UNIVERSITY of York

ASSURING AUTONOMY INTERNATIONAL PROGRAMME

Safety, Security and Autonomy: plus ça change plus c'est même chose?

Prof John McDermid OBE FREng

Overview

Safety, Security and Autonomy

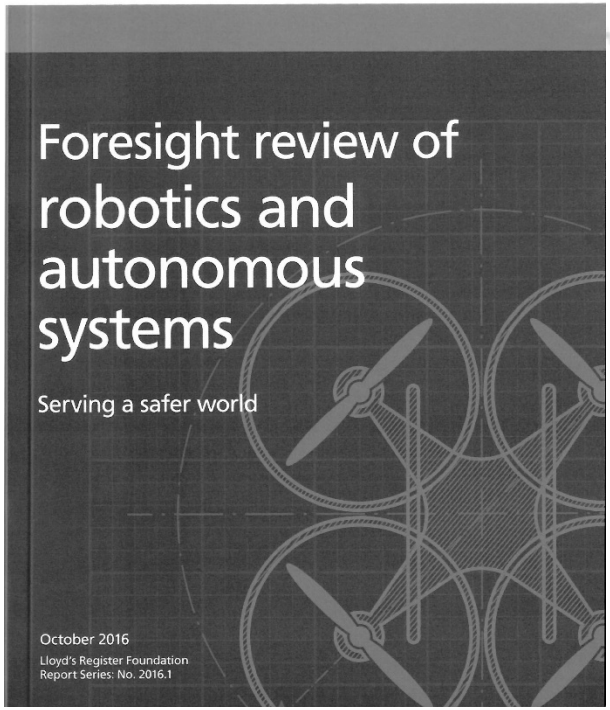
- Assuring Autonomy International Programme
- Safety and Security
  - Concept
  - Analysis approaches – a biased review
- Safety, Security and Autonomy
  - Additional challenges – a tentative view
- Conclusions

Robotics & Autonomy

The Foundation's Review of RAS

- Published in October 2016
  - Key "white spaces" in assurance & regulation that need addressing to realise benefits of RAS
- Five-year York-led Programme
  - £10M from Foundation
  - £2M from York for management
  - A strong focus on 'demonstrators' and working 'bottom up'
  - Linked work, e.g. EU network

Lloyd's Register Foundation



LRF Review of RAS

Priority Research Areas

Suggested priority areas			
Openess and sharing	Assurance and certification	Security and resilience	Public trust, understanding and skills
Open data standards	Asset self certification	Cyber security of RAS	Ethical and trusted RAS frameworks
Open data sets	Assurance of RAS learning systems	Software system integrity	Assured skills for RAS
Shared curation of RAS knowledge			

Programme Activities

Scope

- Four main strands of work
  - Work on assurance and regulation in support of demonstrator projects (real-world prototypes or real deployments – use cases)
  - More fundamental research, e.g. on dynamic risk, and assurance of artificial intelligence/deep learning
  - Education and training, for professionals in RAS/AI and safety (senior level briefings to Master's level material)
  - Support to the international community

  All contributing to or using a Body of Knowledge (BoK)

