CITY UNIVERSITY LONDON

# Evaluating complex adaptive systems

What can we learn from swans, cats, dragons, toads, bugs and oxymorons?

DEOS, OSD Symposium March 2012, Tokyo

Robin E Bloomfield
Adelard LLP and CSR City University London

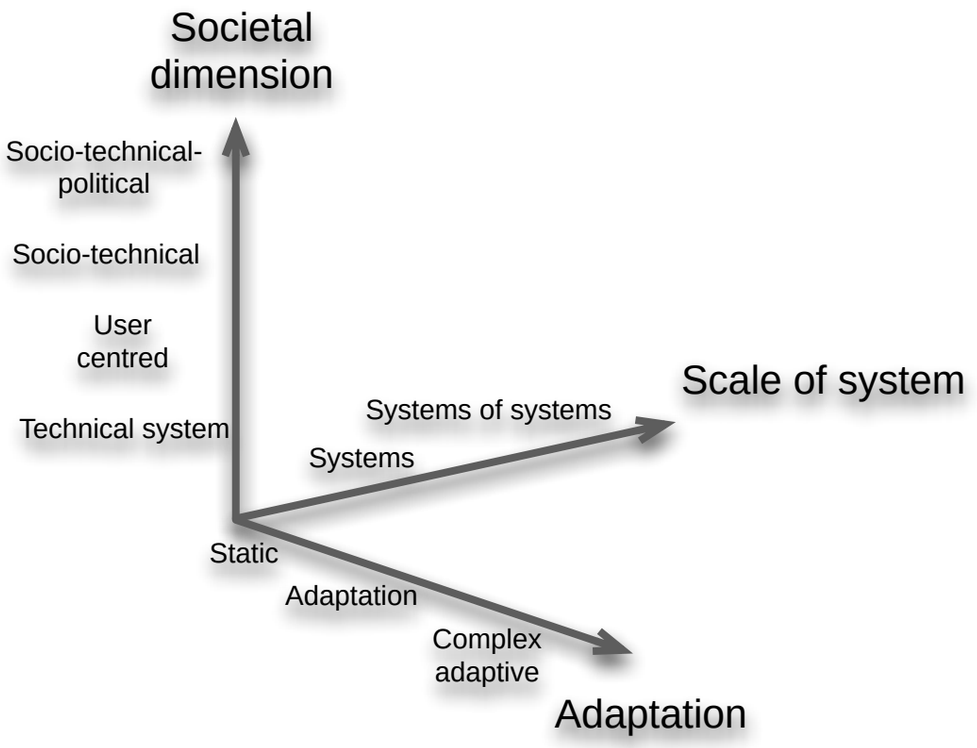
reb@adelard.com
reb@csr.city.sc.uk
*College Building, City University, London EC1V 0HB*
*Tel: +44 20 7832 5850 (sec Adelard)*
*Tel: +44 20 7040 8420 (sec CSR)*

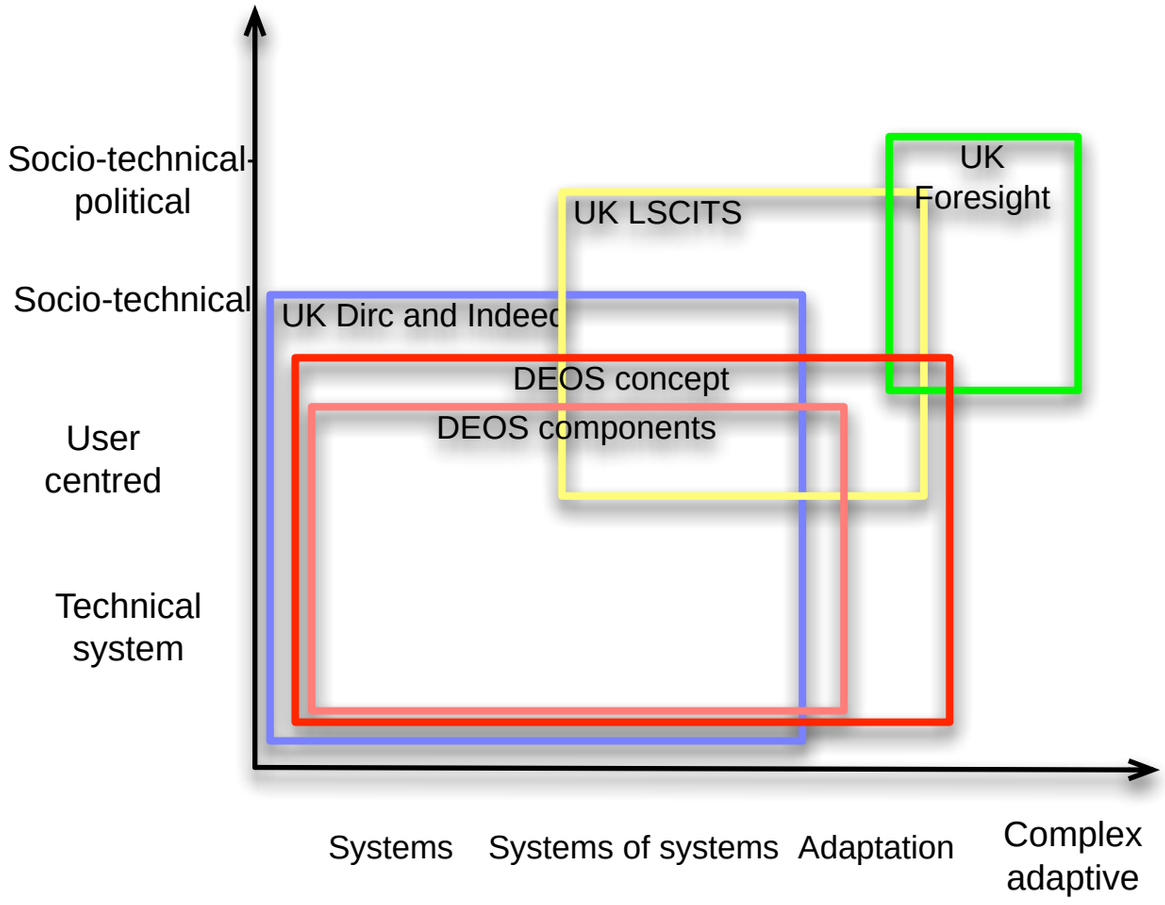
CSR Building confidence in a computerised world
www.csr.city.ac.uk

ADELARD

# Introduction

- Focus on evaluation
- Challenge of complex adaptive systems
  - What is current engineering state of practice
  - What are particularities of complex, adaptive systems
- Future directions
  - Engineering complex adaptive systems




