ADELARD

The Challenge of Complex Adaptive System Safety

Dragons, Swans, Cats, Toads, Bugs

Robin E Bloomfield
Adelard LLP and City University London
29 June 2012

Exmouth House 3–11 Pine Street London EC1R 0JH
T +44 20 7832 5850 F +44 20 7832 5853 E office@adelard.com W www.adelard.com

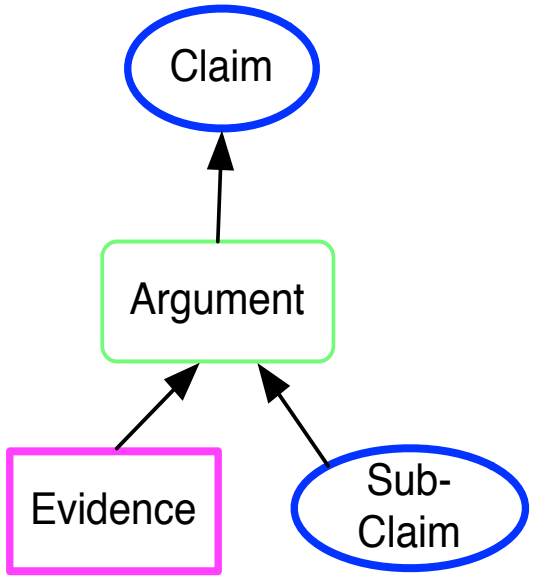
Introduction

1. Background
   - Focus on evaluation, practitioner, nuclear, ATC
   - UK Foresight –Computer trading and systemic risk: a nuclear perspective
   - ERTMS, CIP, Defence in depth
   - Confidence modelling
   - DEOS – Dependable Open Systems
   - Health foundation
2. Confidence and Assurance Cases
3. Some challenge of complex adaptive systems
   - What is current engineering state of practice
   - What are particularities of complex, adaptive systems
4. Future directions
   - Engineering complex adaptive systems

Assurance Cases

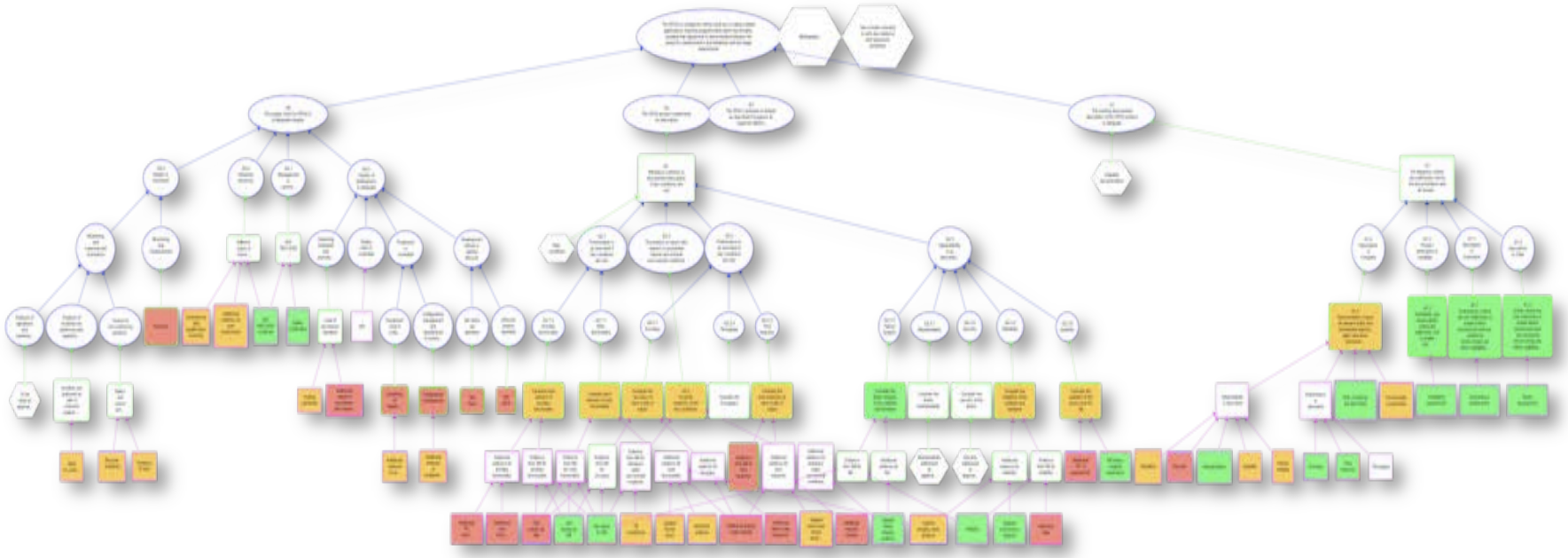
-

Concept

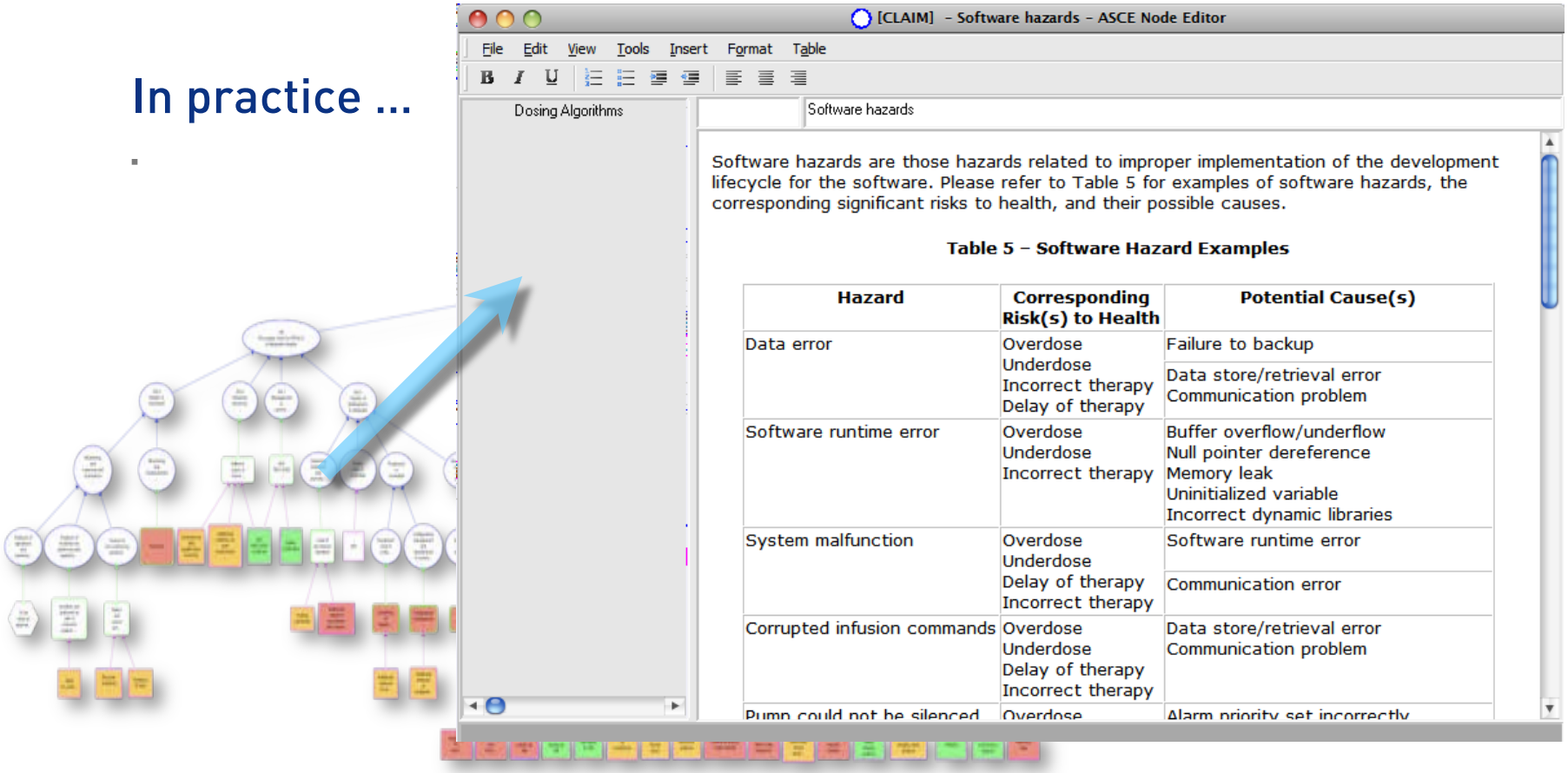


"a documented body of evidence that provides a convincing and valid argument that a system is adequately safe for a given application in a given environment"