Demonstrators

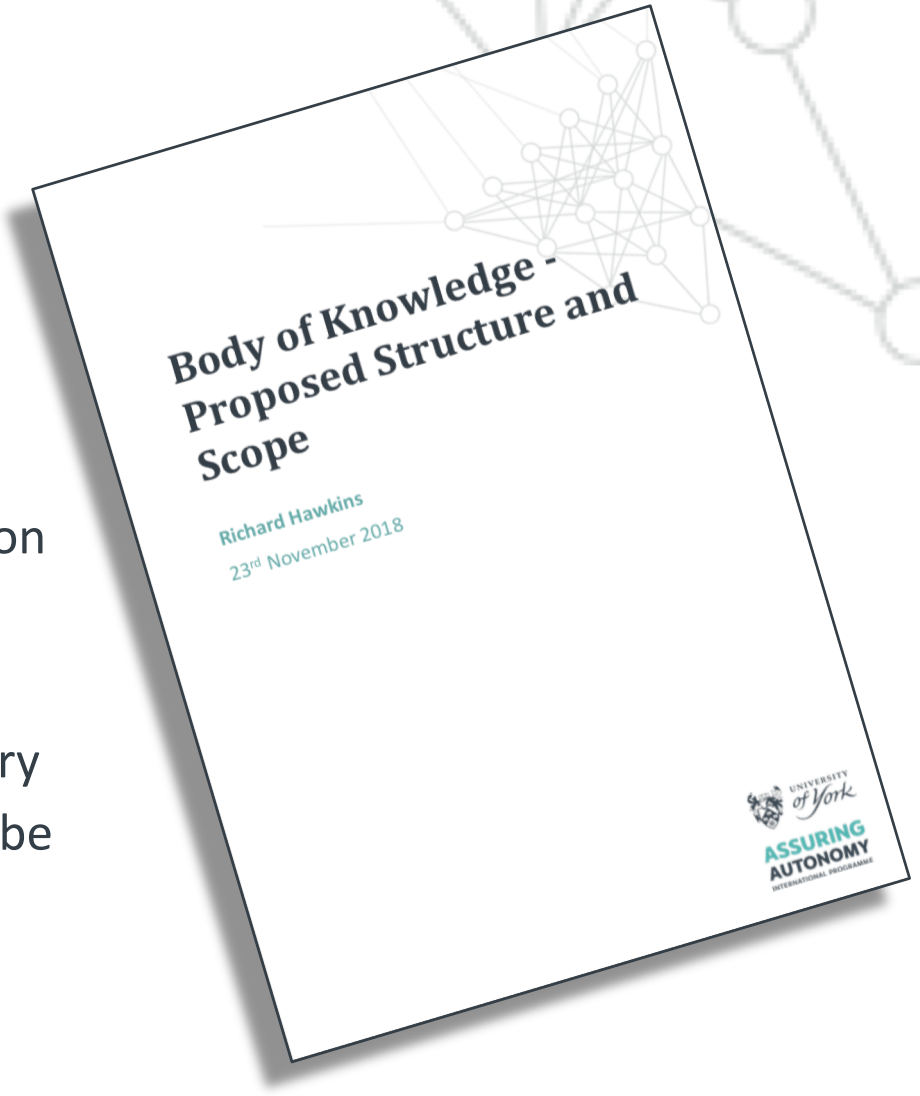
Ongoing Demonstrators

- RECOLL (MCM, Italy, manufacturing)
  - Started 01/07/18
- TIGARS (Adelard, UK & Japan, autonomous vehicles)
  - Started 01/09/18
- SAM (Derby ICU, UK, healthcare)
  - Started 01/09/18
- Assistive Robotics (Bristol Robotics, UK, assistive)
  - Started 01/12/18
- SUCCESS (Mälardalen, Sweden, quarrying)
  - Started 01/12/18

Research Consolidation

Body of Knowledge

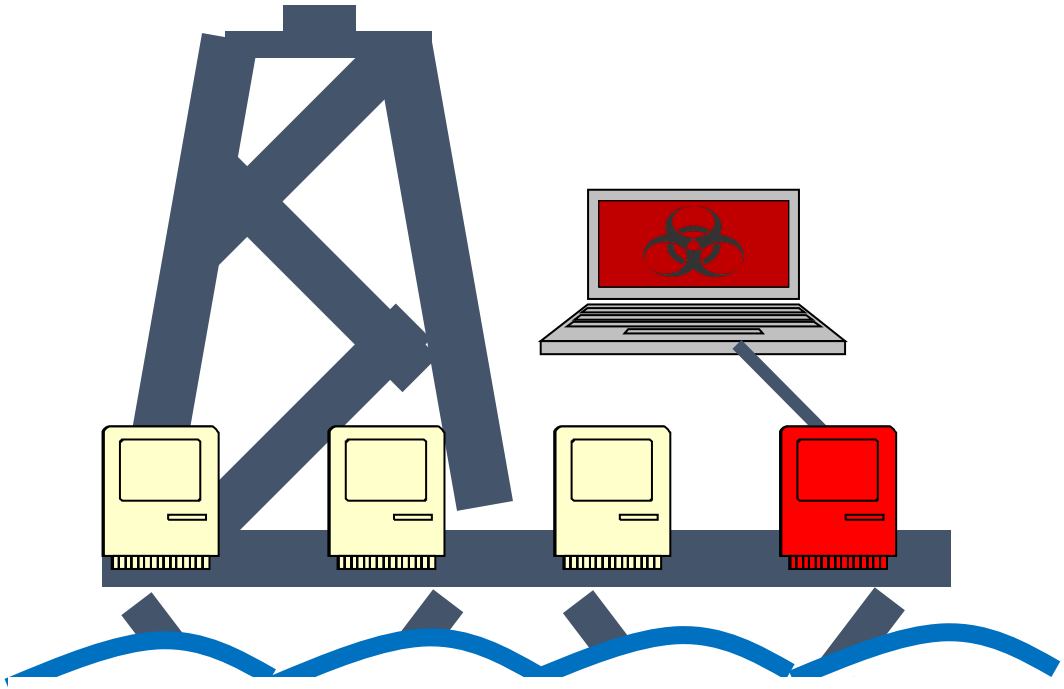
- Development of the BoK
  - Structured to address assurance and regulation challenges – for each
    - Objectives
    - Approaches to demonstration
    - Contextual information
  - Initial web-based version
    - Partially populated in January
    - More interactive version to be developed later in 2019