CSR Building confidence in a computerised world
www.csr.city.ac.uk

ADELARD

# Flash crash - May 2010, ~$1tr


**Preliminary Findings Regarding**
**the Market Events of May 6, 2010**

**Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues**


Building confidence in a computerised world
www.csr.city.ac.uk

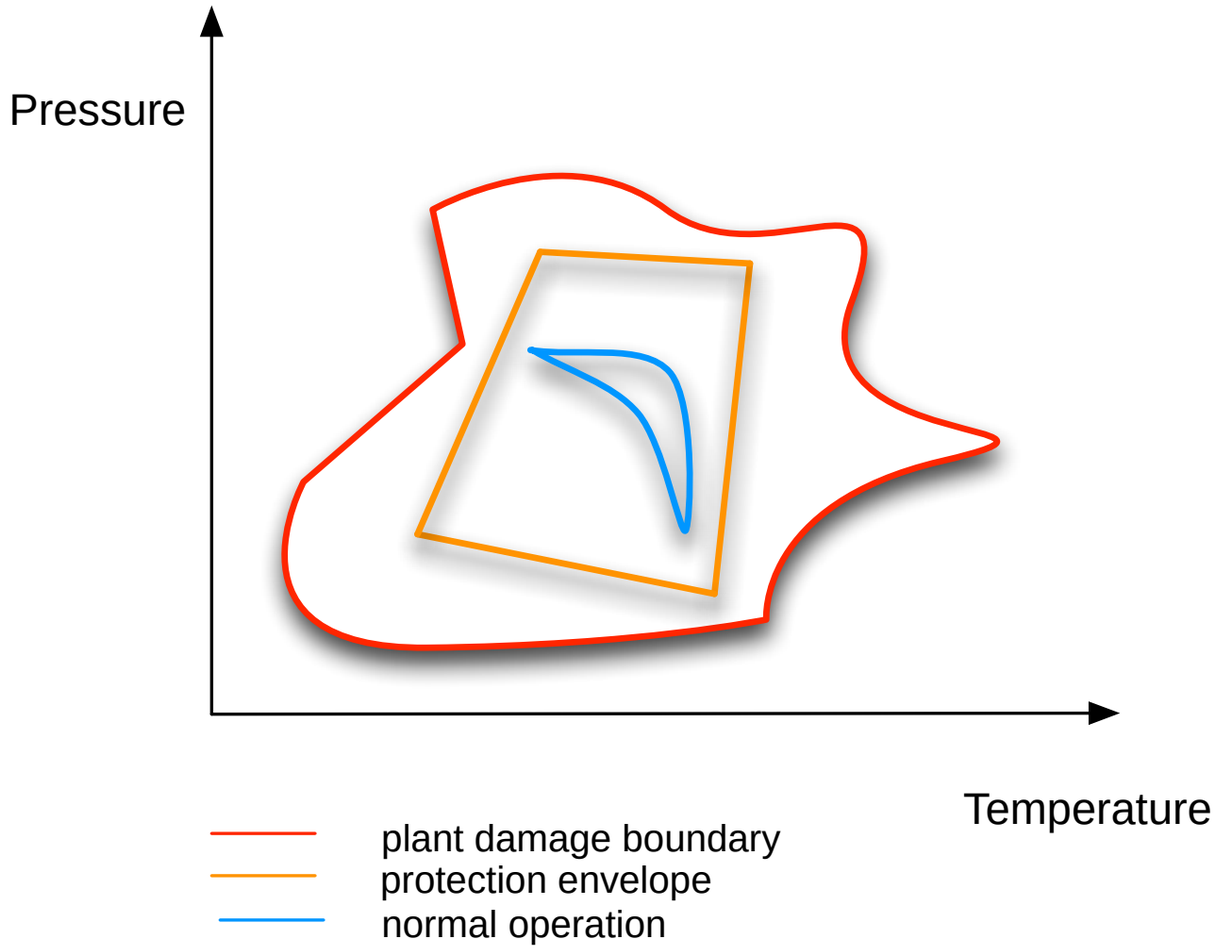
ADELARD

# Complex adaptive systems

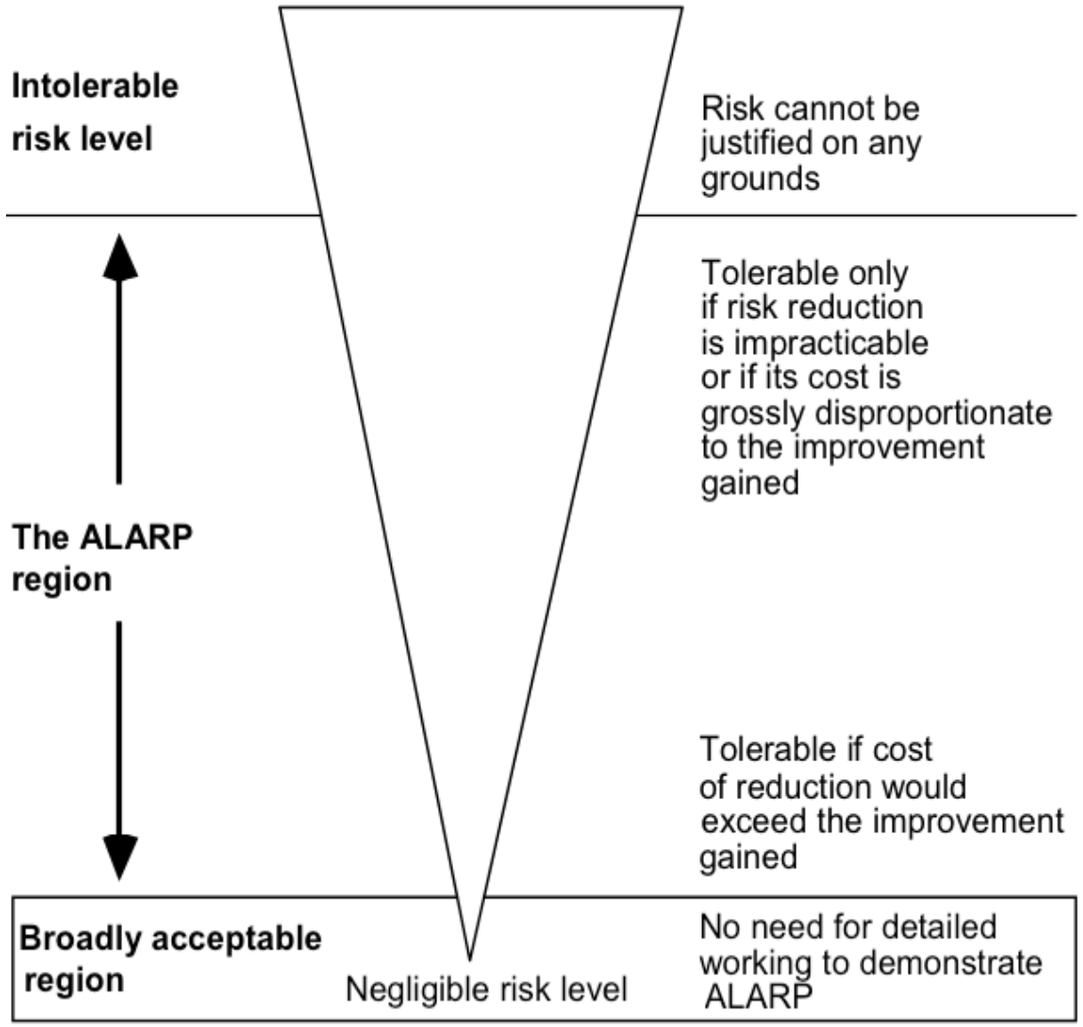
- Complex adaptive
  - Large scale, complicated
  - System properties and emergent behaviours, tipping points, phase transitions, cascades
  - Adaptive with people and society as part of the system or part of openness
  - Limits of reductionism, potential for surprise
- Finance sector and computer based trading
  - “Meltdown in the markets: a nuclear perspective”
  - Swans, Dragons, Cats, Toads and Bugs

# Nuclear reactor protection and control

# Protection parameters

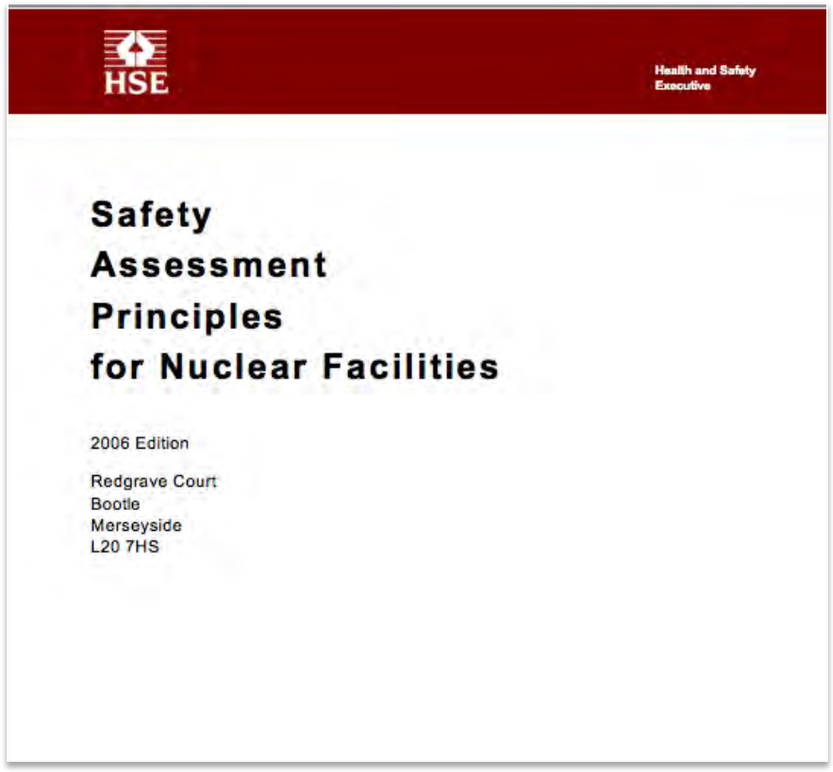


# Systemic risk definition and evaluation



# Safety Assessment Principles (SAPS)

- SAPS cornerstone of UK licensing approach
  - Goal based, based on eight fundamental principles
  - responsibility,
- leadership,
- understanding,
- fairness,
- inter-generational issues,
- protection,
- recovery and mitigation.



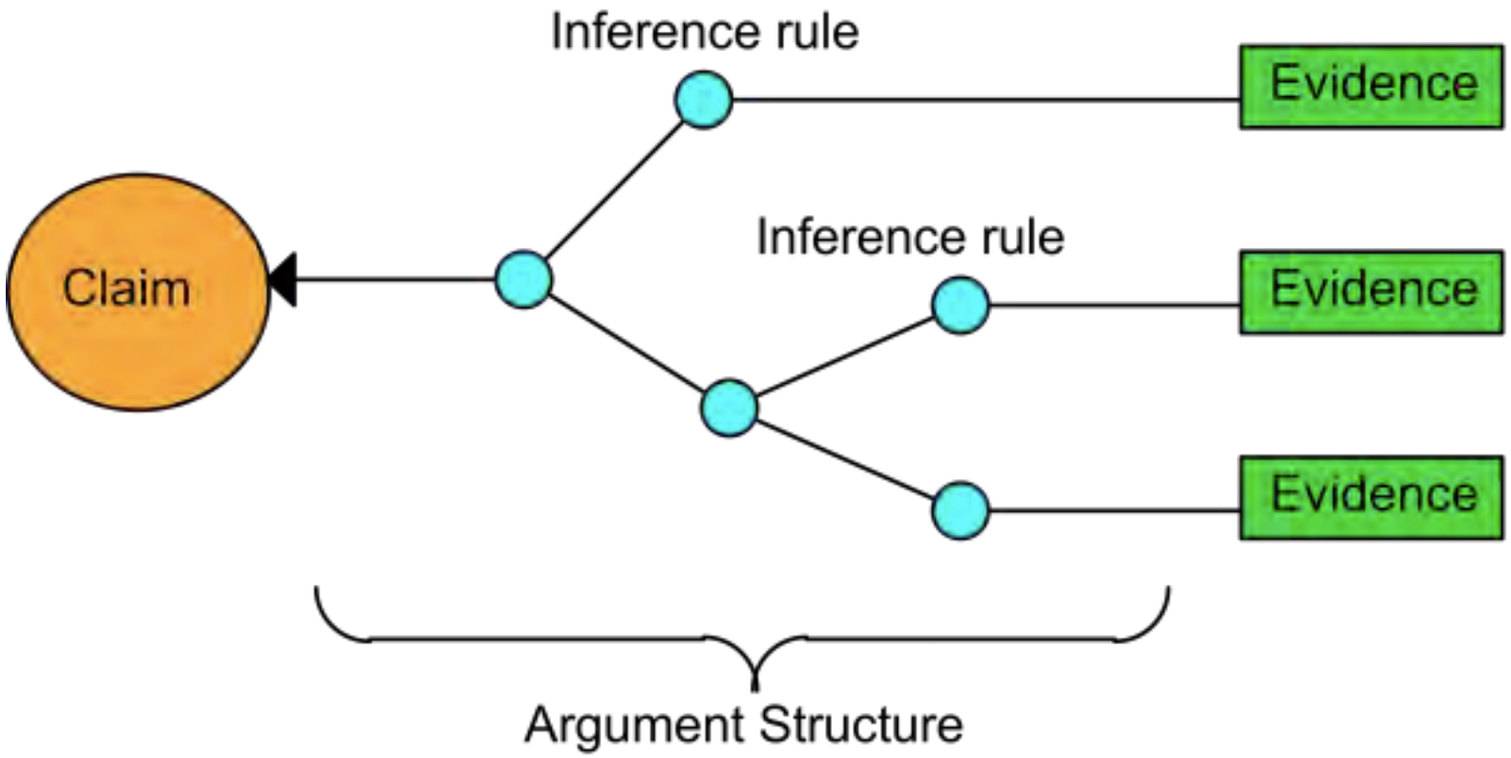
# Safety cases – regulatory obligation

- Safety cases are required by licence conditions
- The Conditions are non-prescriptive and set goals that the licensee is responsible for meeting, amongst other things by applying detailed safety standards and establishing safe procedures for the facilities.
- A "safety case" is defined as
  - the document or documents produced by the licensee documentation to justify safety during the design, construction, manufacture, commissioning, operation and decommissioning phases of the installation.
- Safety Assessment Principles (SAPs) describe safety case process and principles to be covered
  - “.... the establishment of and compliance with appropriate standards and practices throughout the software development life-cycle should be made, commensurate with the level of reliability required, by a demonstration of ‘production excellence’ and ‘confidence-building’ measures.”