In practice … the engineering



In practice …

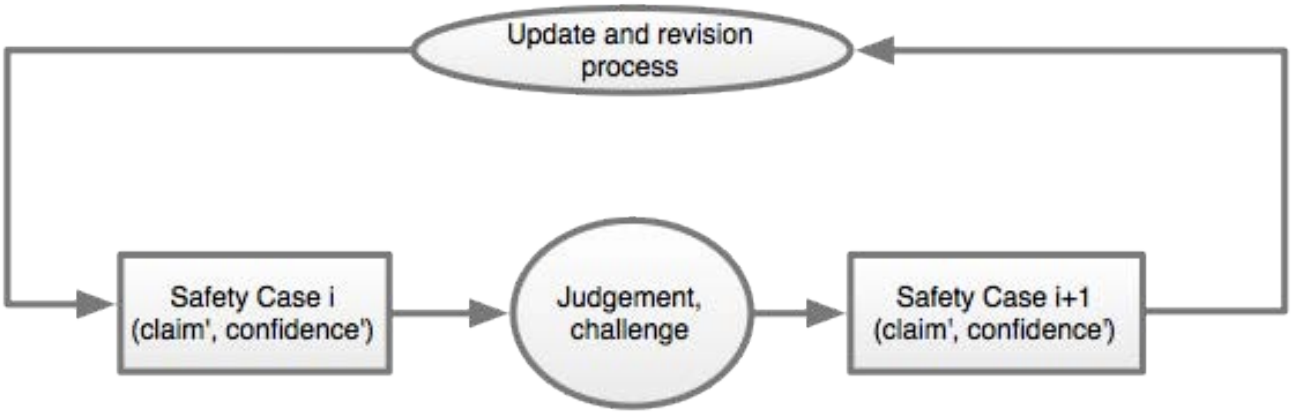


Communication and reasoning

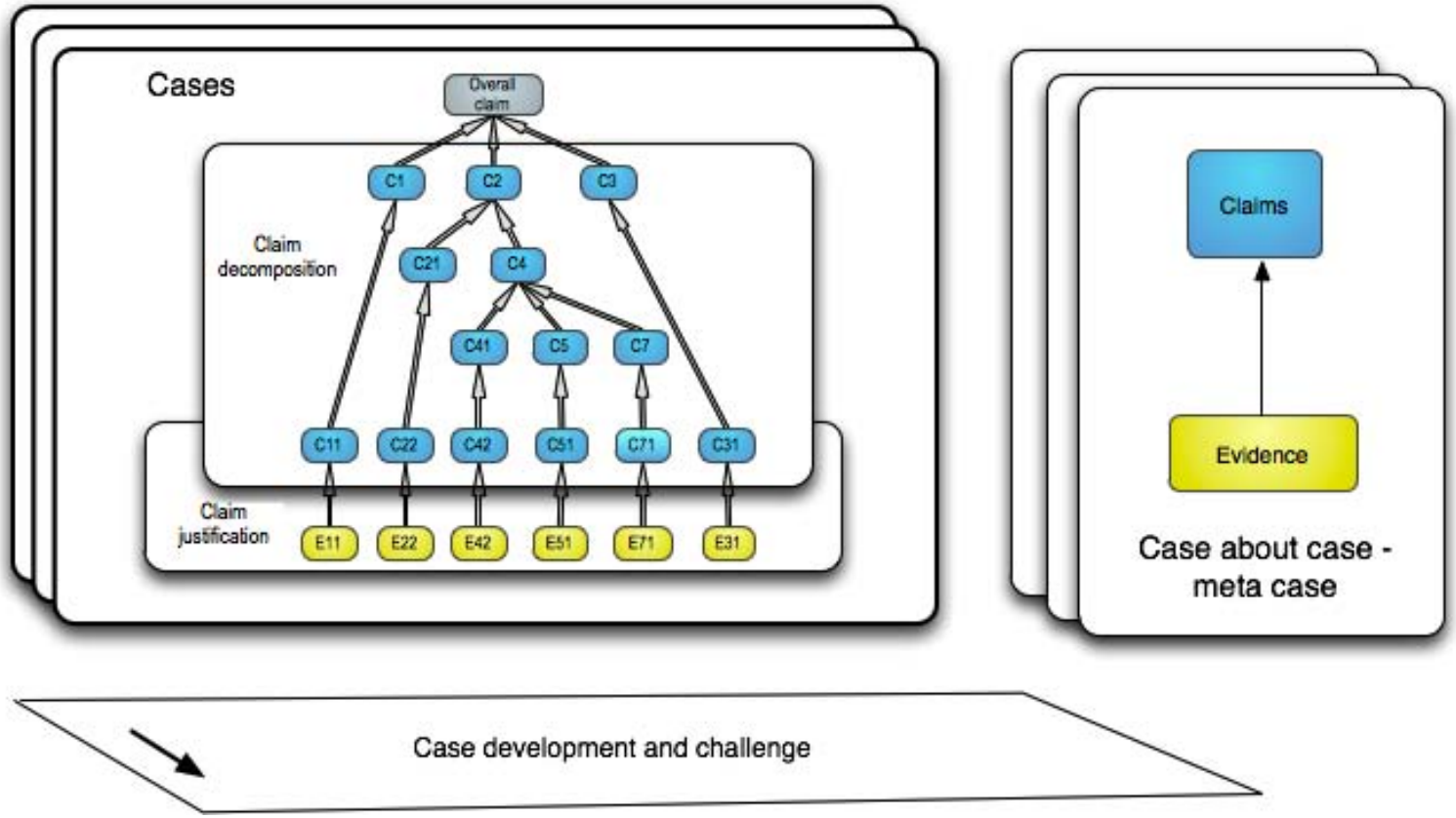
1. Structured safety justification has two roles:
   - communication is an essential function of the case, from this we can build confidence
     - boundary objects that record the shared understanding between the different stakeholders
   - a method for reasoning about dependability (safety, security,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

Safety justification process – building confidence, challenging assumptions

1. Captured in safety management system and in meta-case
2. Challenge and response cycle essential
3. Proof as a social, technical, adversarial process

