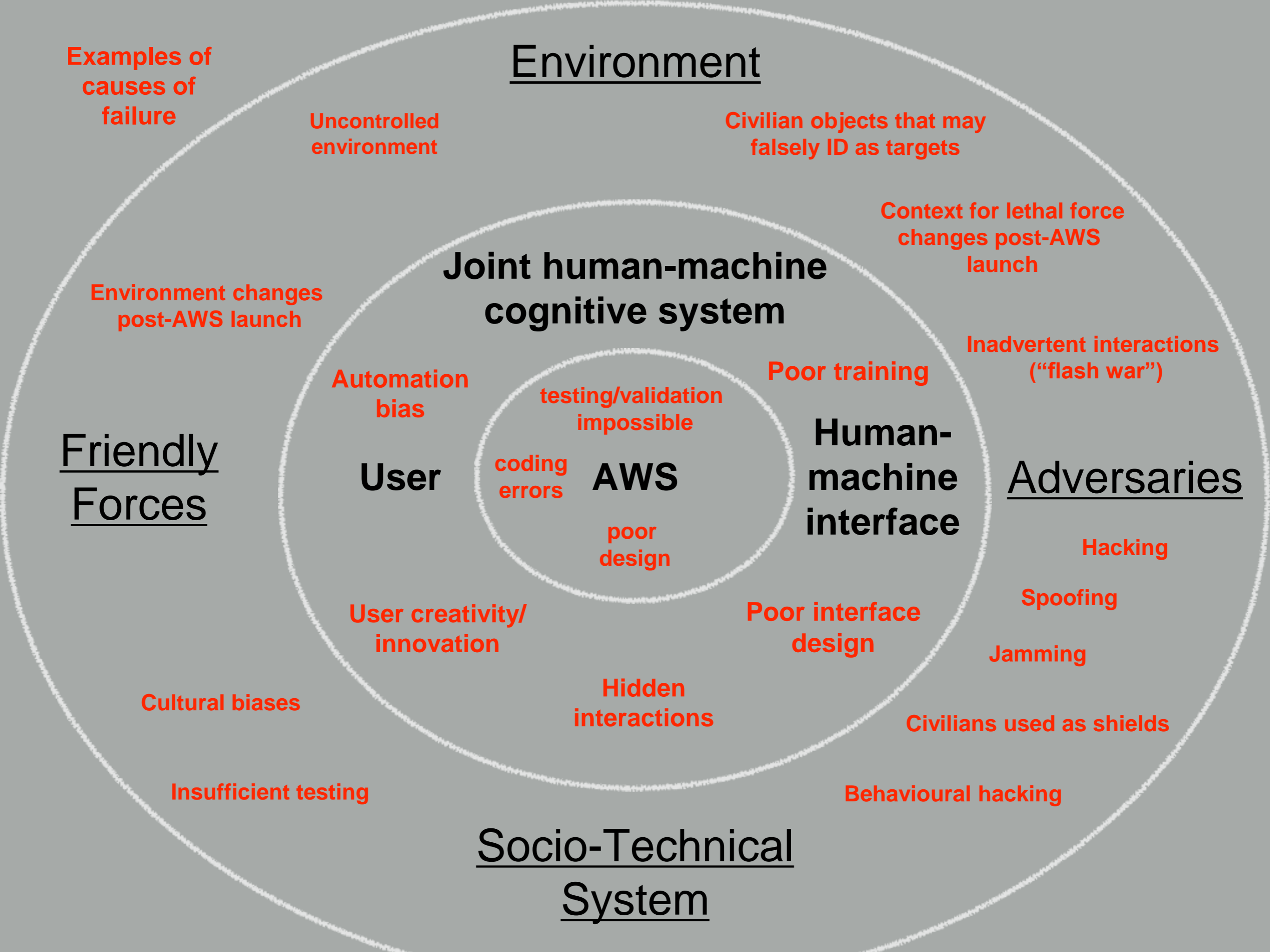


# **Security, Unintentional Risk, and System Accidents**

**John Borrie, UNIDIR**



- **Risk:** probability x consequences
- **Unintentional risk:** sub-set of total risk in which machine systems that have targeting and attack functions fail to behave in ways intended by their designers and operators.



# Environment

Examples of causes of failure

Uncontrolled environment

Civilian objects that may falsely ID as targets

Context for lethal force changes post-AWS launch

## Joint human-machine cognitive system

Environment changes post-AWS launch

Automation bias

testing/validation impossible

Poor training

Inadvertent interactions ("flash war")

Friendly Forces

User

coding errors

AWS

Human-machine interface

Adversaries

poor design

Hacking

User creativity/innovation

Poor interface design

Spoofing

Cultural biases

Hidden interactions

Jamming

Insufficient testing

Behavioural hacking

Civilians used as shields

# Socio-Technical System

- **Accidents are inevitable in complex and tightly coupled systems.**
- **Safety is always one of a number of competing objectives (not least in conflict)**
- **Redundancy is often a cause of accidents**
- **Development (and testing) aren't value-free processes**
- **Hidden interactions might have particular risk potential in systems reliant on machine learning.**