# Communication and reasoning

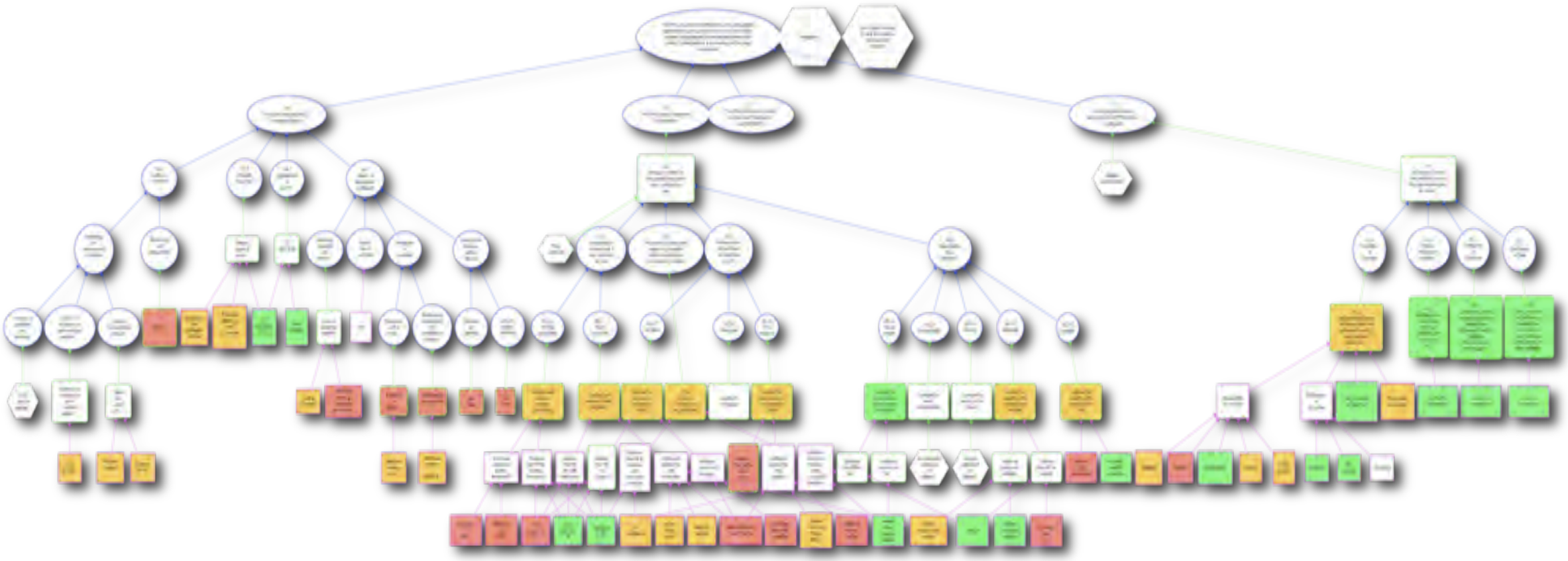
- Structured safety and assurance cases have two essential roles:
  - communication is an essential function of the case, from this we can build confidence
    - boundary objects that record the shared understanding between the different stakeholders
  - a method for reasoning about dependability (safety, security, reliability, resilience ...) properties of the system
- Both are required to have systems that are trusted and trustworthy

# Concept

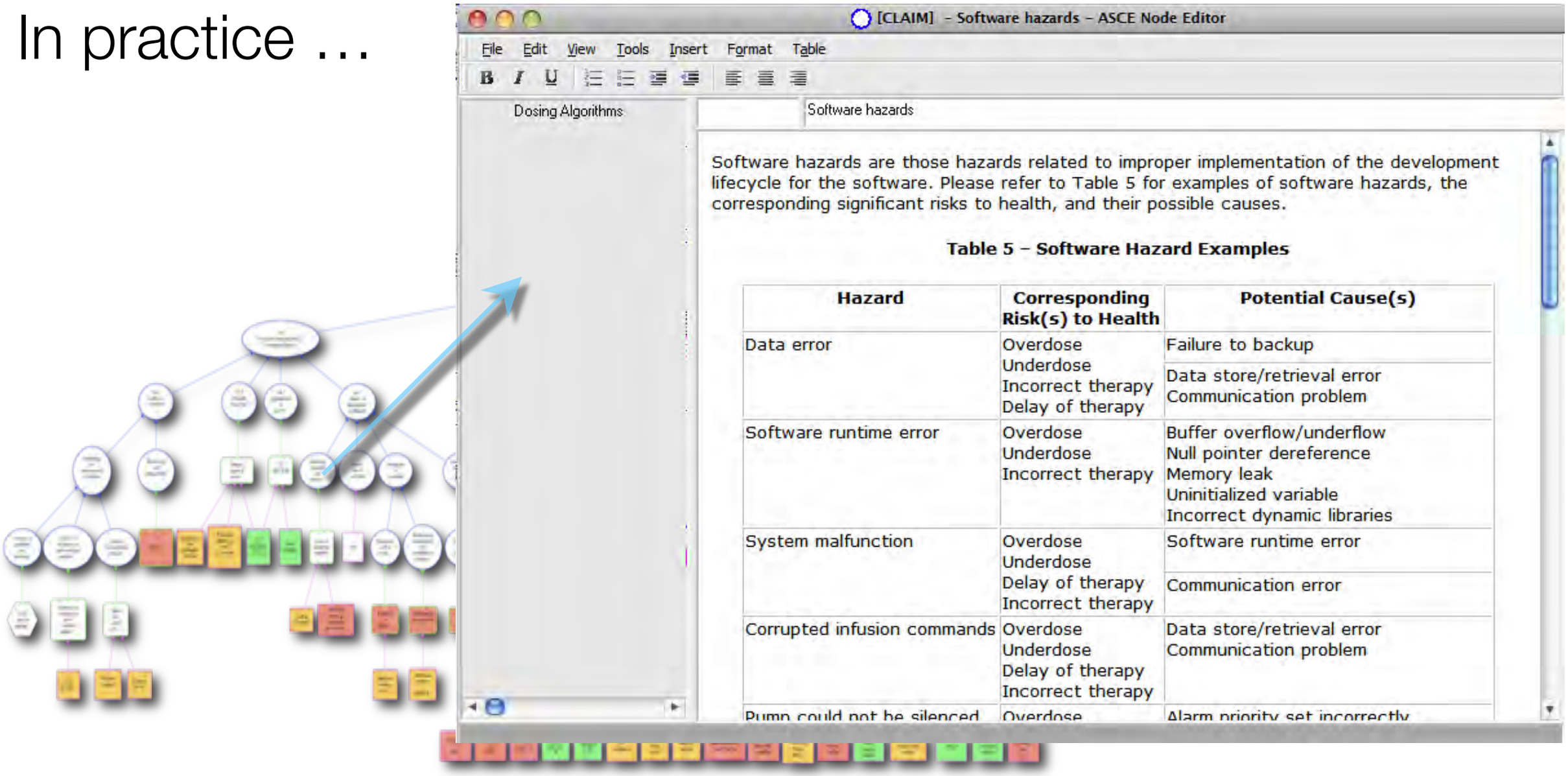



- "a documented body of evidence that provides a convincing and valid argument that a system is adequately safe for a given application in a given environment"

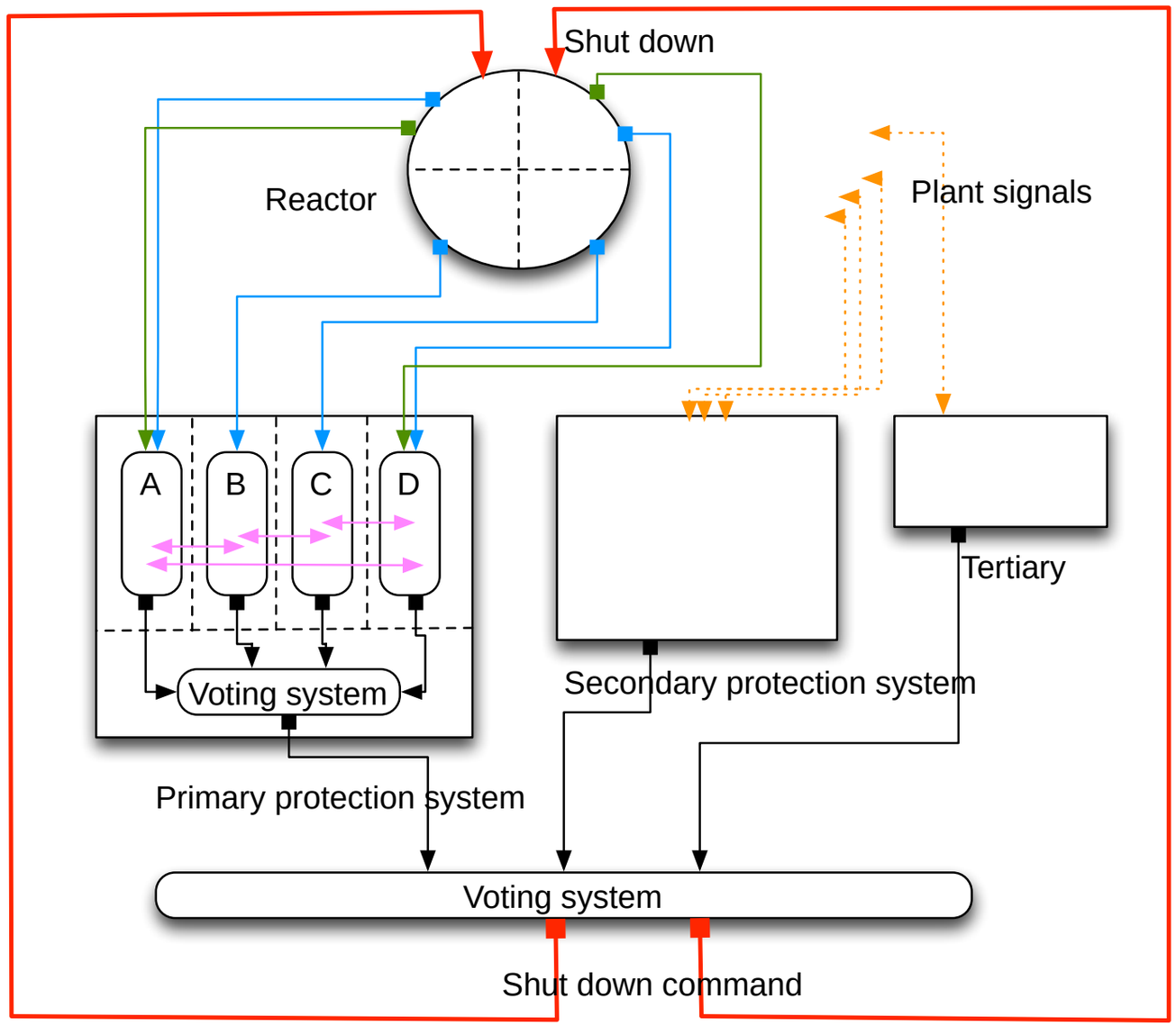
# In practice … the engineering



# In practice …


[CLAIM] – Software hazards – ASCE Node Editor
File Edit View Tools Insert Format Table
Dosing Algorithms
Software hazards
Software hazards are those hazards related to improper implementation of the development lifecycle for the software. Please refer to Table 5 for examples of software hazards, the corresponding significant risks to health, and their possible causes.
Table 5 – Software Hazard Examples
Hazard
Corresponding Risk(s) to Health
Potential Cause(s)
Data error
Overdose
Underdose
Incorrect therapy
Delay of therapy
Failure to backup
Data store/retrieval error
Communication problem
Software runtime error
Overdose
Underdose
Incorrect therapy
Buffer overflow/underflow
Null pointer dereference
Memory leak
Uninitialized variable
Incorrect dynamic libraries
System malfunction
Overdose
Underdose
Delay of therapy
Incorrect therapy
Software runtime error
Communication error
Corrupted infusion commands
Overdose
Underdose
Delay of therapy
Incorrect therapy
Data store/retrieval error
Communication problem
Pump could not be silenced
Overdose
Alarm priority set incorrectly