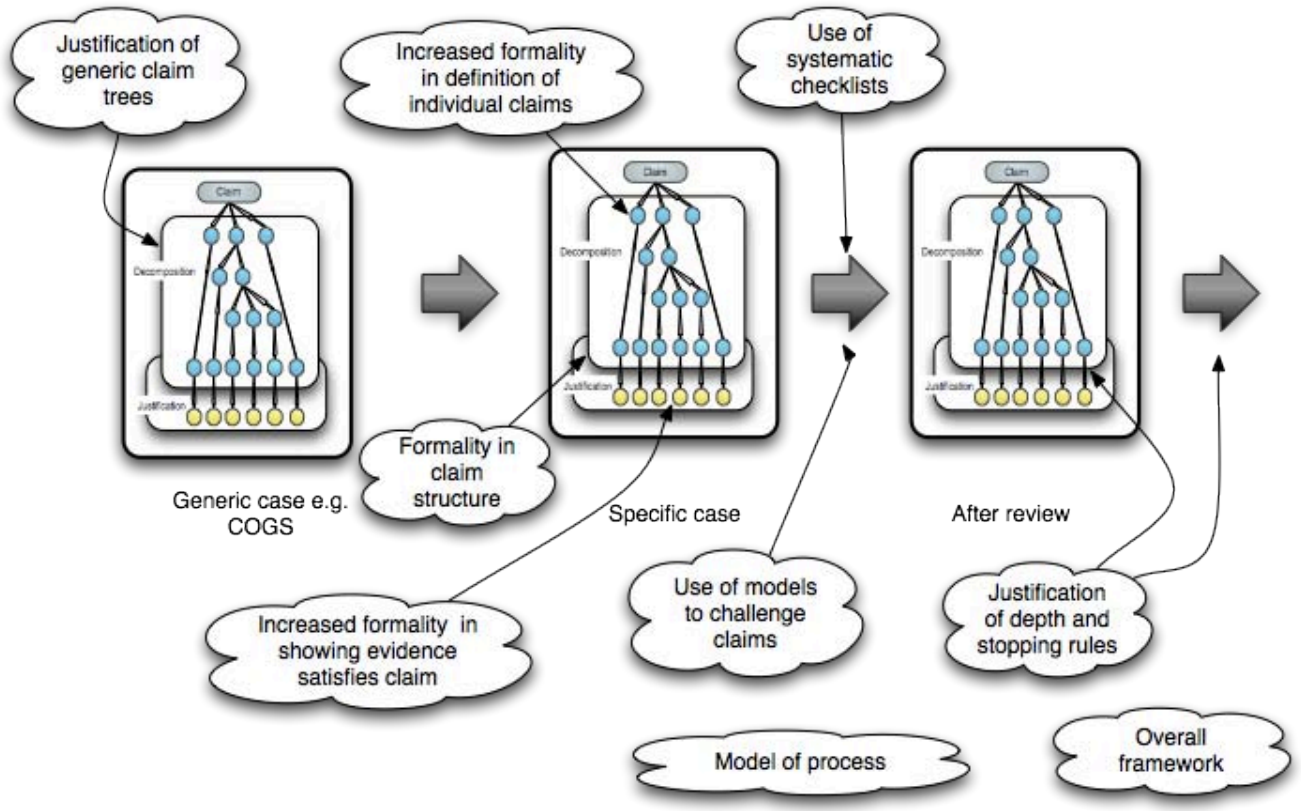


Reasoning, communication, confidence



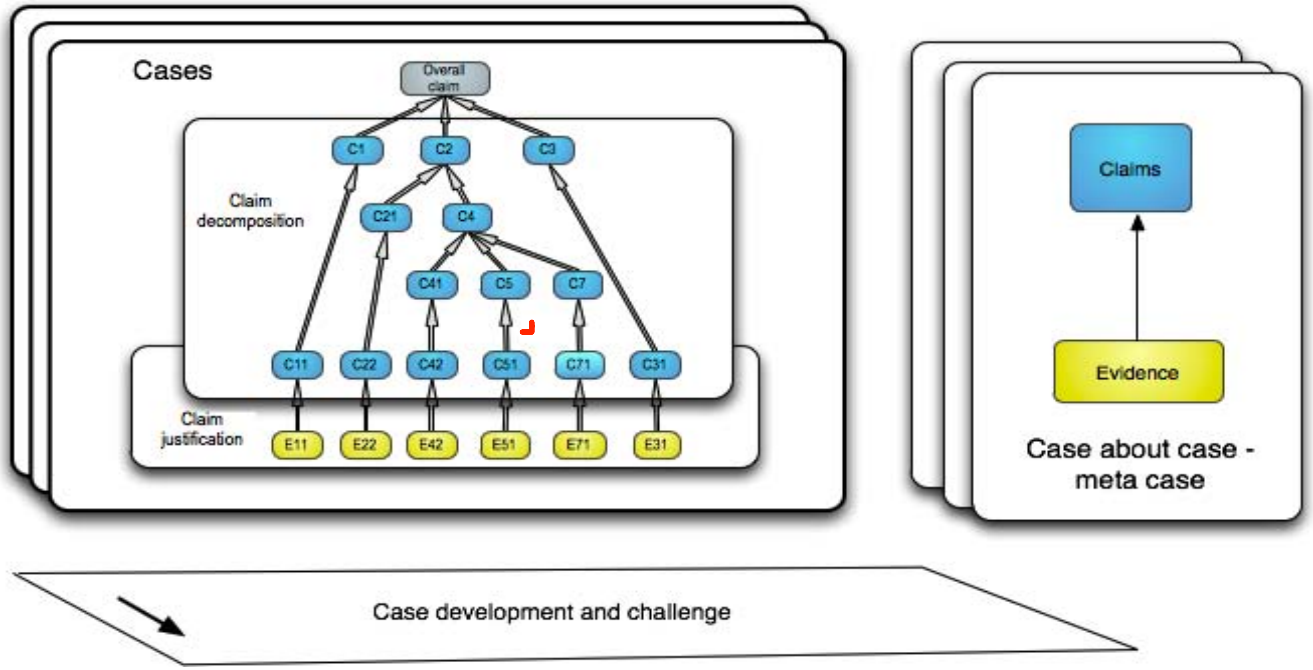
Strategies

Role of formality



Map evidence to claims

- iterative selection of techniques that generate evidence

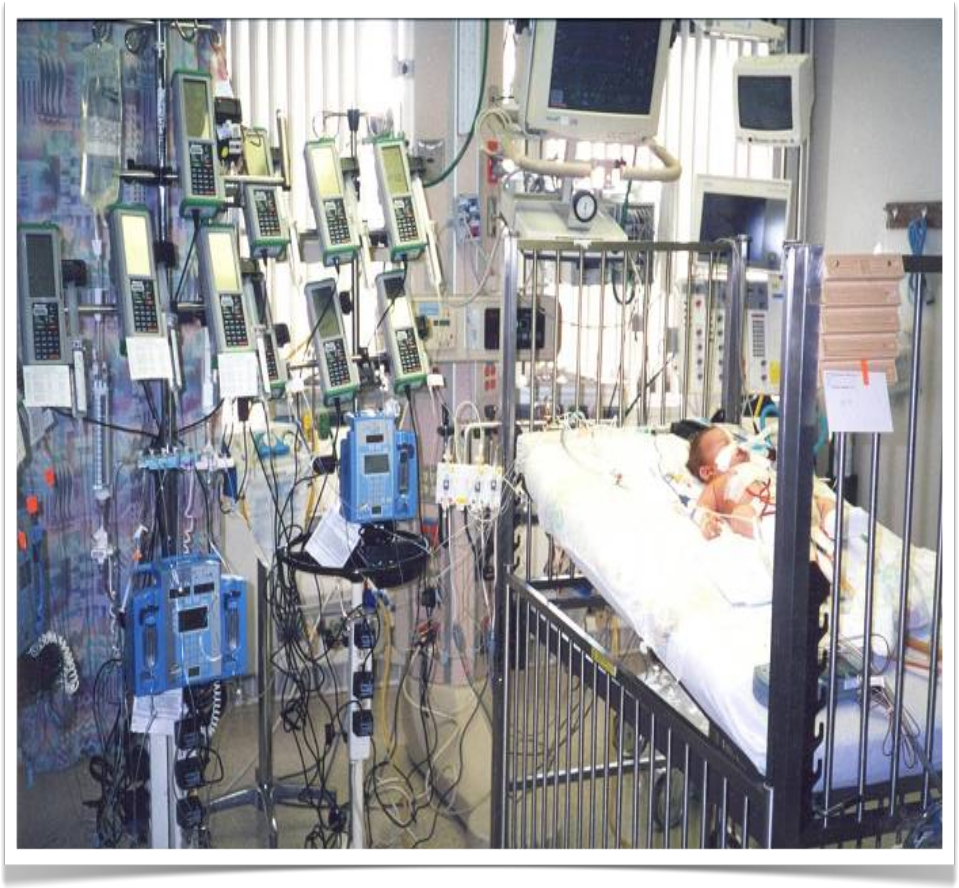


Medical device assurance cases

1. Component/ device cases should be solvable now
   - In terms of doing "what it says on the tin"
   - Safety and software engineering
   - Architecting for safety
   - Formal analysis techniques and statistical testing
   - Field experience models
   - Interfaces ….
   - .. If not too critical, always doubt
2. Surprising that manufacturers have to supply "case" about system being safe and effective
   - Distance between component and system
3. Yet *system* assurance gap

Medical systems

- Tempo
- Heterogeneous systems
- Patient's own devices
- Accidental systems
- Ad hoc Apps
- Off label
- Local and global
- Multi-stakeholder