Safety and Security

Mobile Drilling Platform

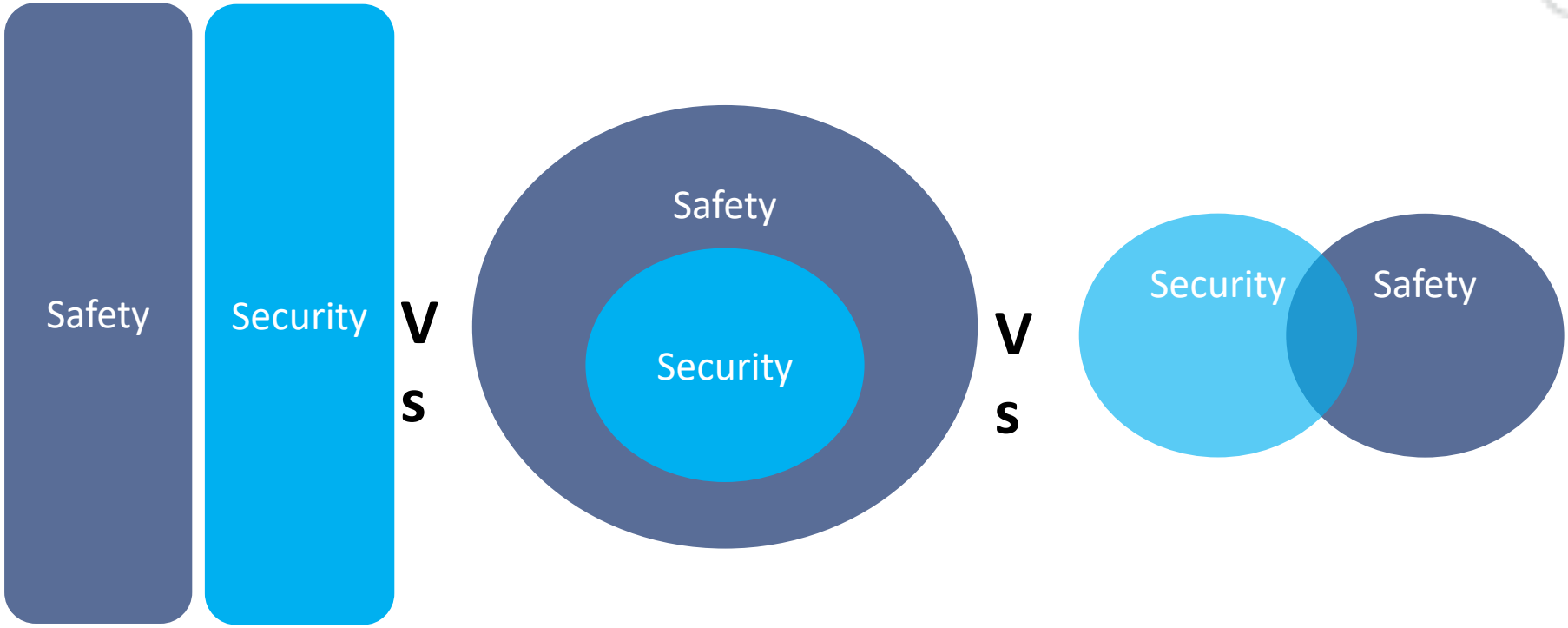
- Impact from financial malware
  - Safety problem as a result of DoS



Safety and Security

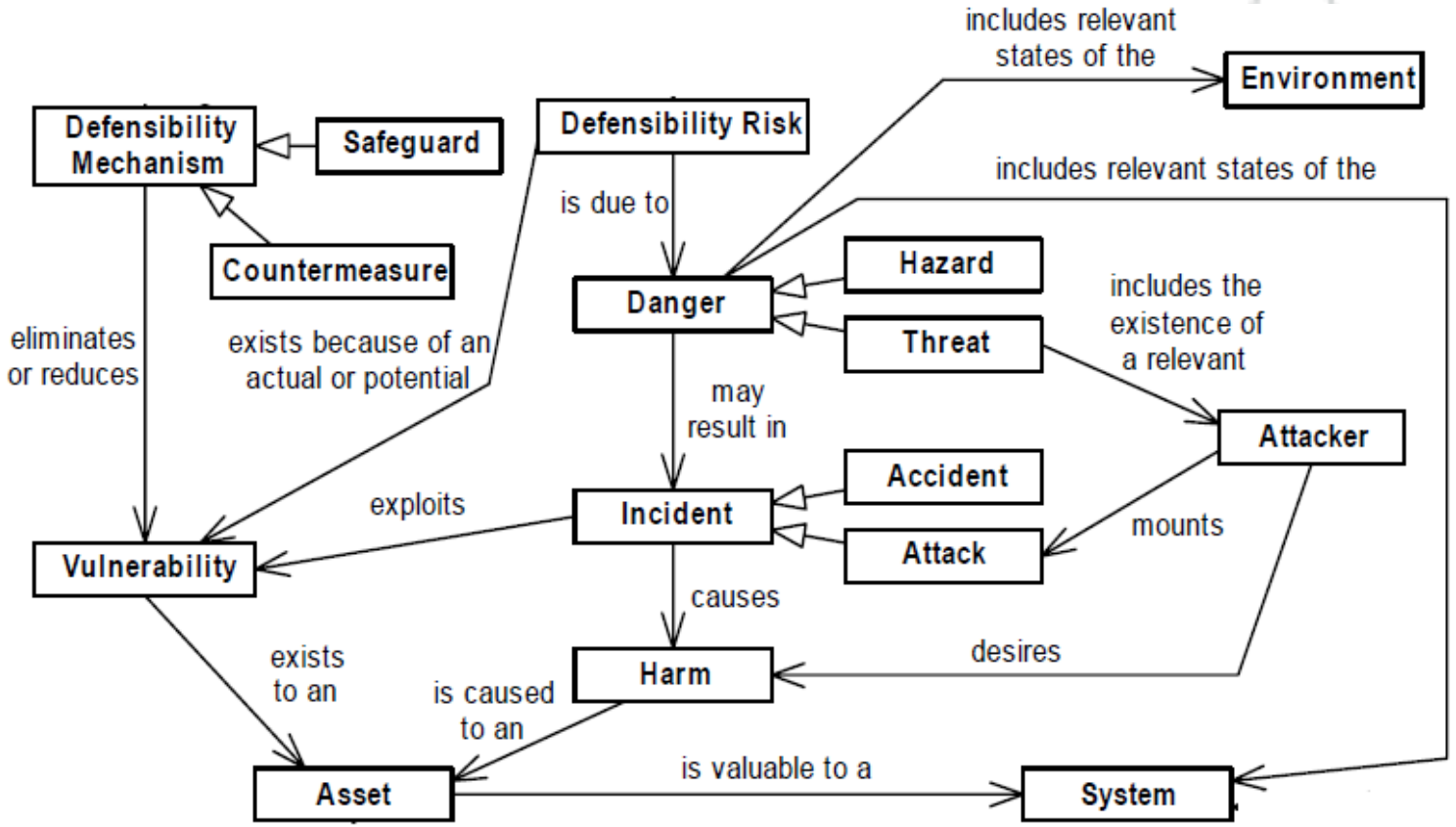
Relationships?