# Reactor protection systems




diverse signals A, B, C, D redundant channels

segregation

communication between channels

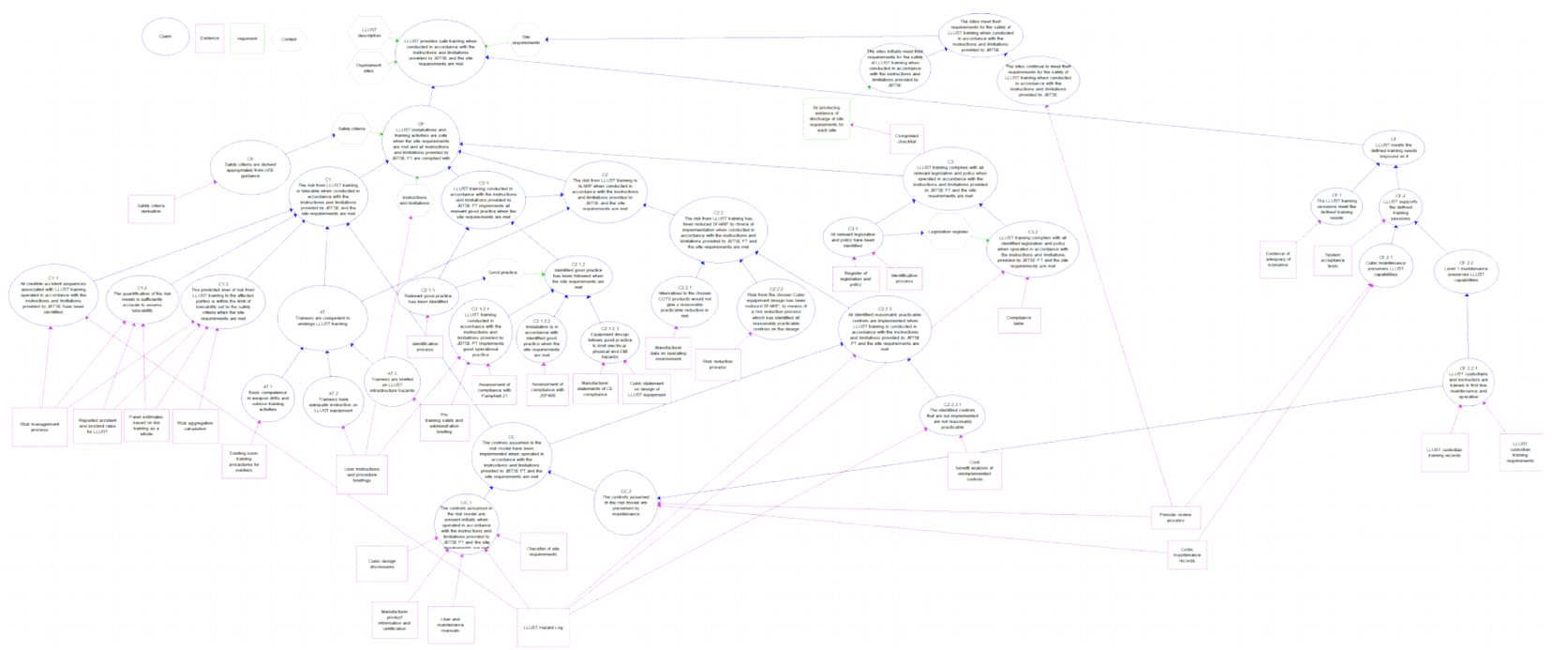
Also ESFAS for post trip control and protection

# Techniques

- goal-based assessments  of attributes (e.g. demonstration of accuracy, reliability, correctness etc.)
  - assessment of potential vulnerabilities and formal demonstration of integrity properties (e.g. deadlock, run time exceptions caused by divide by zero)
  - static analysis of software source code with mathematically formal demonstration of functional correctness of the code with respect to a specification
  - dynamic analysis of the software with test coverage metrics; statistical testing commensurate with the level of reliability required
- demonstration that any software tools do not comprise the integrity of the code.
- a post-installation demonstration, performed independently of the suppliers, that the system performs to requirements
- analysis of experience from previous use of the system (and previous versions of the system) in the nuclear industry or elsewhere

# Do everything and do it at least twice.

- very careful lifecycle
- static analysis and proof
- statistical testing
- negative testing
- compliance

# Some issues and challenges

In engineering trusted systems

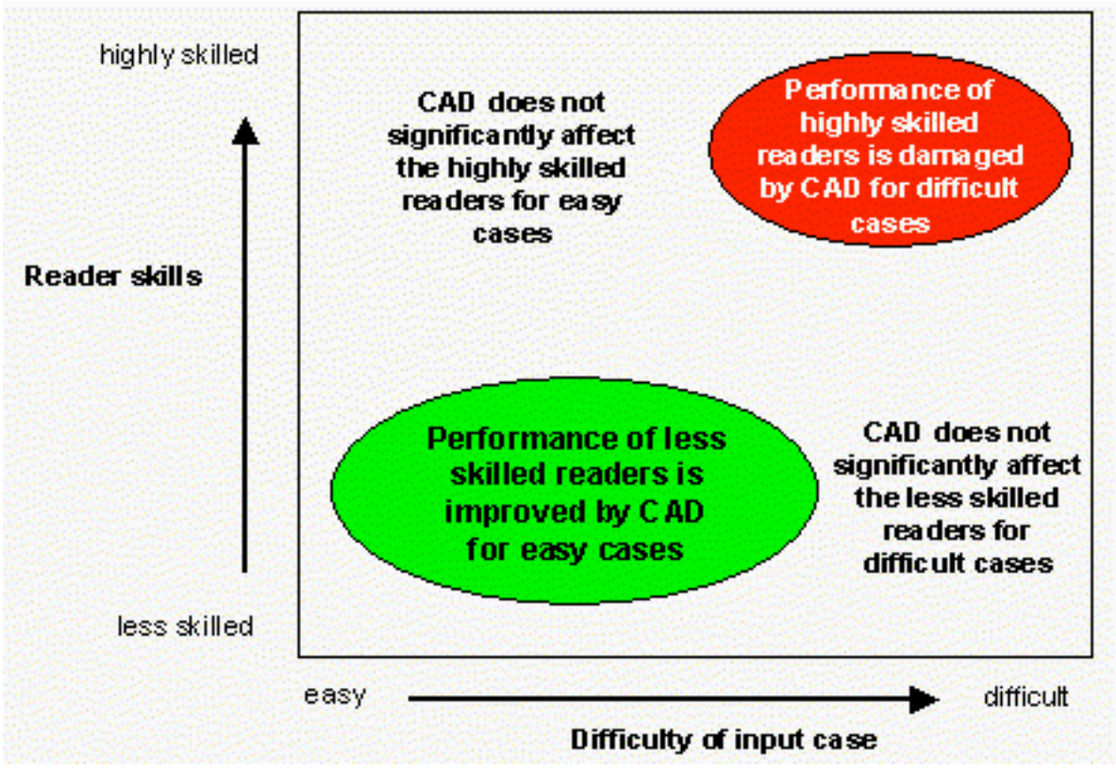
- Defence in depth and diversity
- Interdependencies and resilience
- Adaptation and socio-technical systems
- Confidence and doubt
- Learning form experience – difficult and important
- Security and threats

# The human element

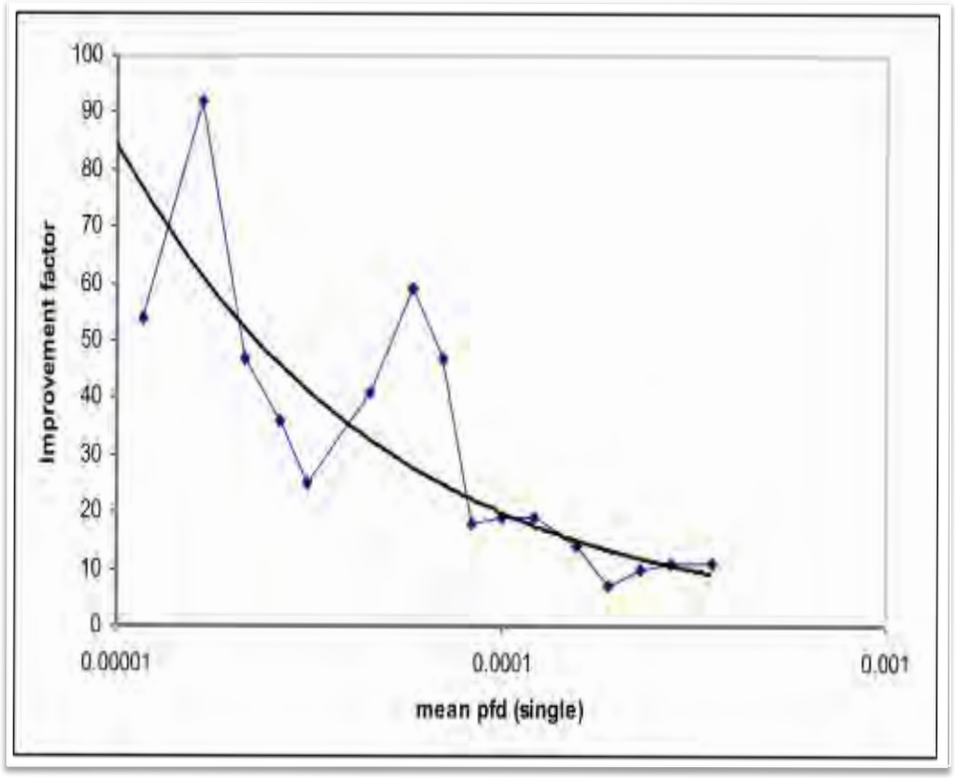
- people as a source of resilience, as a threat and as victims
- don’t blame the pilot – but the system
- 30 min rule, tempo and expertise