COMPLEX ADAPTIVE SYSTEMS

Complex adaptive systems

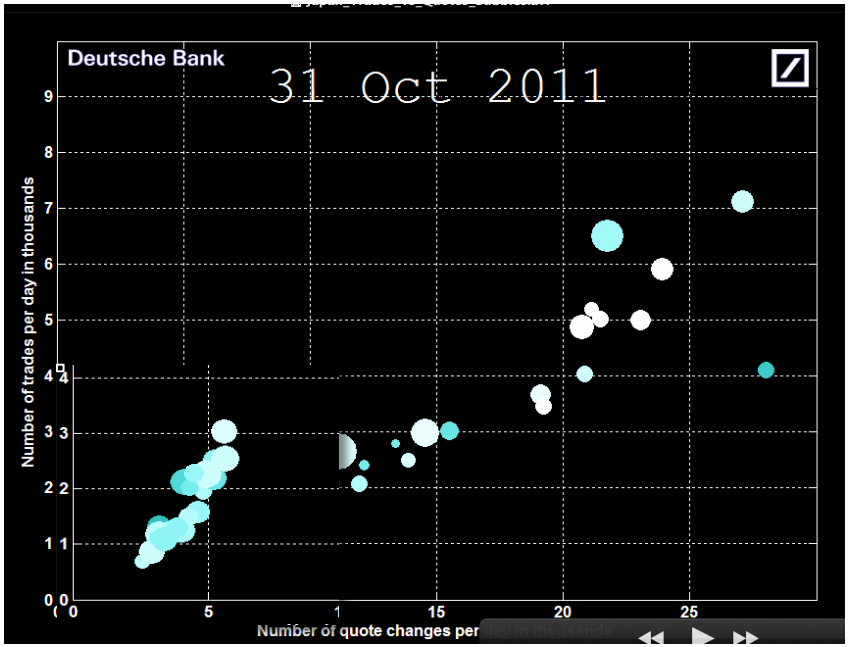
1. Complex adaptive
   - Large scale, complicated
   - System properties and emergent behaviours, tipping points, phase transitions, cascades
   - Adaptive with people and society as part of the system or part of openness
   - Limits of reductionism, potential for surprise
2. Finance sector and computer based trading
   - Computer trading and systemic risk: a nuclear perspective
   - Swans, Dragons, Cats, Toads and Bugs
   - Like medical – socio-technical-political system
   - Societal important
   - Lots of data (potentially)

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

Impact of speed



Trust in computer-based systems (and fpgas)

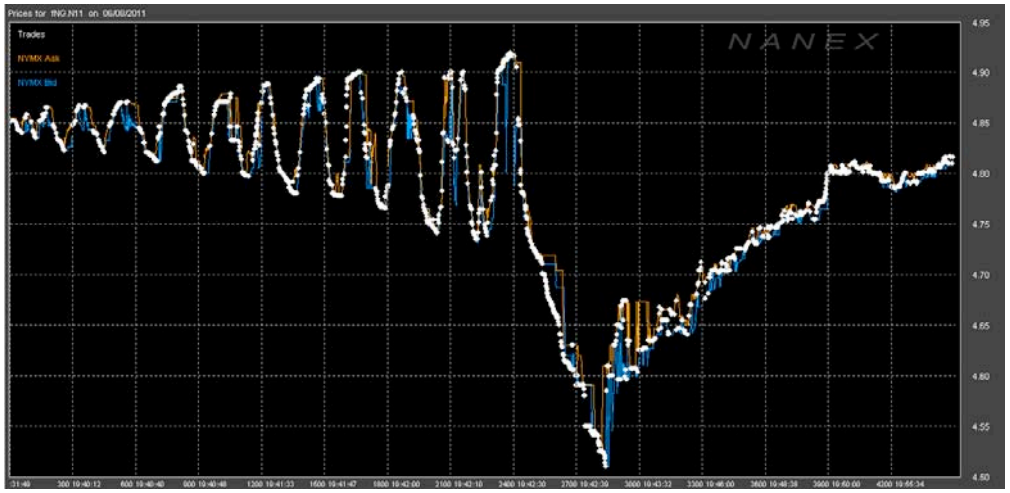
1. Recent work by Johnson
   - analysed a set of 18,520 ultrafast black swan events in stock-price movements between 2006 and 2011
   - empirical evidence for an abrupt transition from a mixed human-machine to a all-machine phase (<650ms for crashes, <950ms for spikes)
2. How much do computers need to 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

Market oscillations

According to nanex highly unusual; pattern possible caused by algorithm buy/sell confusion. Live testing?

Other possible types of issues seen

- fat fingers
- denial of service
- Interdependencies
- Cascade
- Malicious attacks (The SkyNet Wars: How A Nasdaq Algo Destroyed BATS)



Black swans - power laws and fat tails

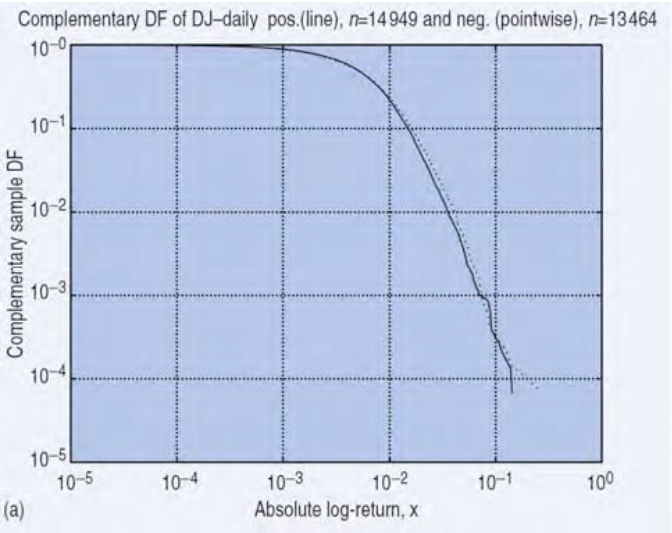
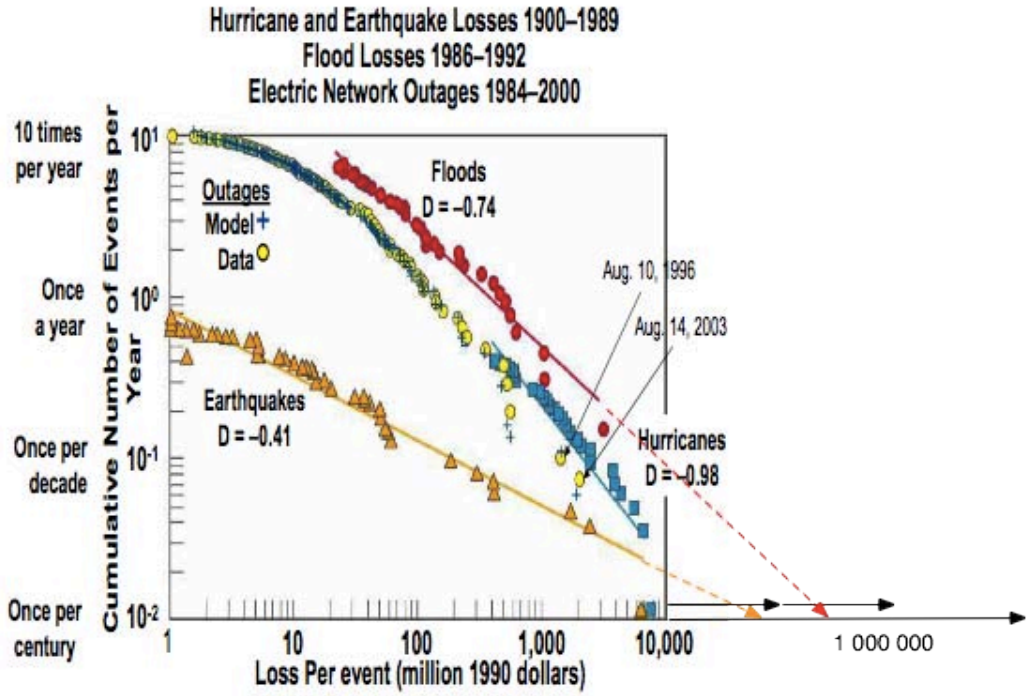
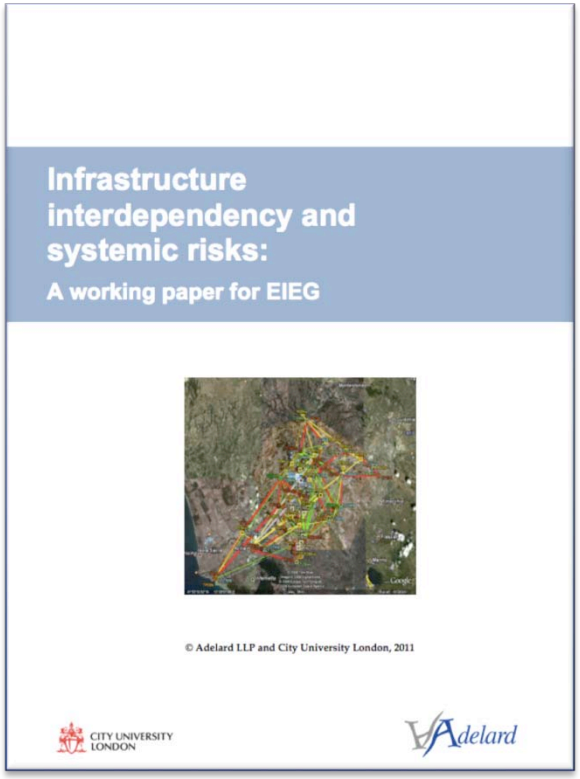