- Cyber attacks can cause safety problems
  - But relationship much wider



Analysis Models

Firesmith's Approach (2003)



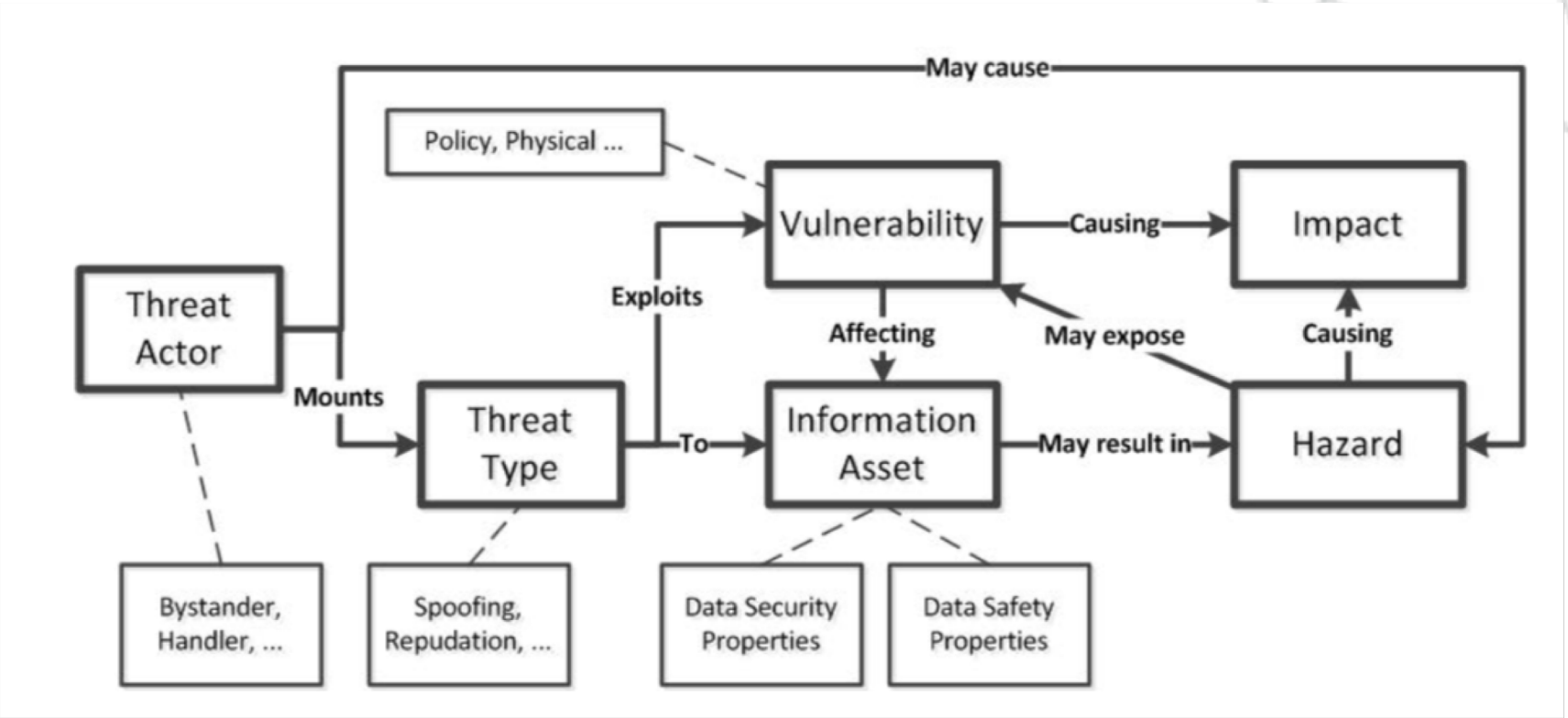
Analysis Approaches

Based on Established Safety Methods

- HAZOP-based approaches (early 2000s)
  - Chemical plant HAZOP adapted to software, e.g. SHARD
    - Guideword based flow analysis (deviations)
    - Extensions to include cyber-security causes of deviations
- STPA-SafeSec (2017)
  - Leveson extended her System Theoretic Process Analysis (STPA) to include cyber-security (STPA-sec)
  - Later work extended further, addressing perceived weaknesses in the approach
    - Essentially integrating safety and security

Early Life Cycle Methods

Cyber Risk Assessment Framework



Safety and Security

Relationships?