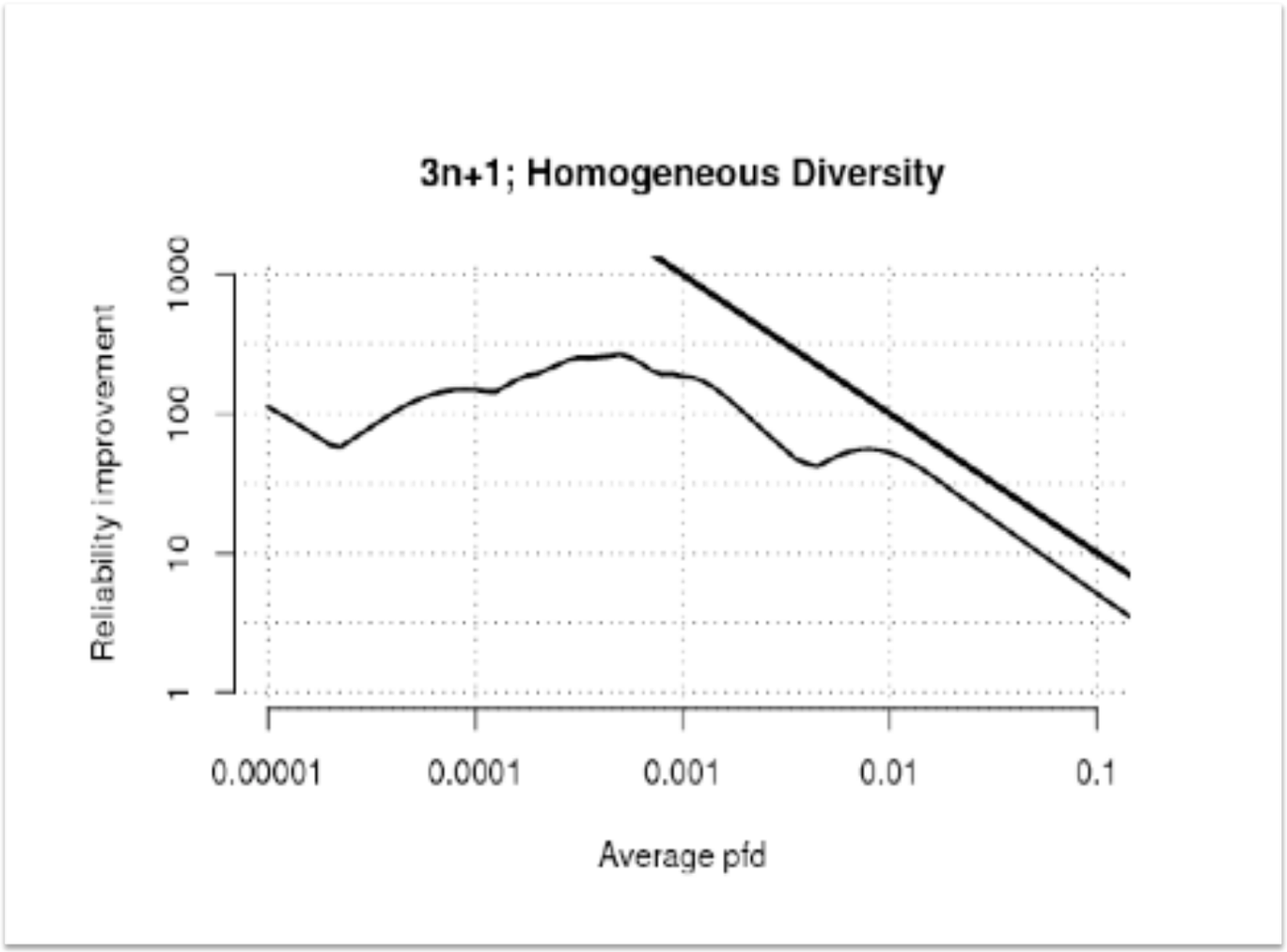
# Socio-technical, adaptation

- A socio-technical perspective on assurance cases:
  - In addition to claims that physical hazards, security threats have been addressed
  - Define a range of vulnerabilities (narrow scope, misaligned responsibilities, undifferentiated users, adaptation, automation biases, non-independence of arguments) and develop arguments of how they might be addressed.
- Develop methods for review wrt socio-technical issues
- Ideas taken from EPSRC INDEED and DIRC projects



# Defence in depth and diversity




- Improvement trend (low pfd subsets) Knight and Leveson

Van der Meulen

# Conservative approach

- Figure 6a shows a typical 'real' distribution, f(p) that satisfies the expert's belief:

$$P(pfd < y) = 1 - x$$

Of all the many such distributions that satisfy his belief, Figure 6b shows the most conservative: here all the probability mass for the interval (0,*y*) is concentrated at *y*, and all the mass for the interval (*y*,1) is concentrated at 1. Then:

$$P(\text{system fails on randomly selected demand }) = \int_0^y p.f(p) + \int_y^1 p.f(p) < y(1-x) + x$$

$$= x + y - xy = z, \text{ say}$$

Figure 6a

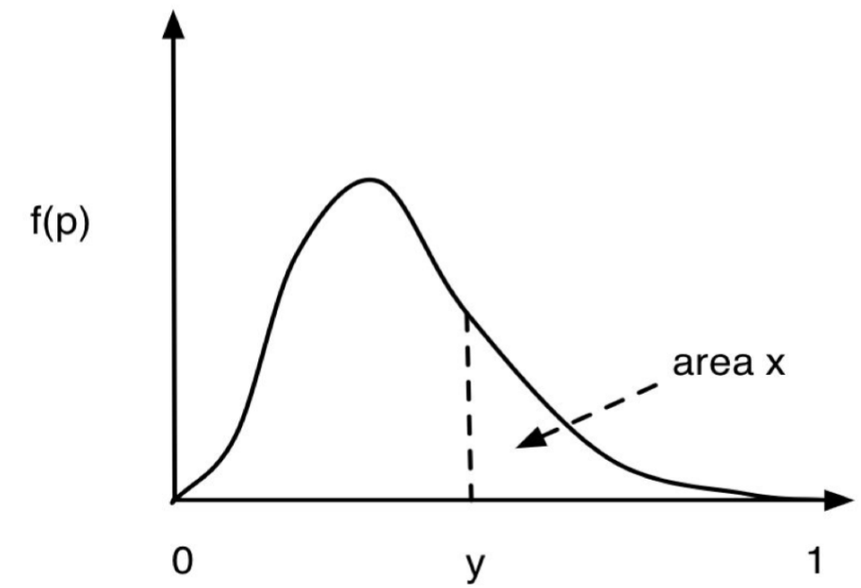
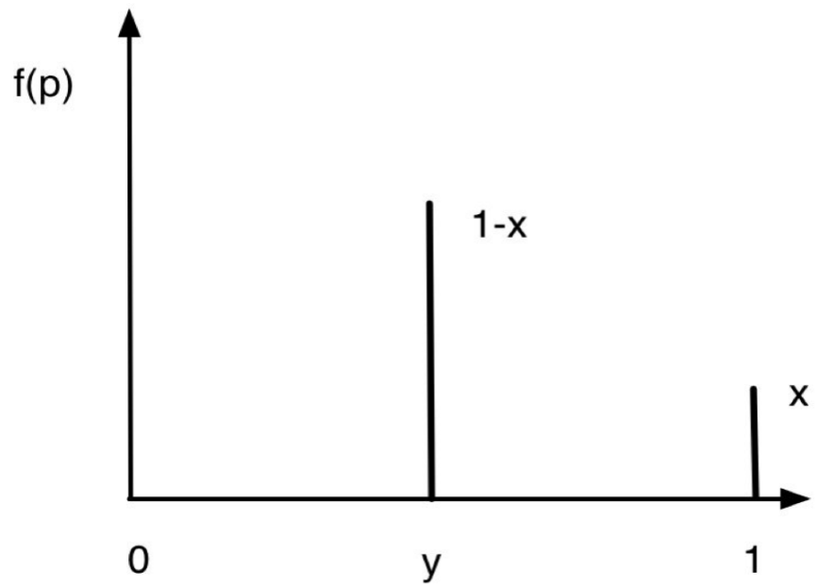



Figure 6b



# Trojan Horse Virus, 28 yrs before Stuxnet

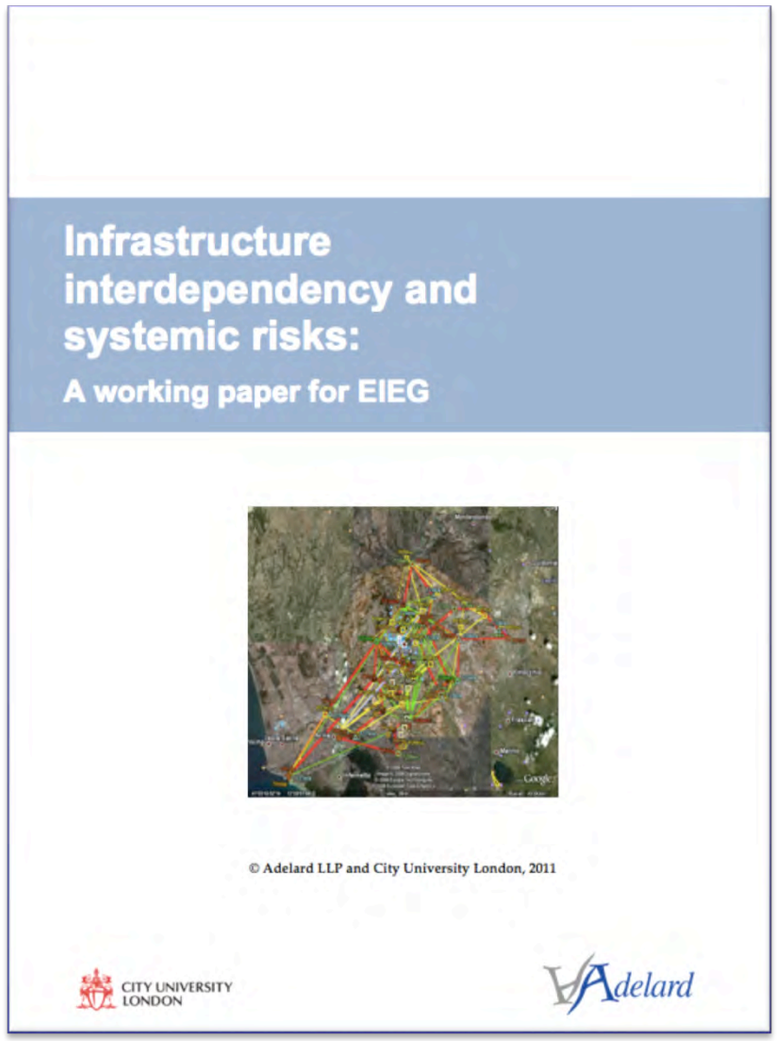
- SCADA Attack In June 1982, malicious code implanted in SCADA software caused a rupture of the trans-Siberian gas pipeline.