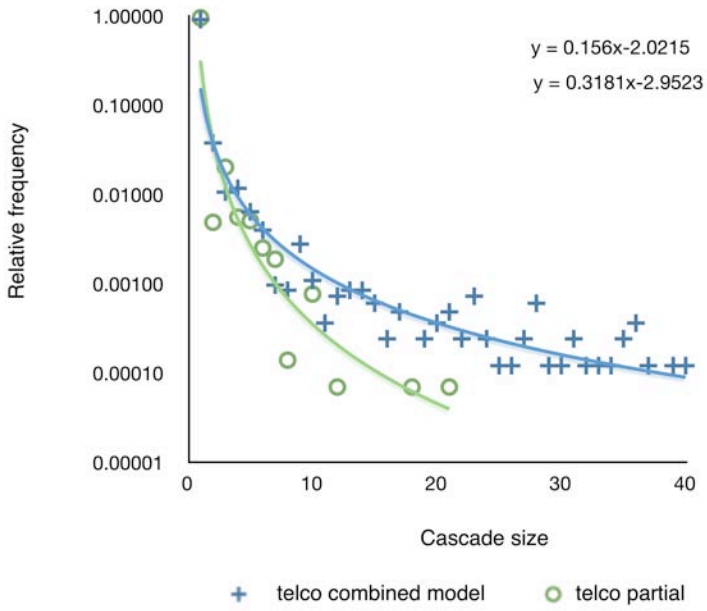
Fig.7 Survival distribution of positive (continuous line) and negative daily returns (dotted line) of the Dow Jones Industrial Average index over the time interval from May 27, 1896 to May 31, 2000, which represents a sample size of n=28 415 data points. The straight part in the tail in this log-log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



Visually power laws but see critiques

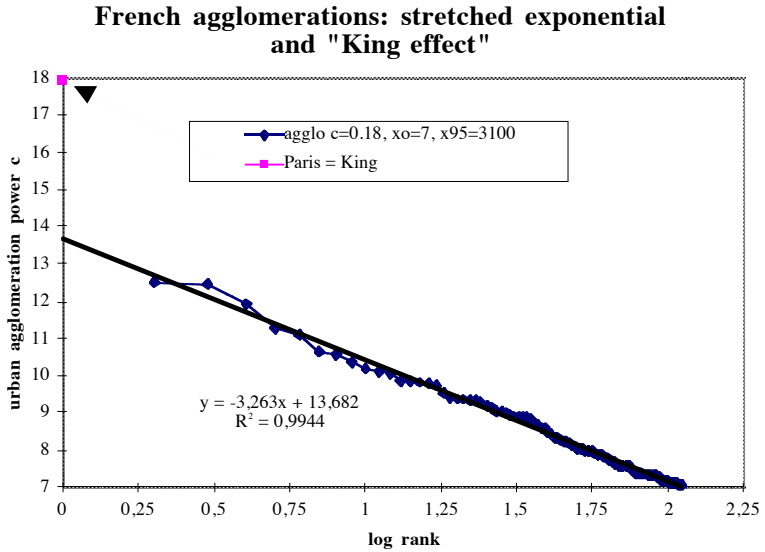
Critical infrastructure interdependencies





Dragons

1. Even with power laws and other fits get outliers



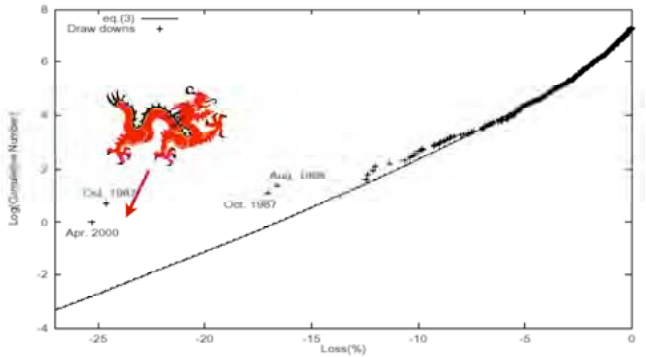


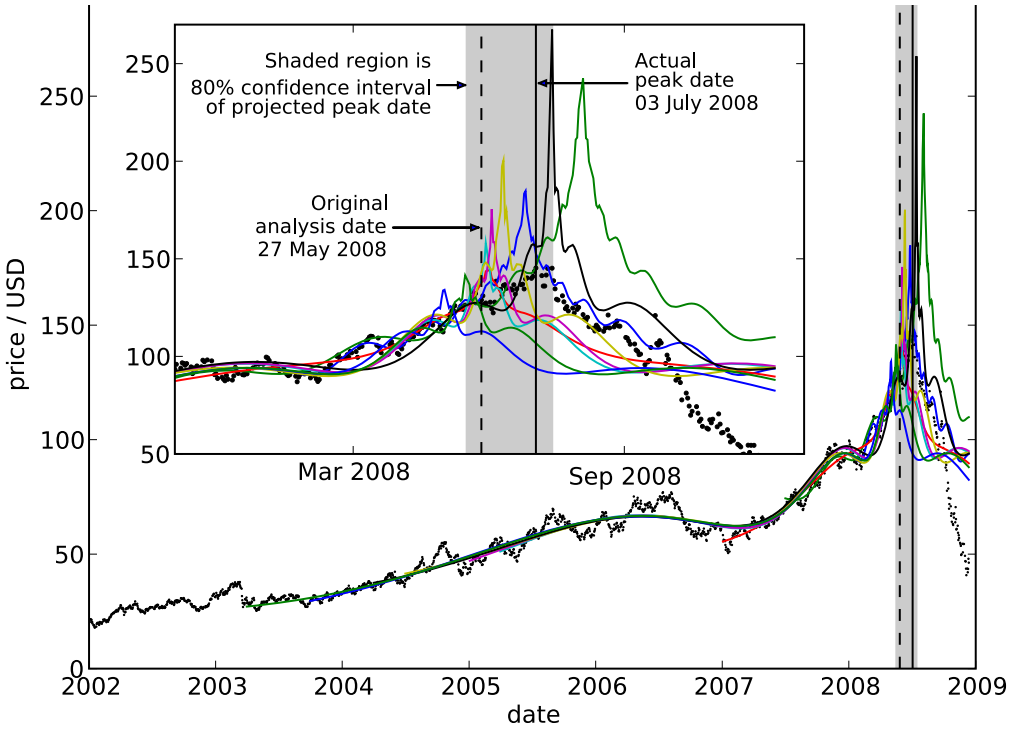
Fig.15 Distribution of drawdowns D for the Nasdaq Composite index, showing several "outliers" in the tail, that qualify as dragon-kings. It turns out that these anomalous events are

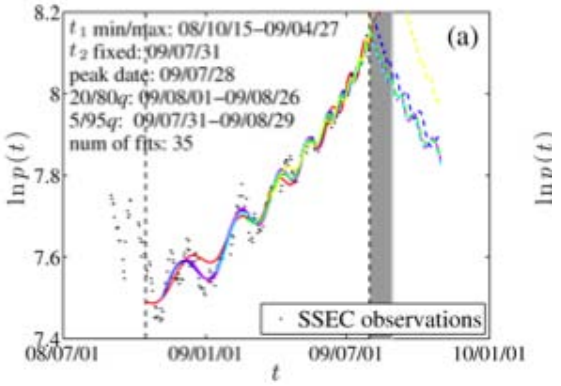
Financial crashes – no precursors?

1. Power law distributions embody the notion that extreme events are not exceptional events.
2. 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).
3. 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*
4. 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)

Tolstoy "All happy families resemble each other, each unhappy family is unhappy in its own way"

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

Toads

.

