Australian Government
Department of Defence
Defence Science and Technology Organisation

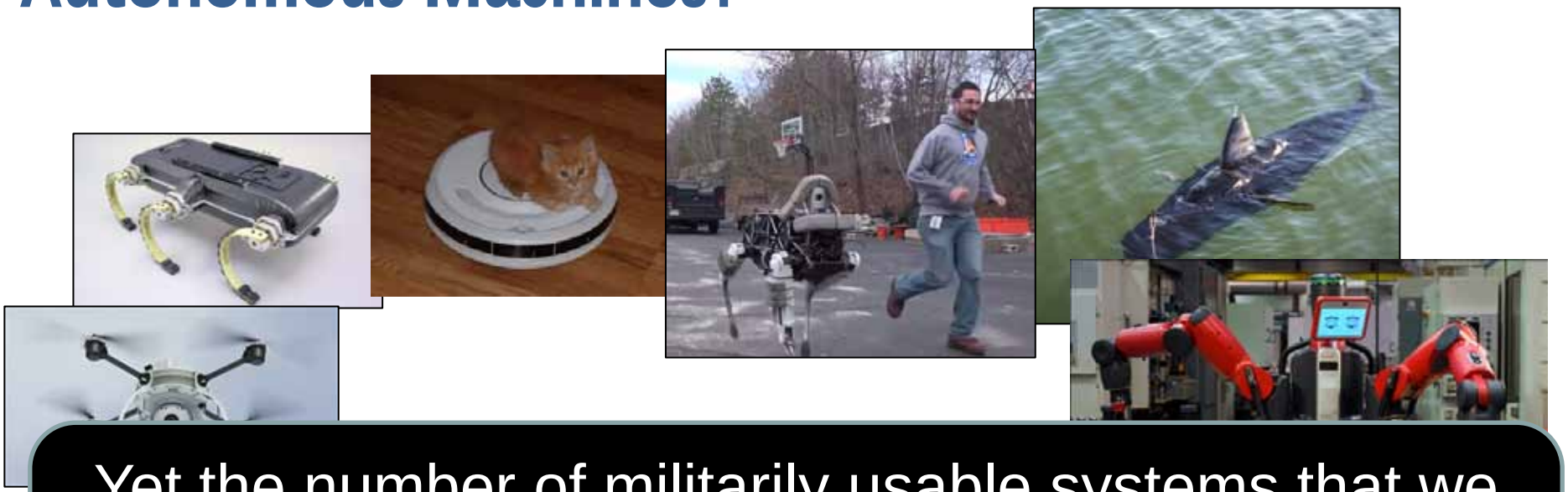
Towards Machine Plasticity

Dr Jason Scholz
Research Leader Decision Sciences

DSTO | Science and Technology for Safeguarding Australia

3 July version

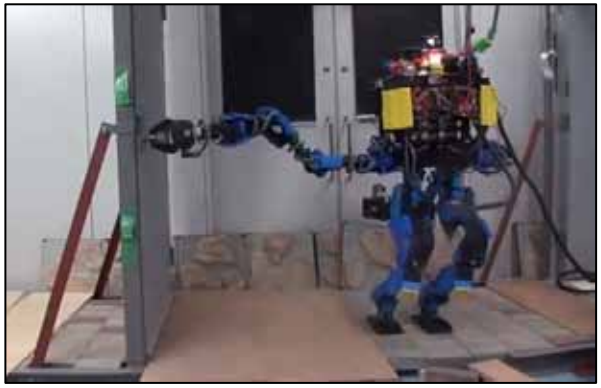
Autonomous Machines?



Yet the number of militarily usable systems that we can truly regard as autonomous is precisely zero – Darryn Reid

or just Automations?