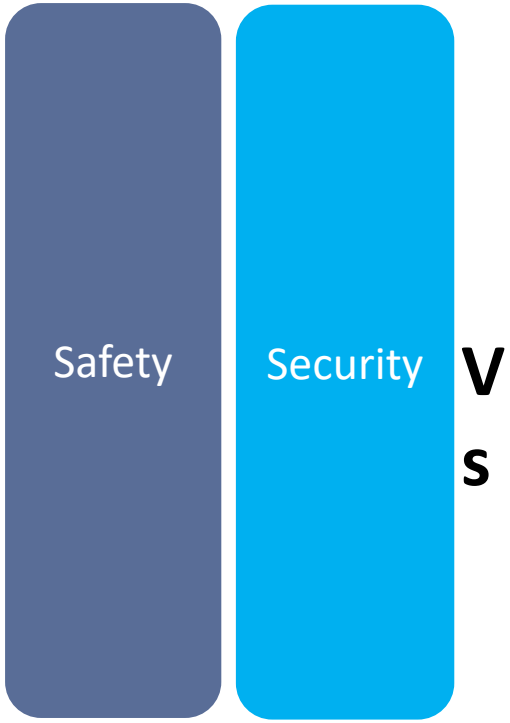
STPA/STPA-sec

SafSec

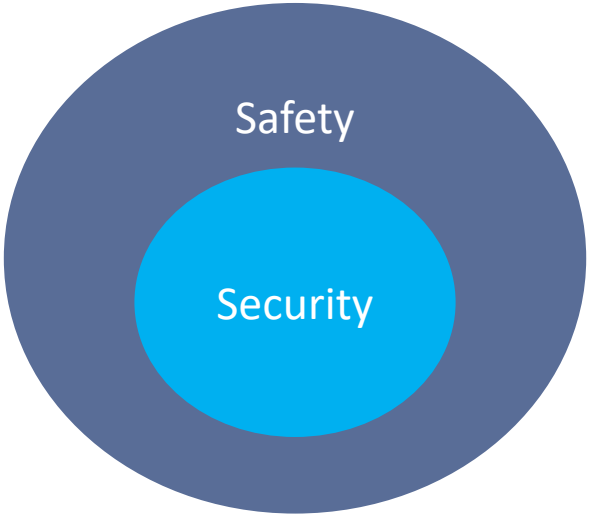
STPA-SafeSec

CRAF

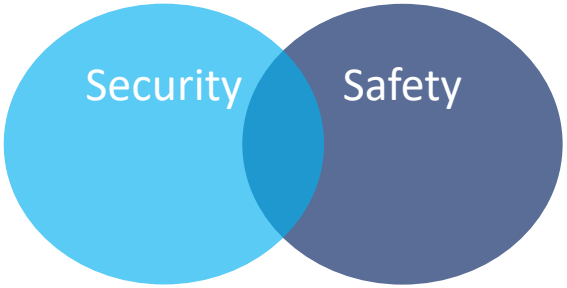
Others?



V
s



V
s



Autonomy

All the same?

- Autonomy
  - The ability of a person to make his or her own decisions (or self-government, independence …)
  - Autonomous systems make the decisions, not the humans (but judgmental – robot vs kettle)
- Autonomy does not change concepts of safety and security
  - Hazard, threat, vulnerability … all the same
  - So methods such as CRAF, STPA-SafeSec can be applied

So plus c'est meme chose?