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

-

The human element

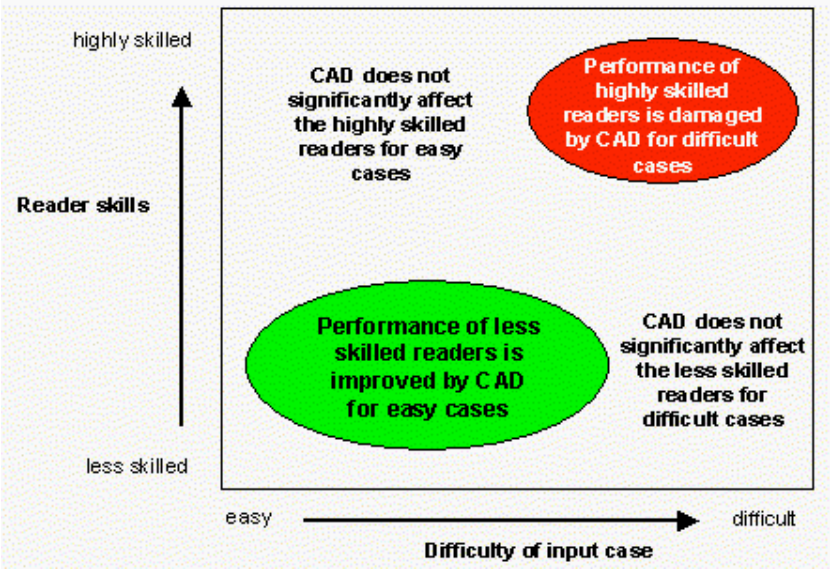
1. People as a source of resilience, as a threat and as victims
2. Don't blame the pilot – but the system
3. "fat fingers", attacks, market abuse, fraud

Socio-technical, adaptation

A socio-technical perspective:

- 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 alignment of incentives so system evolution is shaped



Performative models

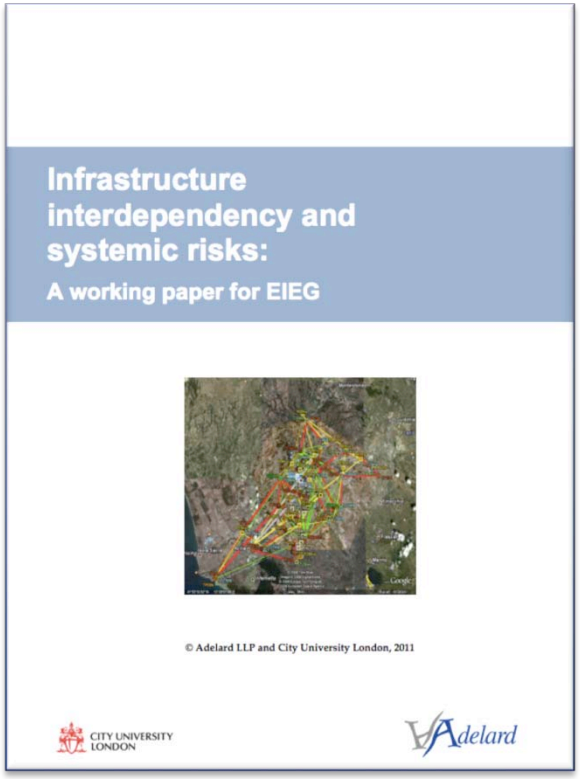
1. In the past, in the engineering domain, the models used to design and assess the risks do not affect the threats or challenges that the system faces
2. Modelling severe weather does not change the wind speed in London
   - (except perhaps via a slow political process and peoples' behaviour)
3. In the financial (and security) 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
4. engineered systems and security risks
5. knowledge and access to design models may inform an adversary and hence have a potential impact on the threats a system faces

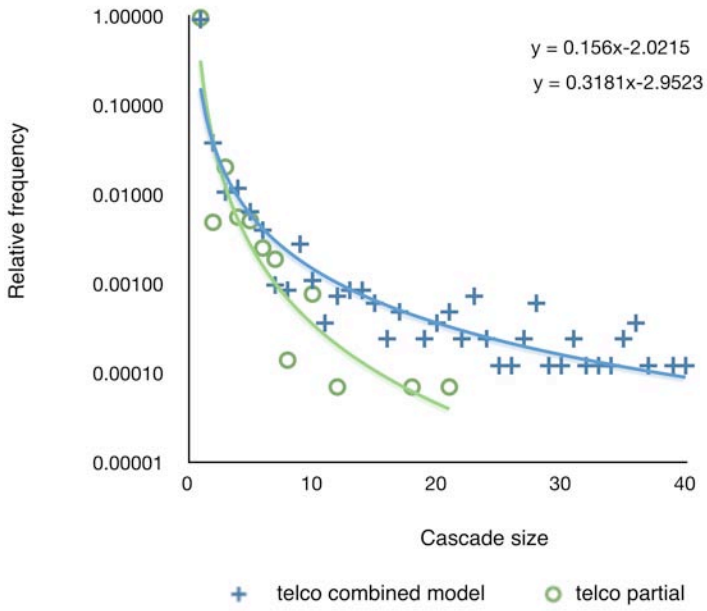
Donald MacKenzie , An Engine, Not a Camera: How Financial Models Shape Markets and also Do Economists Make Markets?: On the Performativity of Economics

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, supply chain, design basis threats*

Interdependencies and openness





Not all behaviour is “complex adaptive”

The Nimrod Review

An independent review into the broader issues surrounding the loss of the RAF Nimrod MR2 Aircraft XV230 in Afghanistan in 2006

A Failure of Leadership, Culture and Priorities

Charles Haddon-Cave QC

-

Fukushima

-

- Tsunami hazard not a surprise
  - Guidance 2002, assessment revisited in 2010,awareness among specialists of need to revise this
  - Not sure what the results of re-evaluation (if any) but considerable work on method, publication was due in 2012 Introductory, superficial talk.
- So how did a society sensitised to the initiating events, technologically advanced with world class engineering resources have this accident and what can we learn from it?
  - Key issue is dealing with uncertainty in analysis and communication

"High dwellings are the peace and harmony of our descendants,"

Vincent Yu/Associated Press A centuries-old tablet warned of tsunamis in the town of Aneyoshi, Iwate Prefecture, in northern Japan.

-

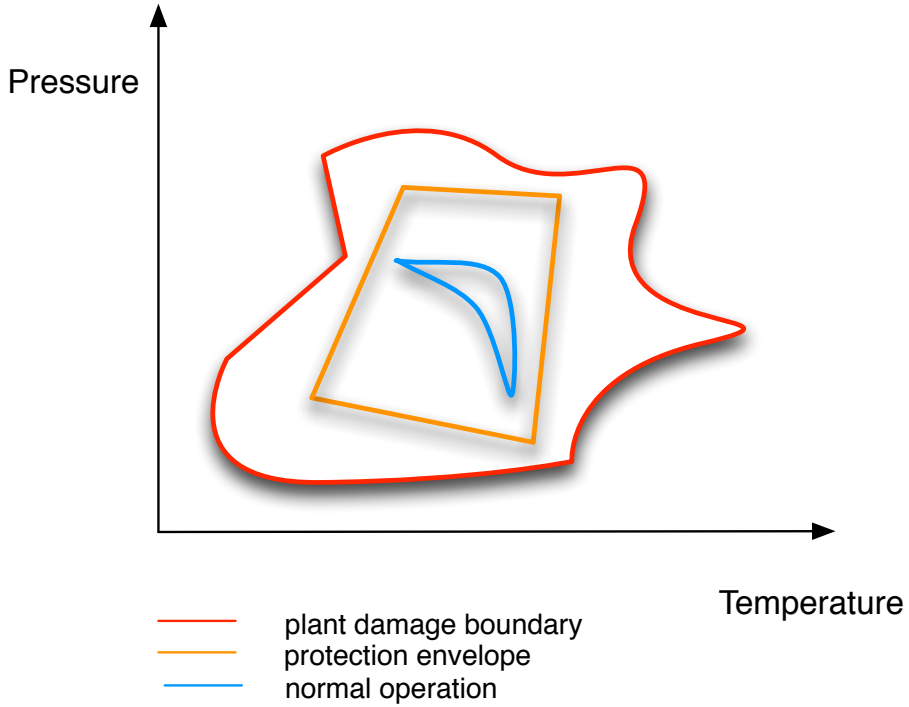
Engineering complex adaptive systems

Some basic questions

1. 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?