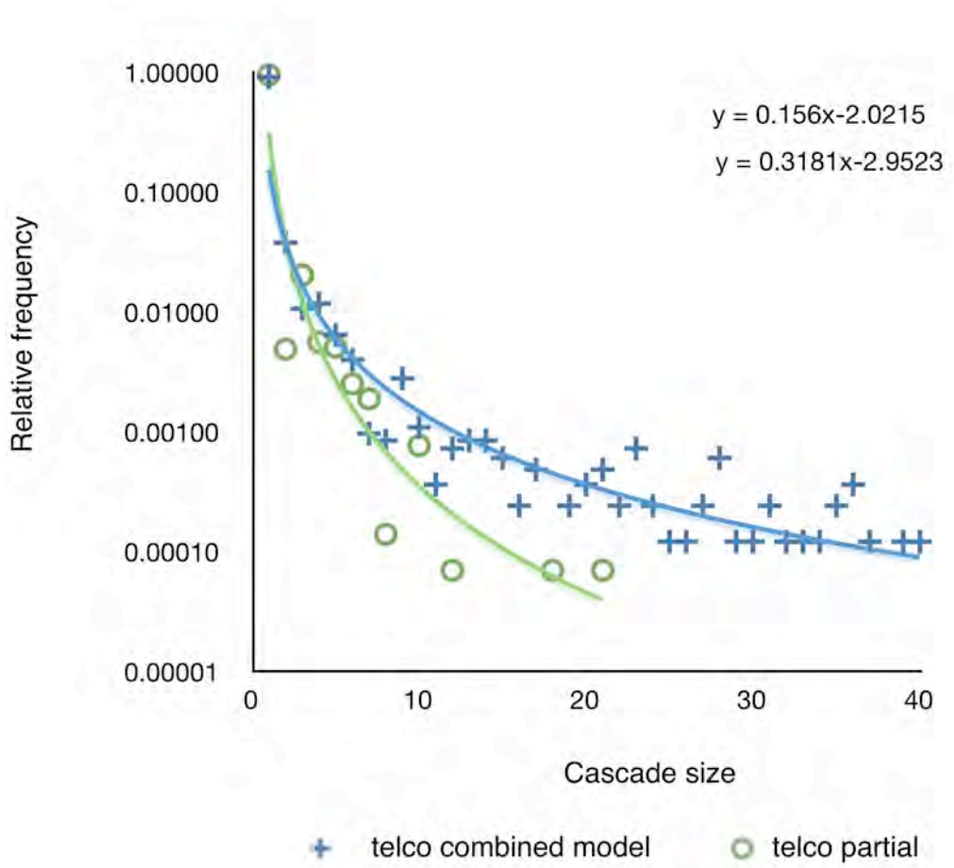
- According to Reed, “the pipeline software that ran the pumps, turbines, and valve settings was programmed to produce pressures far beyond those acceptable to the pipeline joints and welds.” This caused the largest non-nuclear explosion on record (3KT).

- Thomas Reed, At the Abyss: An Insider’s History of the Cold War

- This is disputed by other sources, issue of *attribution*

# Critical infrastructure interdependencies



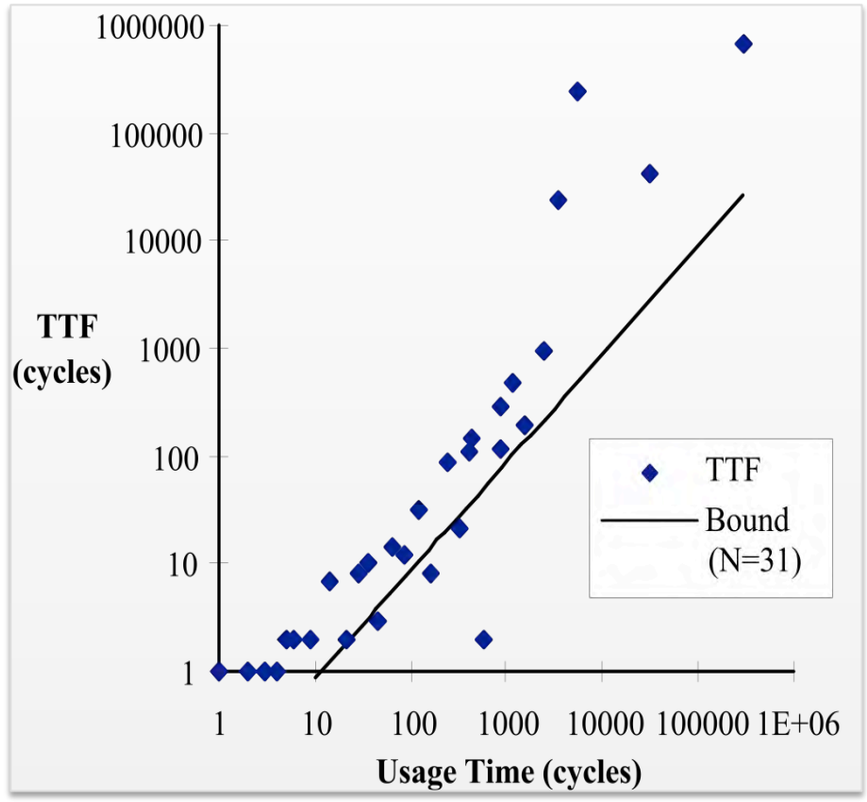




CSR Building confidence in a computerised world
www.csr.city.ac.uk

ADELARD

# Analysis of field experience

- Use of field experience to complement the justification
  - When information is lacking
  - To gain confidence
  - Identify vulnerabilities
- Review of defects to check impact on application
- Looking for evidence of reliability growth
  - Long-term, conservative models
- Depends on quality of data
  - Normally only part of case



CSR Building confidence in a computerised world

www.csr.city.ac.uk

ADELARD

# Summary

- Established approaches for evaluating and communicating trust in critical computer based systems
- Plenty of issues
  - Defence in depth and diversity
  - Interdependencies and resilience
  - Adaptation and socio-technical systems
  - Confidence and doubt
  - Learning form experience – difficult and important
  - Security and threats

# Flash crash - May 2010


**Preliminary Findings Regarding**
**the Market Events of May 6, 2010**

**Report of the Staffs of the CFTC and SEC to the Joint Advisory Committee on Emerging Regulatory Issues**


Building confidence in a computerised world
www.csr.city.ac.uk

ADELARD

# Black swans - power laws and fat tails

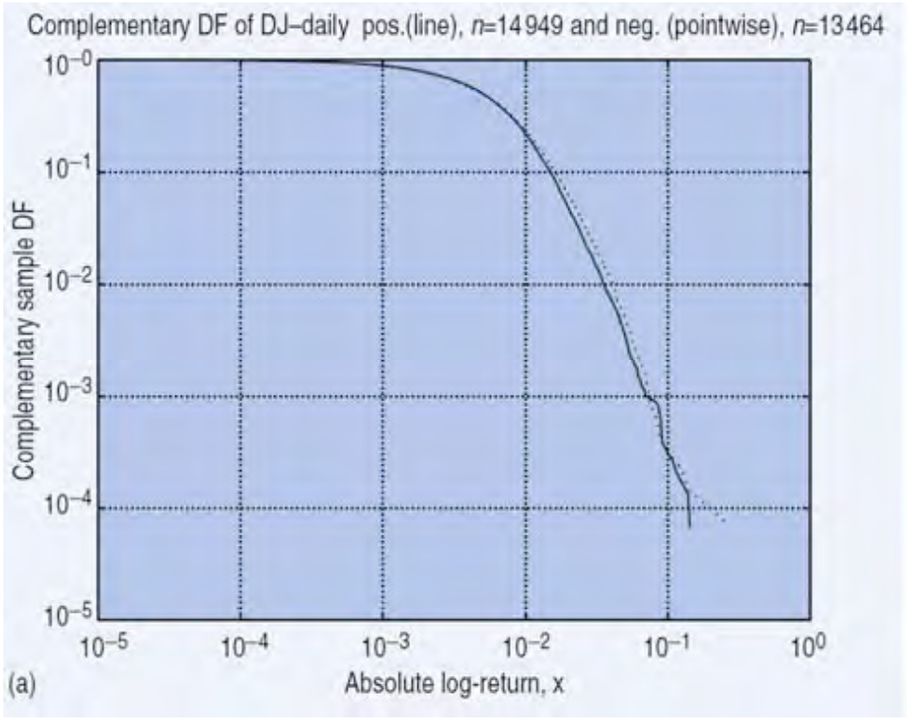
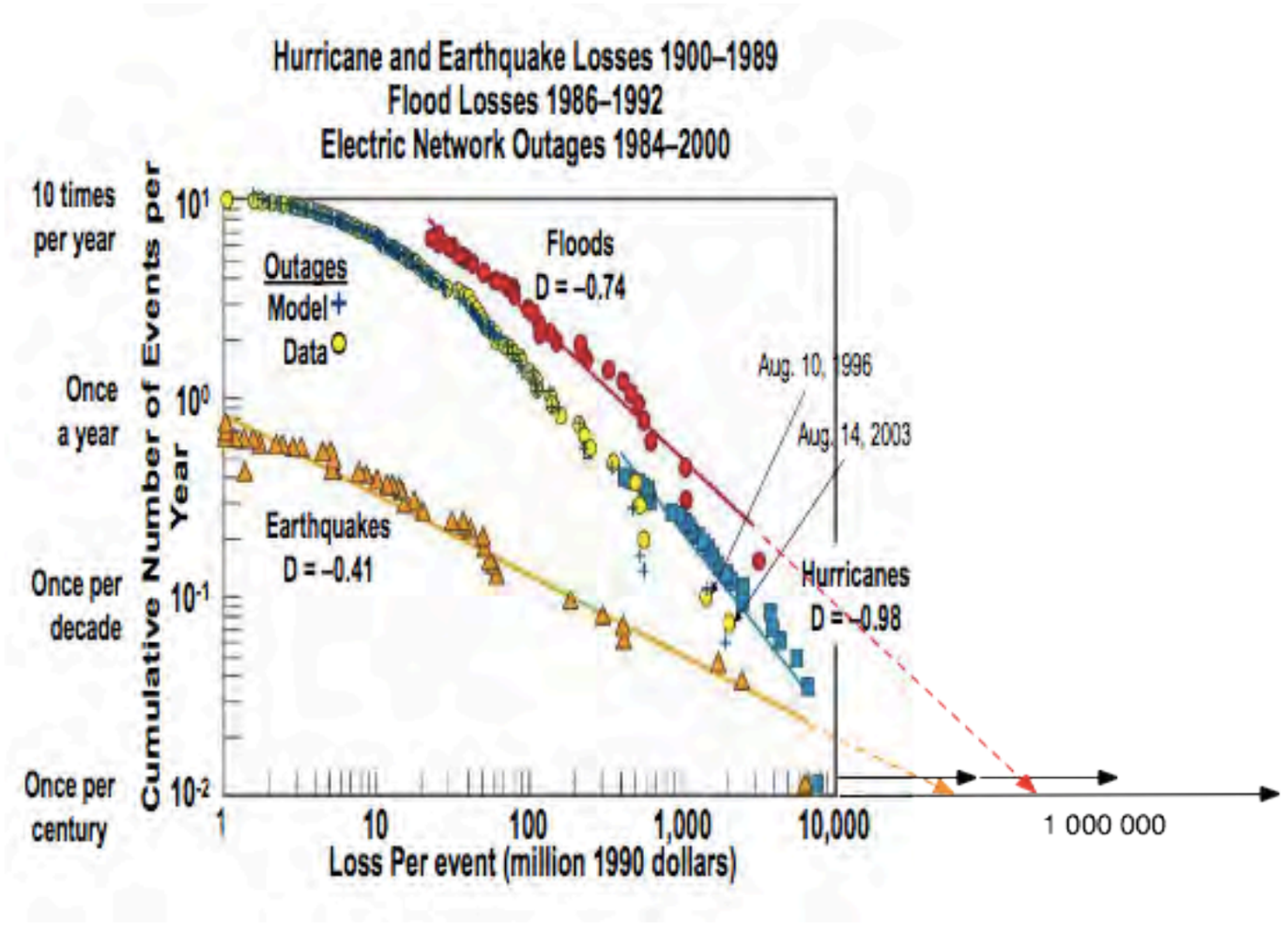