Autonomy

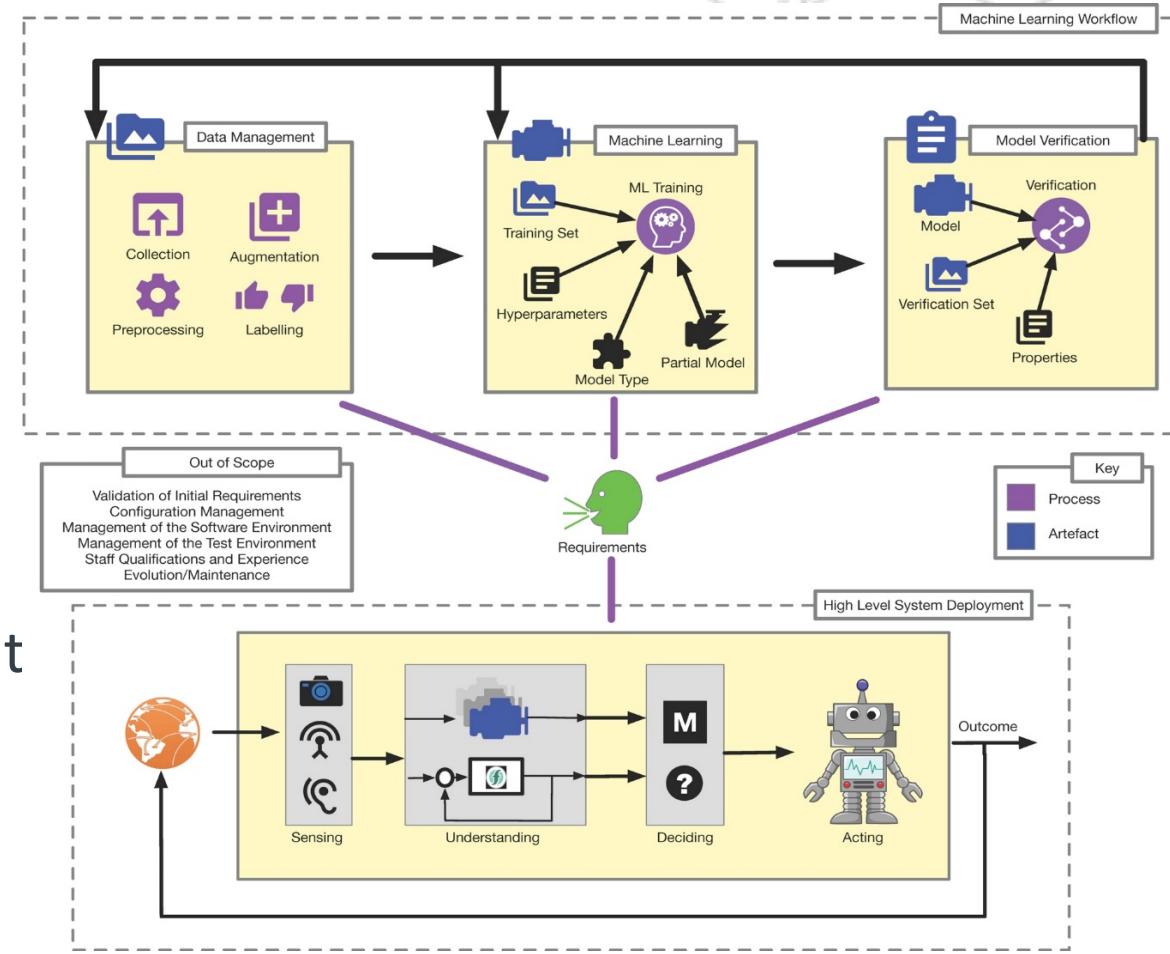
Challenges

- Classical safety and security builds in defences or barriers
  - To (detect and) mitigate risks
  - Often require redundancy or diversity for the defences
    - Other sources of, and means of processing, data
- Autonomy may reduce diversity/redundancy
  - Can we learn image analysis two different ways
    - Similar enough we can “match” objects but different enough there is a level of resilience?
  - May be more single points of failure

RAS Models

Model Underpinning BoK

- Model of systems
  - Sense
  - Understand
  - Decide
  - Act
- Development
  - Data management
  - Machine learning
  - Verification

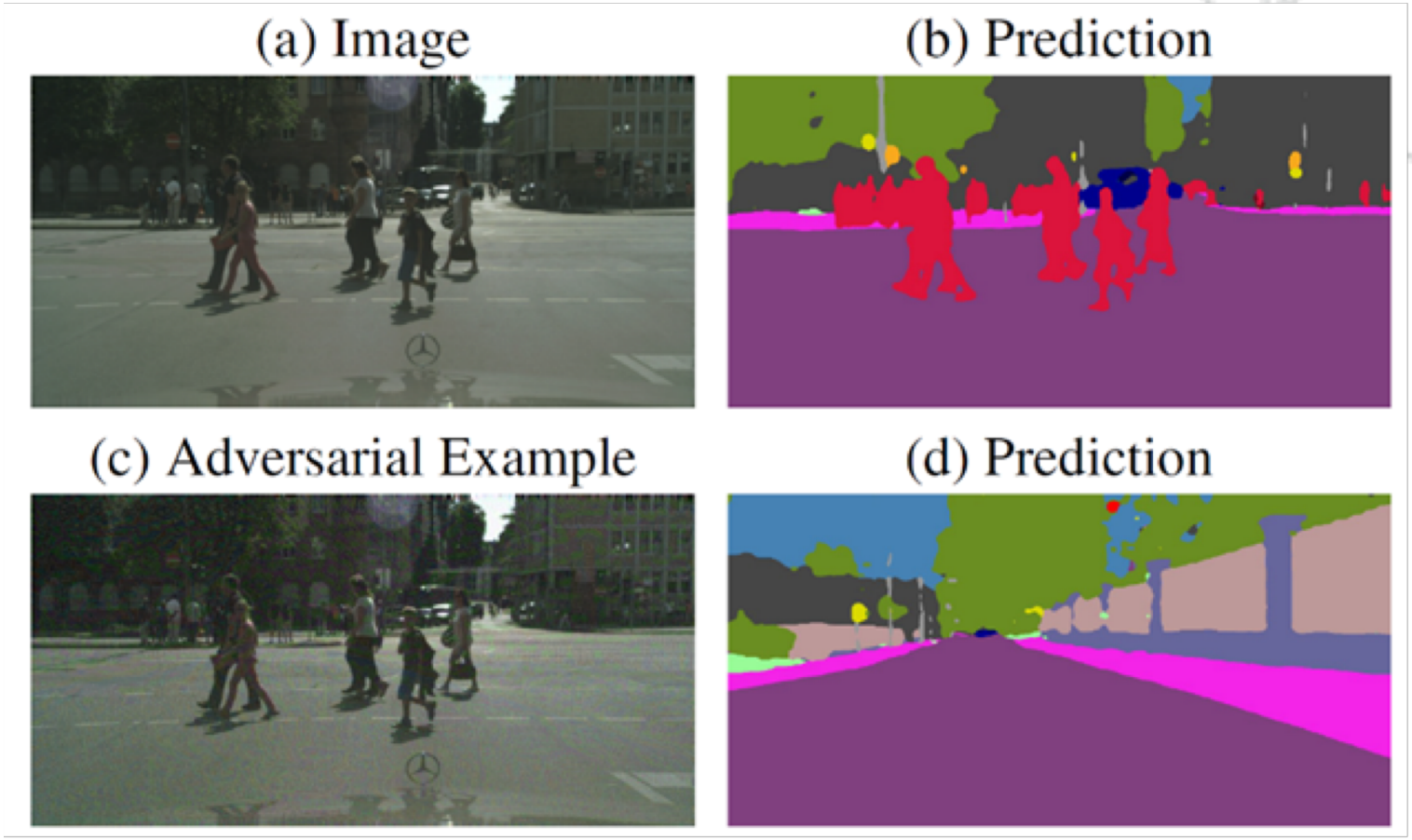


System Sensing

Spoofing (non-malicious)