2. Need to address – adaptation, openness, connectedness

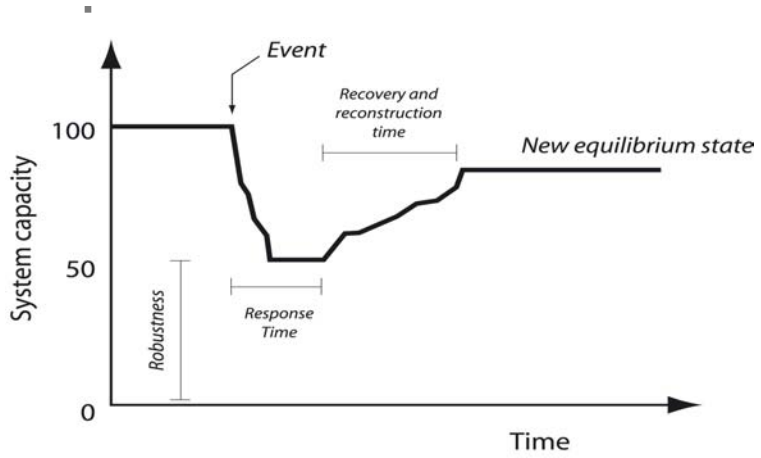
Protection parameters and viability domains



Engineering complex adaptive systems

1. Can simplified envelopes of operation be defined? Can we define control strategy for unstable system?
2. What would be the parameters that need to be measured and what are we trying to infer?
3. 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?
4. What does “safety” and resilience mean in the complex adaptive system context. Where do we draw the boundaries?
5. What is the balance between automation and operator recovery. Who would design this when no centralised authority?
6. If crashes are hard to anticipate should the focus not shift to recovery and resilience?

Concepts - resilience viewpoint- “case”



- Type 1: Resilience to design basis threats. This could be expressed in the usual terms of availability, robustness, etc. It could be bounded by credible worst case scenario.
- Type 2: Resilience to beyond design basis threats. This might be split into those known threats that are considered incredible or ignored for some reason and other threats that are unknowns.

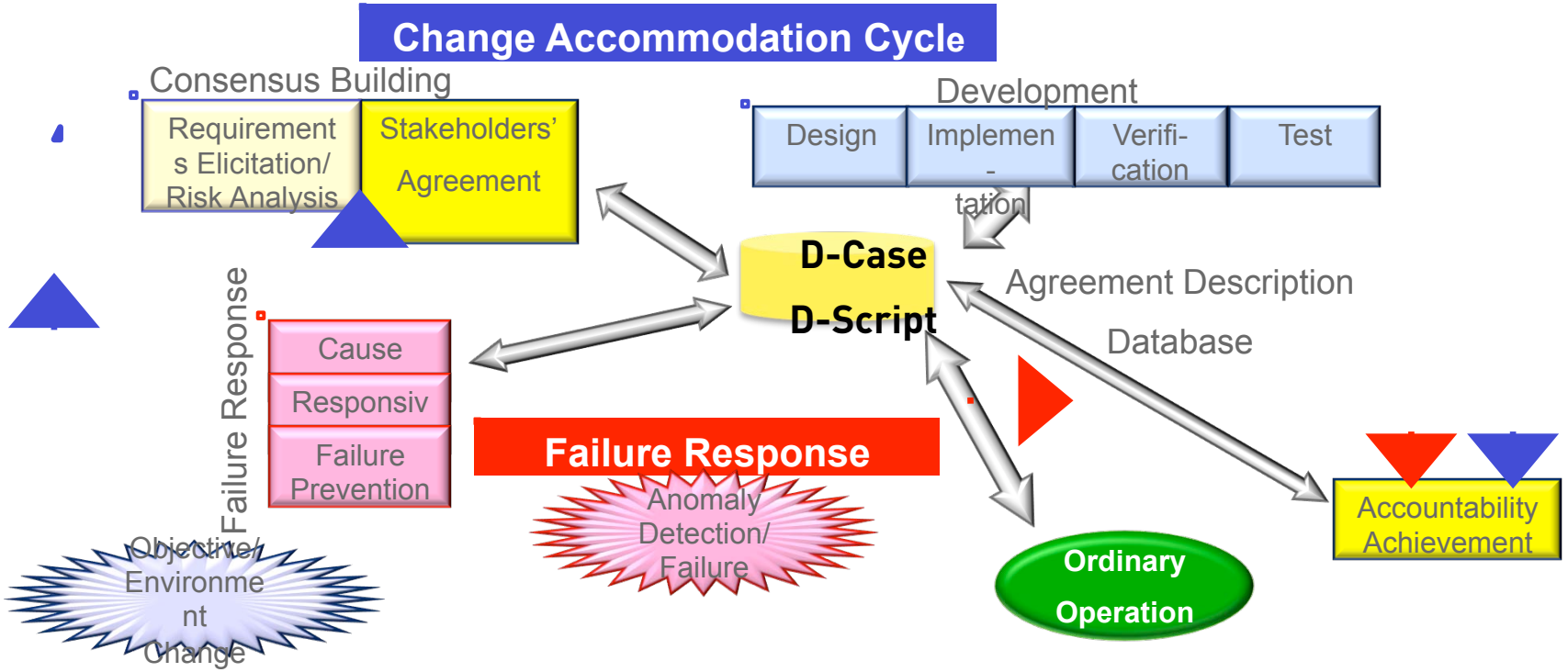
- Attacks on intangibles - these are also societal assets, not just CIP
- Does addressing Type 2 help with Type 1?

-

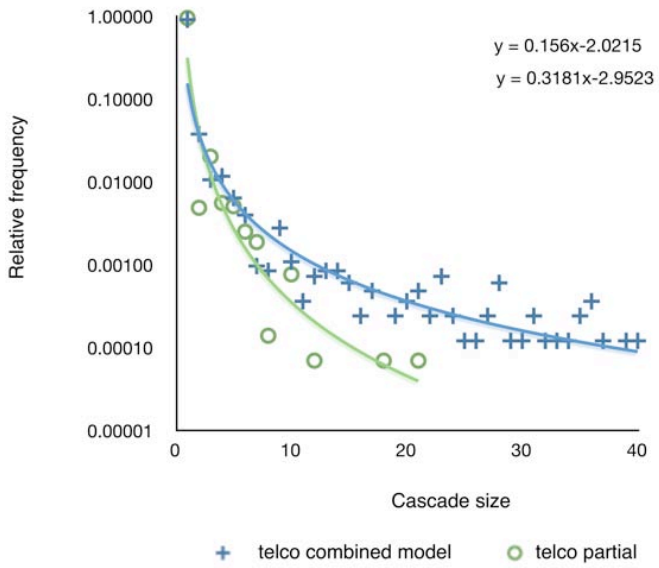
Open Systems Dependability - the DEOS Process

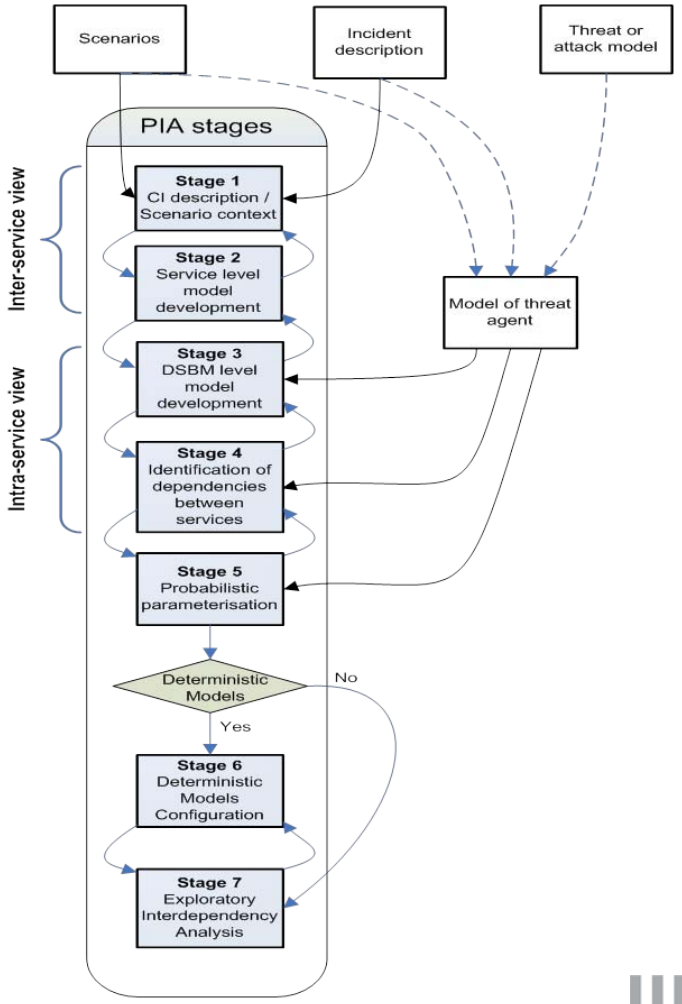
A process to achieve Open Systems Dependability

- Iterative process
  - Change Accommodation Cycle to accommodate requirement changes in service objectives and environments
  - Failure Response Cycle to respond quickly and properly to failures
- A process of processes that are organically connected

