for the purpose.”

“He stated that he had always done such work in this manner and had never before been hurt. He had performed similar operations on an average of twenty times a day for three months and had therefore exposed his hand in this way over one thousand five hundred times.” (Heinrich, 1931 “Industrial accident prevention”)
One way of managing time and resource limitations is to think only one step back and/or one step ahead.

- Confirm that input is correct
- Trust that input is correct
- Consider secondary outcomes and side-effects
- Assume someone else takes care of outcomes

For distributed work it is necessary to trust what others do; it is impossible to check everything.
M/S Finnbirch (1 November 2006)

En route from Helsinki to Århus, with a load of wood, paper rolls, plywood, and steel products.

The next day, the ship began to roll in a heavy storm and suddenly heeled heavily to the port with a total shifting of the cargo, until it finally capsized. 12 men were eventually rescued but 2 lost their lives.

Sailing Finnbirch sailed on an unfavourable course and speed, in conditions with high and long waves. This led to a loss of stability and severe heeling. IMO guidelines on how to manage a ship in adverse sea conditions were unknown.

Working as usual. Textile lashings were probably used instead of chains when securing cargo on two of the decks. Textile lashings are easier to use so lashing and unlashing is faster. Severe heeling that lead to shifting of the cargo.

Working as usual. Typical loading rate was 30 - 40 units/hour. One officer had to check all cargo units visually, manage the computerised stability program, ballast and deballast the ship, and keep the ship ramps in position.
Principle of emergence

The variability of normal performance is rarely large enough to be the cause of an accident in itself or even to constitute a malfunction. The variability from multiple functions may combine in unexpected ways, leading to consequences that are disproportionally large, hence produce non-linear effects. Both failures and normal performance are emergent rather than resultant phenomena, because neither can be attributed to or explained only by referring to the (mal)functions of specific components or parts.

The Small World Problem

Socio-technical systems are intractable because they change and develop in response to conditions and demands. It is therefore impossible to know all the couplings in the system, hence impossible to anticipate more than the regular events. The couplings are mostly useful, but can also constitute a risk.
The small world problem

What is the probability that any two persons, selected arbitrarily from a large population, will now each other, or be linked via common acquaintances?

A “target person” (Boston) and three groups of “starting persons” were selected (Nebraska: n=296, Boston: n=100). Target was identified by name, address, occupation, place of work, college & graduation year, military service, wife’s maiden name, hometown. Each starter was given a document and asked to move it by mail toward the target, via first-name acquaintances, who was asked to repeat the procedure.

Six degrees of separation

Technical systems consist of components organised to achieve a specific objective. The structure is fixed and does not normally change.

Network size and links are fixed (tractable).

Socio-technical systems include humans organised in social networks that change and develop according to needs and demands.

Network size and links may vary (intractable).

Six degrees of separation is the idea that, if a person is one “step” away from each person (s)he knows and two “steps” away from each person who is known by one of the people (s)he knows, then no one is more than six “steps” away from each person on Earth.

Facebook users average 3.74 degrees of separation (BBC 23 November 2011)
Six degrees of (Kevin) Bacon

The trivia game Six Degrees of Kevin Bacon is based on a variation of the concept of the small world phenomenon and states that any actor can be linked, through their film roles, to actor Kevin Bacon. The game requires a group of players to try to connect any film actor in history to the Kevin Bacon as quickly as possible and in as few links as possible. The game was played across various college campuses as early as the early 1990s.

Other examples:

www.findsatoshi.com

SixDegrees.org
Stable vs. transient causes

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

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 rather than ‘found.’ Causes are ‘elusive.’

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. Outcomes are emergent because the conditions that can explain them were transient.
The future as non-linear events

Non-linear events have been likened to Brownian movements or random walks. Risk assessment requires something that at the same time is non-linear (non-trivial) and systematic (predictable).
The variability of a number of functions may every now and then resonate, i.e., reinforce each other and thereby cause the variability of one function to exceed normal limits. The consequences may spread through tight couplings rather than via identifiable and enumerable cause-effect links, e.g., as described by the Small World Phenomenon. This can be described as a resonance of the normal variability of functions, hence as functional resonance. The resonance analogy emphasises that this is a dynamic phenomenon, hence not attributable to a simple combination of causal links.

Ways of looking at the future:

As a repetition or recurrence of the past (deterministic or probabilistic).
As a linear extrapolation of the past (combinatorial, probabilistic).
As randomly occurring events (defaitism).
As a non-linear but also non-random development (functional resonance),
Resonance

Natural oscillation + forcing function

Resonance, same frequency but increased amplitude

Time

Natural frequency, fixed amplitude

Forcing function with same frequency as natural oscillation
Stochastic resonance is the enhanced sensitivity of a device to a weak signal that occurs when random noise is added to the mix.
For each function, the others constitute the environment.

Every function has a normal weak, variability.

The pooled variability of the “environment” may lead to resonance, hence to a noticeable “signal.”

Functional resonance is the detectable signal that emerges from the unintended interaction of the normal variabilities of many signals.
Tacoma Narrows Bridge

July 1, 1940

November 7, 1940
London Millennium Bridge

Opened June 10, 2000


Reopened after reconstruction, January 2002
Analyses indicated that the movement was caused by the *sideways loads* we generate when walking. *Chance correlation* of our footsteps when we walk in a crowd generated slight sideways movements of the bridge. It then became more comfortable for people to *walk in synchronization* with the bridge movement.

This instinctive behaviour means that the sideways forces match the *resonant frequency* of the bridge, and are timed such as to increase the motion of the bridge. As the magnitude of the motion increases, the total sideways force increases and movements become even more correlated. The sway movement can occur on any bridge, future or existing, with a lateral frequency under ~1.3 Hz and with a sufficient number of pedestrians.

During the investigations it was discovered that other bridges with completely different structures to the Millennium Bridges have swayed sideways when crowded: The Auckland Harbour Road Bridge (1975), Toda park footbridge, Tokyo (1993) Pont Solferino, Paris (2000)
## Conclusions

<table>
<thead>
<tr>
<th>Safety models must go beyond simple cause-effect relations</th>
<th>Accidents result from alignment of conditions and occurrences</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Errors” – and their causes – are constructed rather than found</td>
<td>Human actions cannot be understood in isolation</td>
</tr>
<tr>
<td>All systems try to balance efficiency and thoroughness</td>
<td>More important to understand nature of system dynamics (variability) than to model individual technological or human failures.</td>
</tr>
<tr>
<td>Safety management must model how variability can create resonance</td>
<td>System as a whole adjusts to absorb everyday performance adjustments (dynamic accommodation) based on experience.</td>
</tr>
<tr>
<td></td>
<td>Accidents are consequences of everyday adjustments, rather than of failures. Without such adjustments, systems would not work</td>
</tr>
</tbody>
</table>