Fig.7 Survival distribution of positive (continuous line) and negative daily returns (dotted line) of the Dow Jones Industrial Average index over the time interval from May 27, 1896 to May 31, 2000, which represents a sample size of n=28 415 data points. The straight part in the tail in this log-log scale qualifies a power law distribution with exponent μ≈3. Reproduced from Malevergne et al. [8].



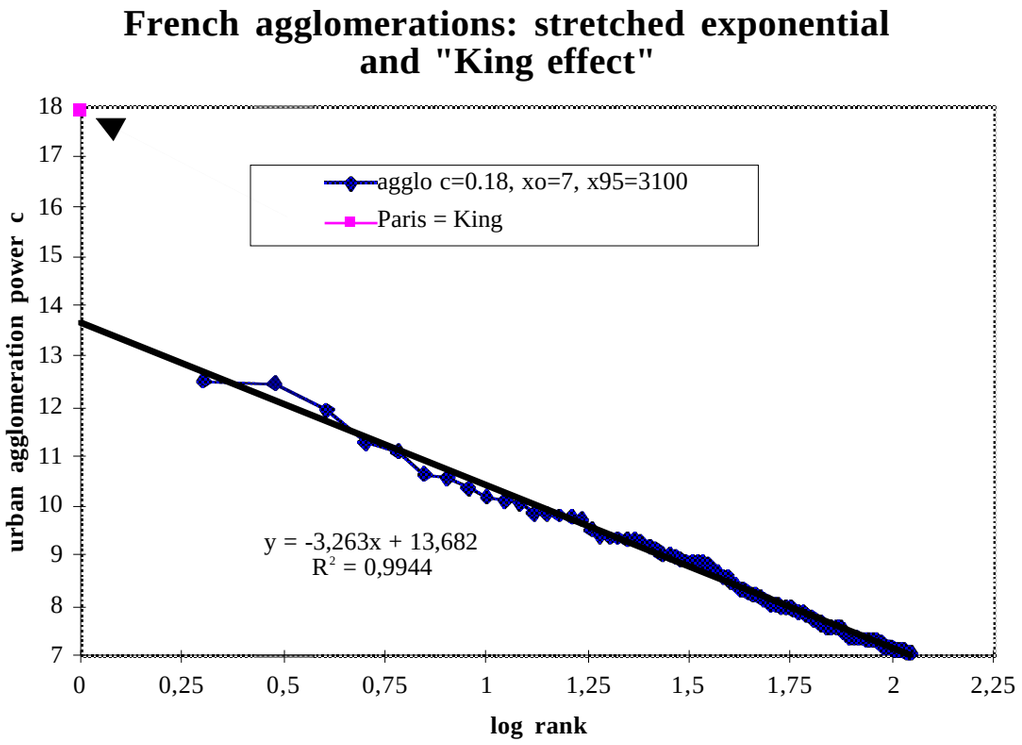

Visually power laws but see critiques

# Financial crashes – no precursors?

- power law distributions embody the notion that extreme events are not exceptional events.
- instead, extreme events should be considered to be rather frequent and to result from the same organization principle(s) as those generating other events: because
    - they belong to the same statistical distribution, this suggests common generating mechanism(s).
- a great earthquake is just an earthquake that started small ... and did not stop; it is inherently unpredictable due to its sharing of all the properties and characteristics of smaller events (except for its size), so that no genuinely informative precursor can be identified
- view expounded in formulation of the concept of self-organized criticality and also the espoused by the “Black Swan Theory” which views high-impact rare events as unpredictable.
- (adapted from D Sornette, Why stock markets crash )

# Dragons

- Even with power laws and other fits get outliers



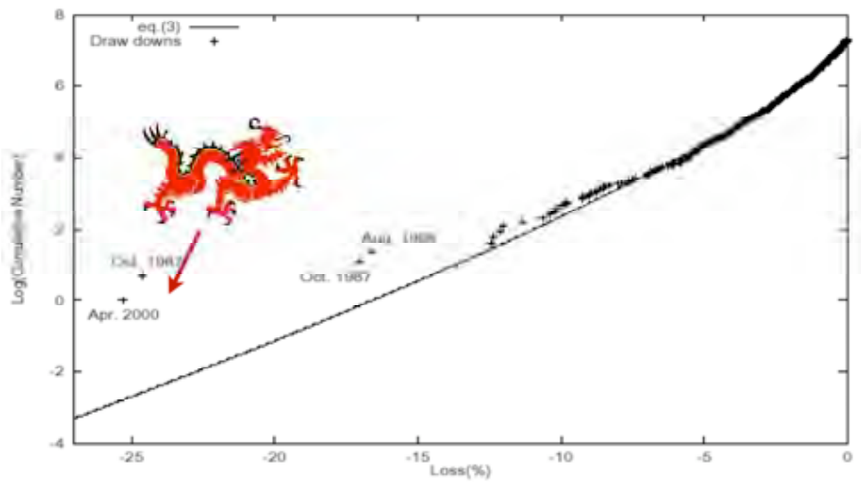
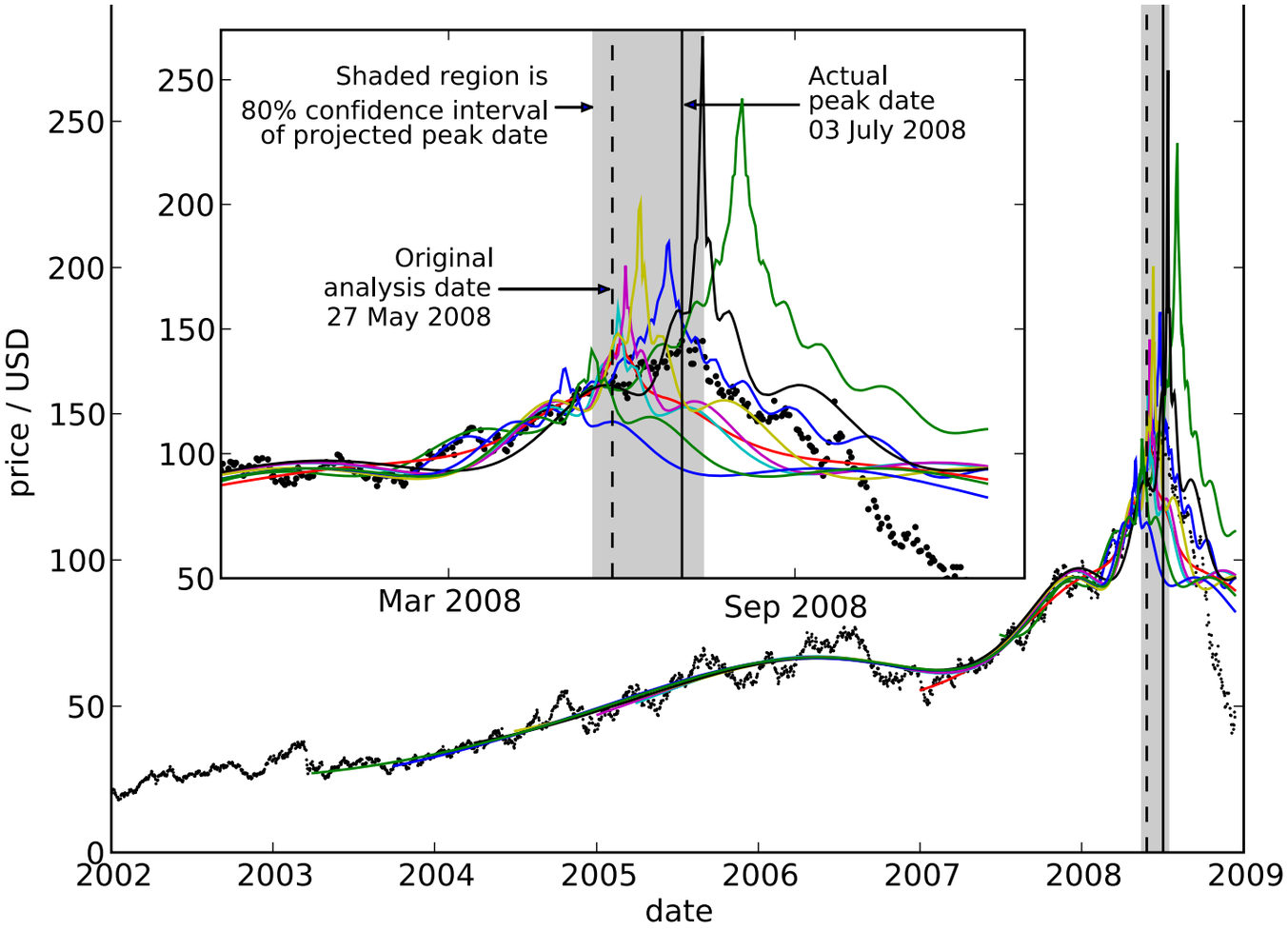
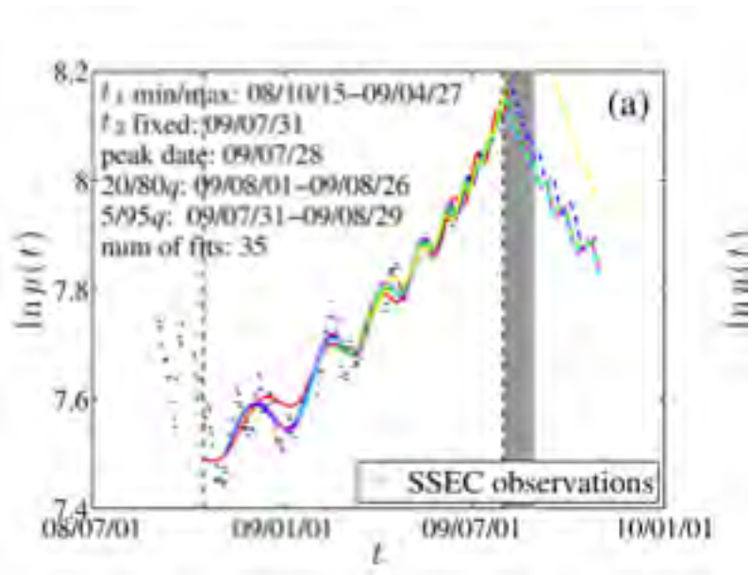