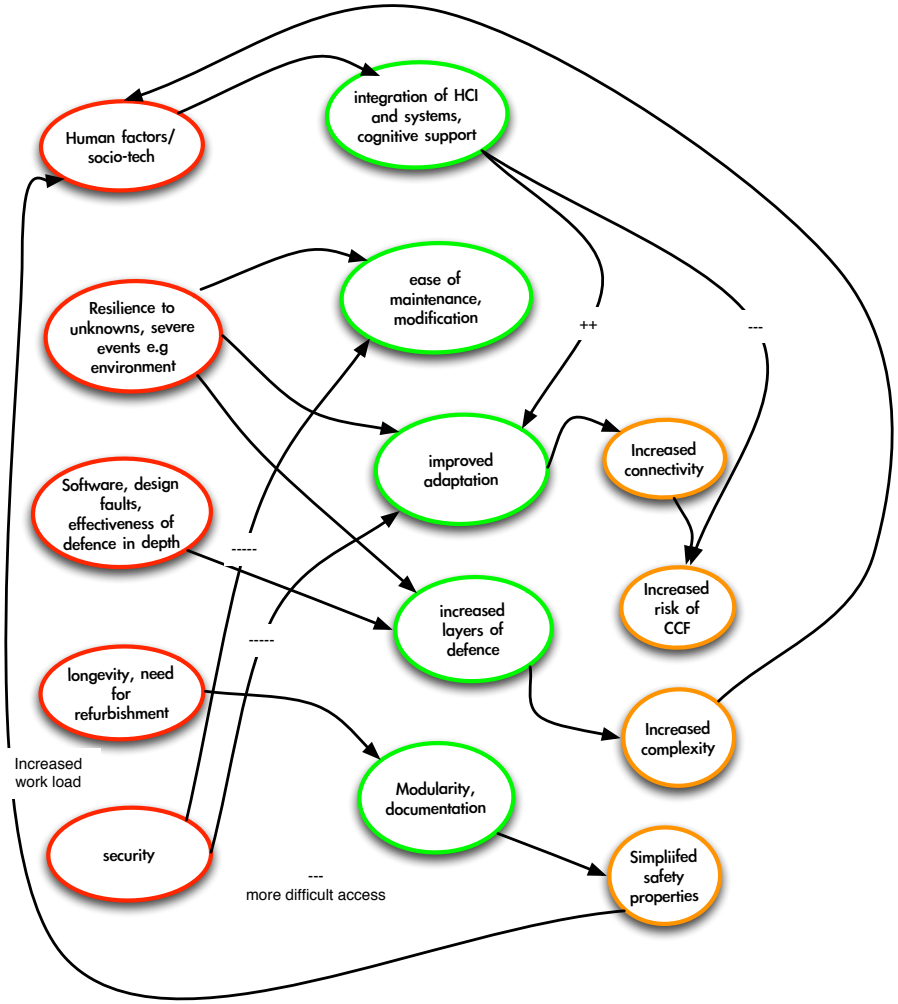


Explicit approach to interdependencies



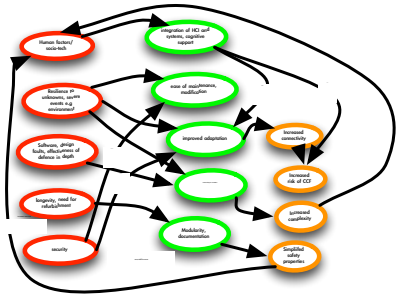


Drivers and trade-offs

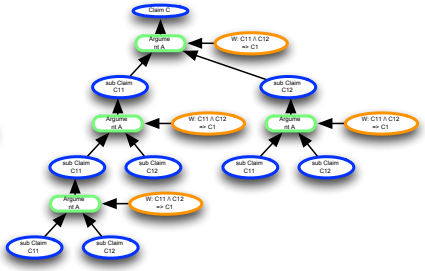


From influence diagrams to claims

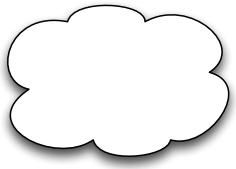
Influence diagram



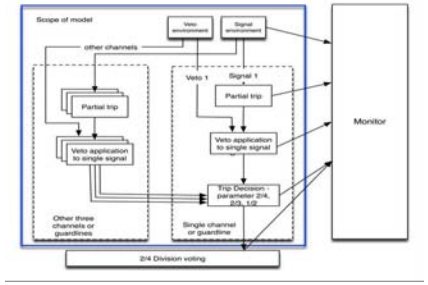
CAE structure



Mental models

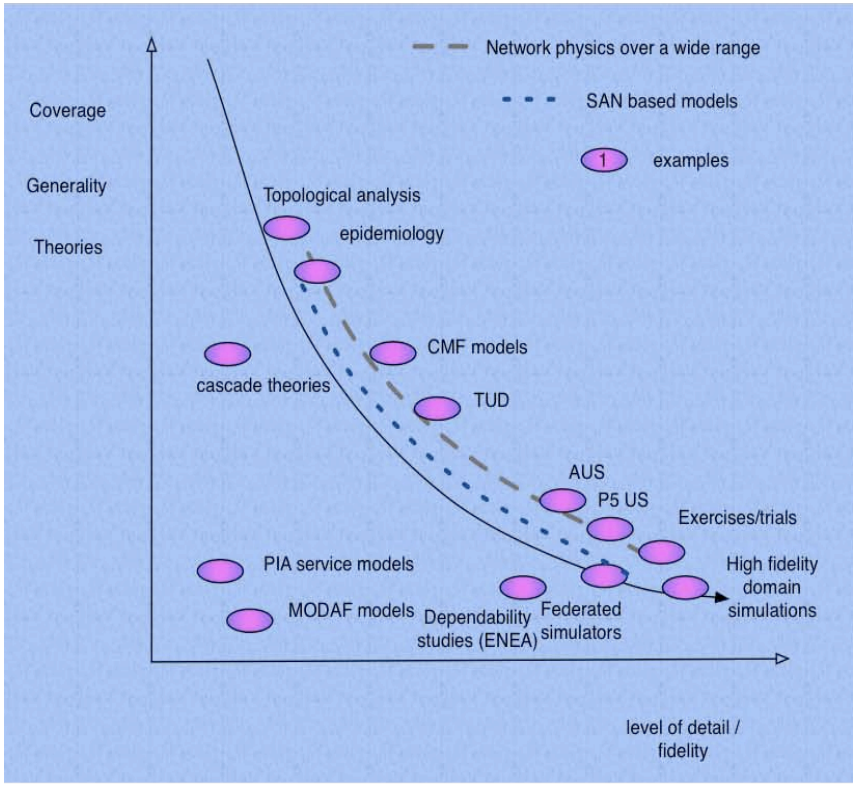


Engineeering models



Need for models and *theories*

1. A plethora of research on infrastructure interaction modelling from diverse research communities
2. Partial review and classification available from this study
3. A variety of related research areas: visualisation, simulation architecture, decision support, human factors
4. Any focused research programme is likely to be highly multi-disciplinary, multi-model.
5. Issue of what abstraction brings you and what is misses; research methodology questions e.g. US vs Eu?



Engineering complex adaptive systems

Developing intervention strategies for different time bands, e.g.

- Circuit breakers and viability domain protection
- Forced diversity in ecology, disrupt correlation
- Alignment of economic and other incentives

New approaches to systemic risk assessment

- Investigate viability envelopes and recovery
  - Not just hazard identification and mitigation
- Consider both conventional and complex risks (Fukushima vs flash crash?)
- Type I and 2 resilience – explicit design basis, threats
- Different approaches for different parts of risk curves
- Emphasis on stability in the face of change as a dependability attribute

Strategies

Some challenges and things to draw on

In engineering trusted systems

1. Defence in depth and diversity
2. Interdependencies and resilience
3. Adaptation and socio-technical systems
4. Confidence and doubt
5. Learning from experience
6. Security and threats

Conclusions - challenges and opportunities

Work in progress

1. Finance sector and computer based trading just one example of complex adaptive systems (automotive, air traffic, medical…)
2. 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
3. Opportunities from deploying and adapting our current engineering approaches to systemic risks
4. Work in progress….
   - Look forward to your ideas

Adelard

- 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

Centre for Software Reliability

- Evaluation of socio-technical systems
  - Technical, interdisciplinary
- Research
  - with international community and users
- Education
  - placements, internships, scholarships, courses, MSc and CPD
- Innovation