System Sensing

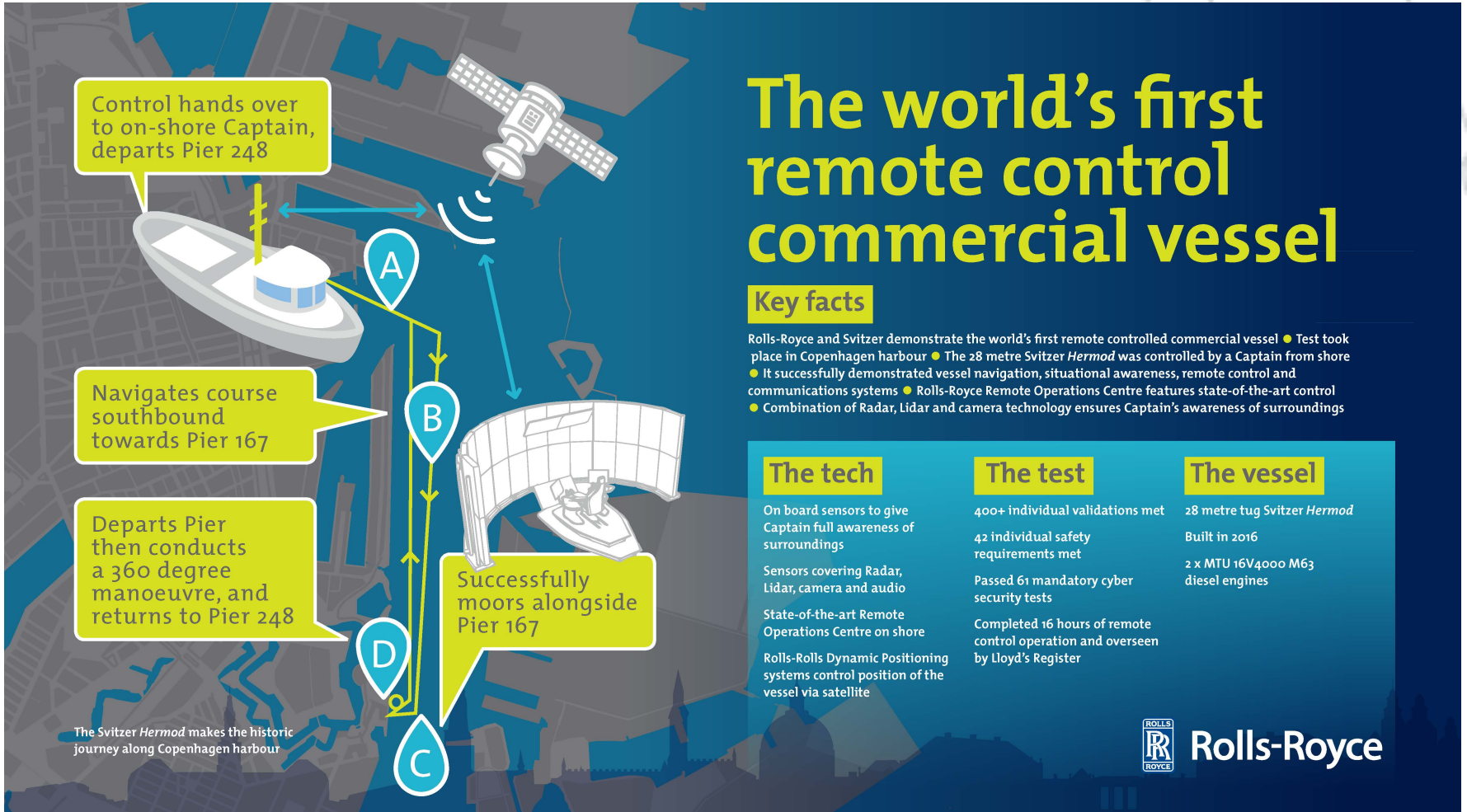
Adversarial Attacks (plausible?)