Fig.15 Distribution of drawdowns D for the Nasdaq Composite index, showing several "outliers" in the tail, that qualify as dragon-kings. It turns out that these anomalous events are

# Predicting bubbles




$$\log p(t) = A + B(t_c - t)^m + C(t_c - t)^m \cos(\omega \ln(t_c - t) - \phi) \qquad (2)$$

# Pencil analogy

Long-term collaborative or herding behavior produces unstable system
Can measure potential energy changes

# Frogs

Common toads appear to be able to sense an impending earthquake and will flee their colony days before the seismic activity strikes.

The evidence from a population of toads which left their breeding colony three days before an earthquake that struck L'Aquila in Italy in 2009.

How toads sensed the quake is unclear, but most breeding pairs and males fled.

They reacted despite the colony being 74km from the quake's epicentre, say biologists in the Journal of Zoology.

http://news.bbc.co.uk/earth/hi/earth_news/newsid_8593000/8593396.stm

Abstraction can remove key indicators

# Performative models

- In the past, in the engineering domain, the models used to design and assess the risks do not affect the threats or challenges that the system faces
- modelling severe weather does not change the wind speed in London
  - (except perhaps via a slow political process and peoples' behaviour)
- in the financial area this is not the case: models can be what is termed *performative*, having a direct and unforeseen impact on the markets and how it fails
- engineered systems and security risks
- knowledge and access to design models may inform an adversary and hence have a potential impact on the threats a system faces
- *Donald MacKenzie , An Engine, Not a Camera: How Financial Models Shape Markets and also Do Economists Make Markets?: On the Performativity of Economics*

# Trust in computer-based systems (and fpgas)

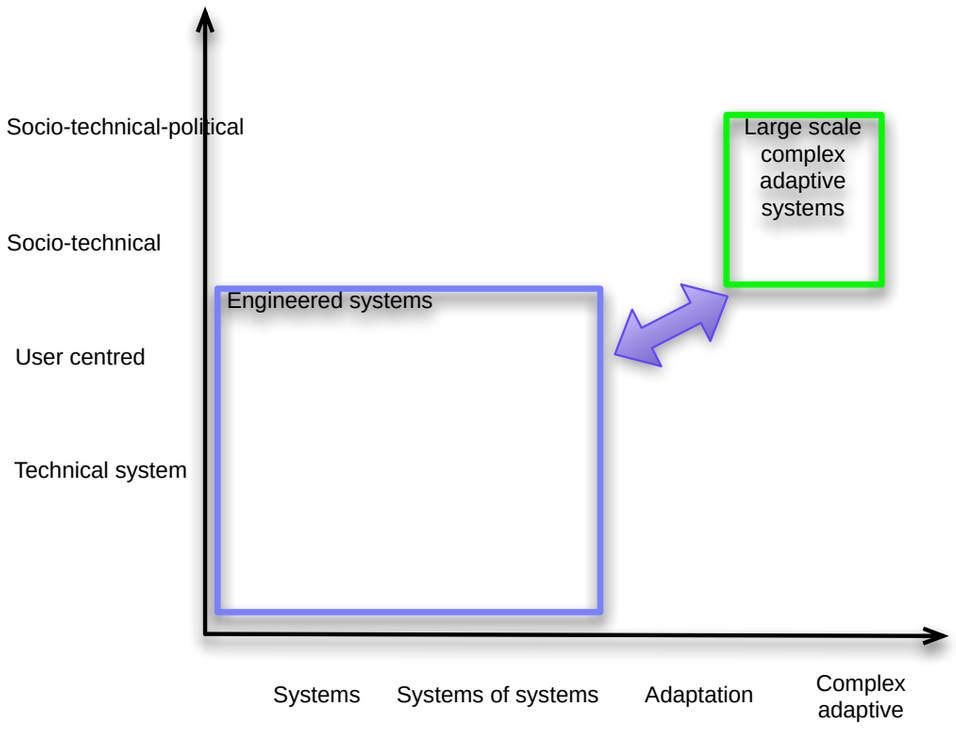
- Recent work by Johnson
  - analysed a set of 18,520 ultrafast black swan events in stock-price movements between 2006 and 2011
  - empirical evidence for an abrupt transition from a mixed human-machine to a all-machine phase (<650ms for crashes, <950ms for spikes)
- How much do computers need to be trusted – the problem of “bugs”
  - Impact on themselves,
  - On platform they are trading in
  - On wider system
    - issues of risk apportionment, of composition and confidentiality

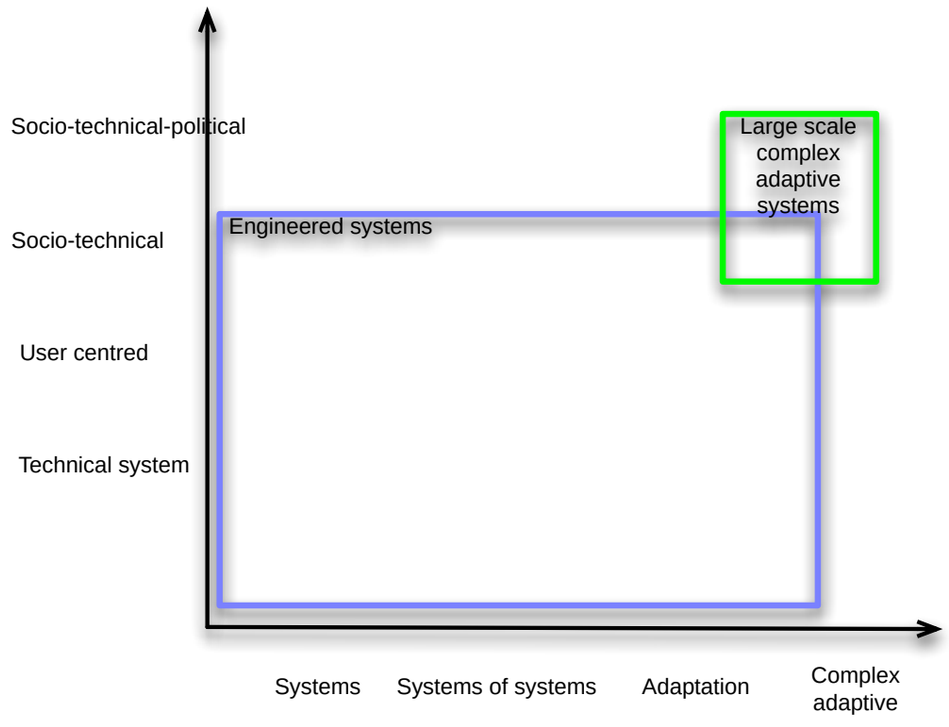
# Engineering complex adaptive systems

- Can simplified envelopes of operation be defined? Can we define control strategy for unstable system?
- What would be the parameters that need to be measured and what are we trying to infer?
- What would the availability and reliability requirement be for such a system. E.g. the probability of failure on demand, the frequency of spurious activation?
- What does “safety” and resilience mean in the HFT context. Would need to consider both the platform and the application?
- What is the balance between automation and operator recovery. Who would design this?
- If crashes are hard to anticipate should the focus not shift to recovery and resilience?

# Engineering complex adaptive systems

- swans, cats, dragons, toads and bugs




Developing understanding to engineer complex adaptive systems will change how we currently engineer systems

CSR Building confidence in a computerised world
www.csr.city.ac.uk

ADELARD

# Some basic questions

- In engineering complex adaptive systems we would like to answer:
- What is the system, what are the risks, who has them, and when, are they tolerable?
  - What is the probability of a crisis and what would the consequences be?
  - What are the risks from the the complex adaptive system to the economy  and society as a whole?
  - What risk levels are tolerable and who/how has this been decided?

# Engineering complex adaptive systems

- *Developing intervention strategies for different time bands*
  - Circuit breakers and trip protection
  - Forced diversity in ecology, disrupt corellation
  - Alignment of economic incentives
- *New approaches to risk assessment*
  - Investigating safe envelopes and recovery
  - Specification for single, group and collective
  - Deriving risk targets for properties - failure and success
  - Separation of conventional and complex risks (Fukushima vs flash crash)
  - Emphasis on stability as a dependability attribute
  - The risk of change or not – focus?

Understanding complex systems - dynamics of change, signals

# Conclusions - challenges and opportunities

- Finance sector and computer based trading just one example of complex adaptive systems (automotive, air traffic, medical…)
- Challenges come from the complex adaptive nature of these systems and their risks and benefits
  - do we need a fundamental change to risk assessments, assess “small” changes, stability and cliff edge identification
- Opportunities from deploying and adapting our current engineering approaches to systemic risks
- Work in progress….
  - Hope to publish public draft in May 2012  “Meltdown in the markets: a nuclear perspective”

# Adelard

# Centre for Software Reliability

- Safety and assurance cases and safety management systems
- Independent safety assessment
- Software assurance, including formal methods and static analysis
- Development, interpretation and application of standards and guidelines
- applied research in safety, security, critical infrastructure interdependencies
- policy to technology
- ASCE – the Assurance and Safety Case Environment
- clients in nuclear, defence, financial, transport sectors

- Evaluation of socio-technical systems
- Technical, interdisciplinary
- Research
- with international community and users
- Education
- placements, internships, scholarships, courses, MSc and CPD
- Innovation
- director, Dr Peter Popov
- DivSQL, PIA-FARA

CSR Building confidence in a computerised world
www.csr.city.ac.uk

ADELARD