Adversarial examples for image recognition with CNNs

System of Systems

Complex & varying attack surfaces



Data Management

Choosing Data Sets

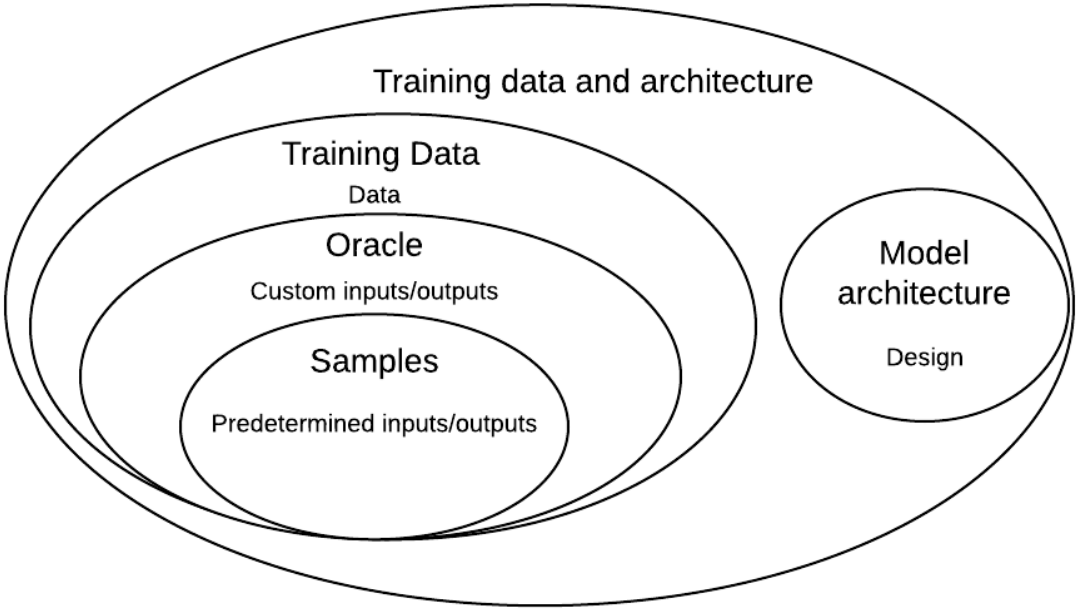
- Main training data
  - Collected from real-world observation
  - Can limitations in training data produce vulnerabilities?
- Augmentation data
  - To "complete" collected data, e.g. accident scenarios
  - Can augmentation data be chosen to trigger unsafe or insecure behaviour?
- Consider non-standard system development process, not covered by existing standards …

So plus ça change?

System Development

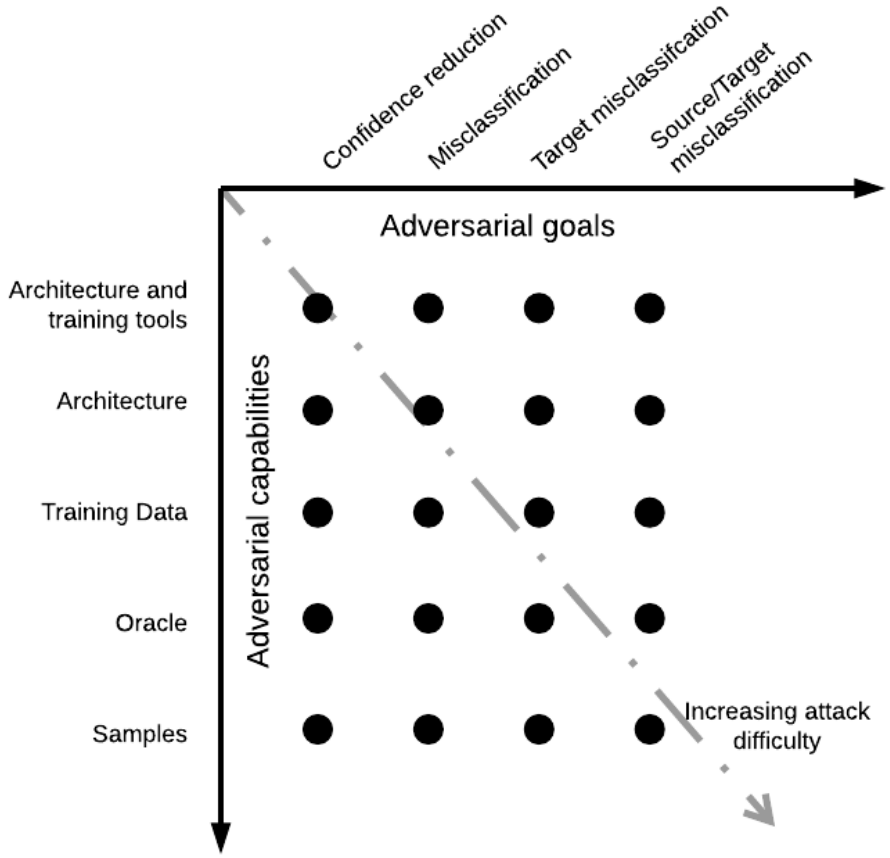
Choosing Data Sets

- Data types and roles in machine learning
  - All potential targets for adversarial activity
  - Potential direct security and indirect safety impacts



System Development

Threat Categories and Data Types



Forensics

How analyse incidents and accidents?

- With a system of systems
  - What is the scope of analysis?
  - How manage data collecting under dynamic change?
    - Cars exchanging data, learning from others …
- With machine learning
  - What information needs to be stored?
  - How can decisions be explained (NB GDPR)?
- How ensure independent of developer?
  - Cf Tesla and Uber fatalities

NB PAS 1885 mentions incident response, but doesn't seem to reflect autonomy …

Conclusions

Safety, Security and Autonomy

- Principle 0 for assuring autonomous systems
  - Apply standard good practice – safety, security, etc.
- Initial bias – that was all we need to do
  - But we do need an integrated approach such as CRAF
- Growing belief
  - Safety and security for autonomy are different
    - Work is needed on product, process and forensic issues
  - Another research strand to augment the Assuring Autonomy International Programme?
    - Collaboratively with Southampton, NCSC, … ?

ASSURING AUTONOMY

INTERNATIONAL PROGRAMME

Funded by

Lloyd's Register Foundation