Hyundai *autonomous car competition called off after rain - Oct 2014*

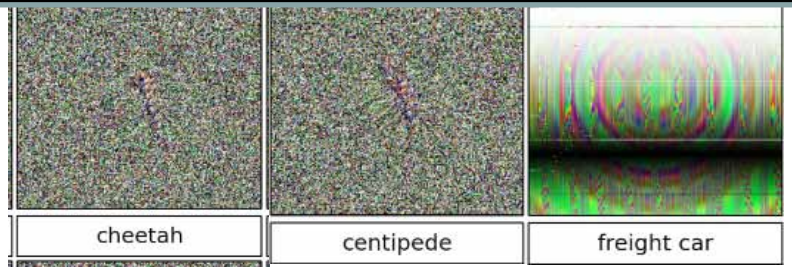


DARPA *Robotic Challenge winner SCHAFT opens a door but gets a surprise - Dec 2014*

Deutsche ***Euro Hawk*** *cancelled deemed unverifiable without massive expense - 2013*

There is the autonomy we *dream* about and there is the automation we possess

14,000,000 + ***Roomba's*** *sold Worldwide but don't work without preconditioning their environment*



Deep Neural Networks *(Google, Facebook & autonomous cars) World benchmark object identification in images and video, yet all these are >99.6% certainty classifications are wrong!*

Strong Expectations & Certainty

People tend to believe...

... that we can precisely define out problem situations.

... that we can accurately describe complete solutions.

... that the path from problems to solutions is a linear matter of efficiency and expected utility.

... that success and failure are crisp and symmetric and similarly accurately definable.

... that it should all be about 'positive' sounding stories of success that make us feel good.

... in justification, prediction and relative certainty.

Military Environments

"The atmosphere in which War moves, is one of danger, physical effort, **chance** and **uncertainty**"

Carl Von Clausewitz
On War 1832
Chapter 3, On Military Genius
Howard & Paret 1976

Current machines are automations

- Designed for **chance** events: ergodic distribution averages over sample sets
- Language of stochastic, efficiency, optimisation, reliability, redundancy, robustness*
- Managed environments: Explicitly structured and closed: dull, dirty and dangerous (localised and immediate)
- Ever more complicated designs to cope with real world situations undermines trust and makes systems unverifiable for operational use

Future machines need to be based on a different R&D problem choice

- Fundamental **uncertainty**, non-ergodic distributions w/out sample sets
- Language of tychastic, resourceful, innovative, agile, resilient, anti-fragile …
- Unmanaged environments: Unstructured and open: unforseen (novel), and dangerous (global and non-immediate)

It simply matters ***nothing*** how well a system works on average if we cannot tolerate the consequences of it failing

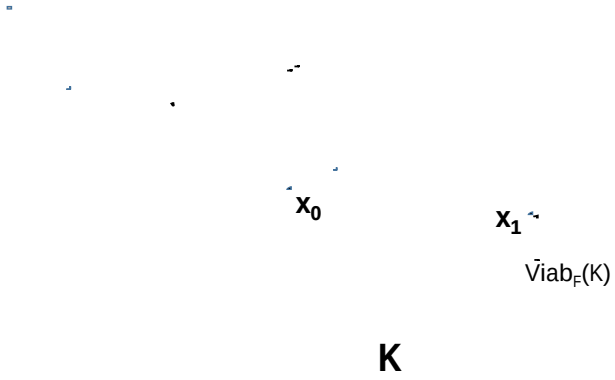
Plasticity Imperative

- Plasticity is concerned with events in the tails – not the middle – of *arbitrary*, *non-ergodic* and fundamentally *unknowable* distributions.
- Symbolic processing is critical
- This is about *effectiveness* under irreducible uncertainty: usual *efficiency and utility-maximising concepts are irrelevant*!
- Autonomous systems require control without strong prediction.
- New ways of applying existing techniques?
- Entirely new techniques?

Foundations of Autonomy

Example: Viability Theory

- Can guarantee bounded action under fundamental uncertainty.
- "**a run time parallel oracle function to determine verify/validate (police) the autonomous function decisions**"
- From a point x_1 in the viability kernel of the environment K there starts at least one evolution viable in K forever. All evolutions starting from x_0 ε K outside the viability kernel leave K in finite time.



$x'(t) \in F(x(t))$ *for uncertainty*
$x(t) \in K$ *for necessity*

Jean-Pierre Aubin, Alexandre M. Bayen and Patrick Saint-Pierre (2011). *Viability Theory: New Directions*. Springer.

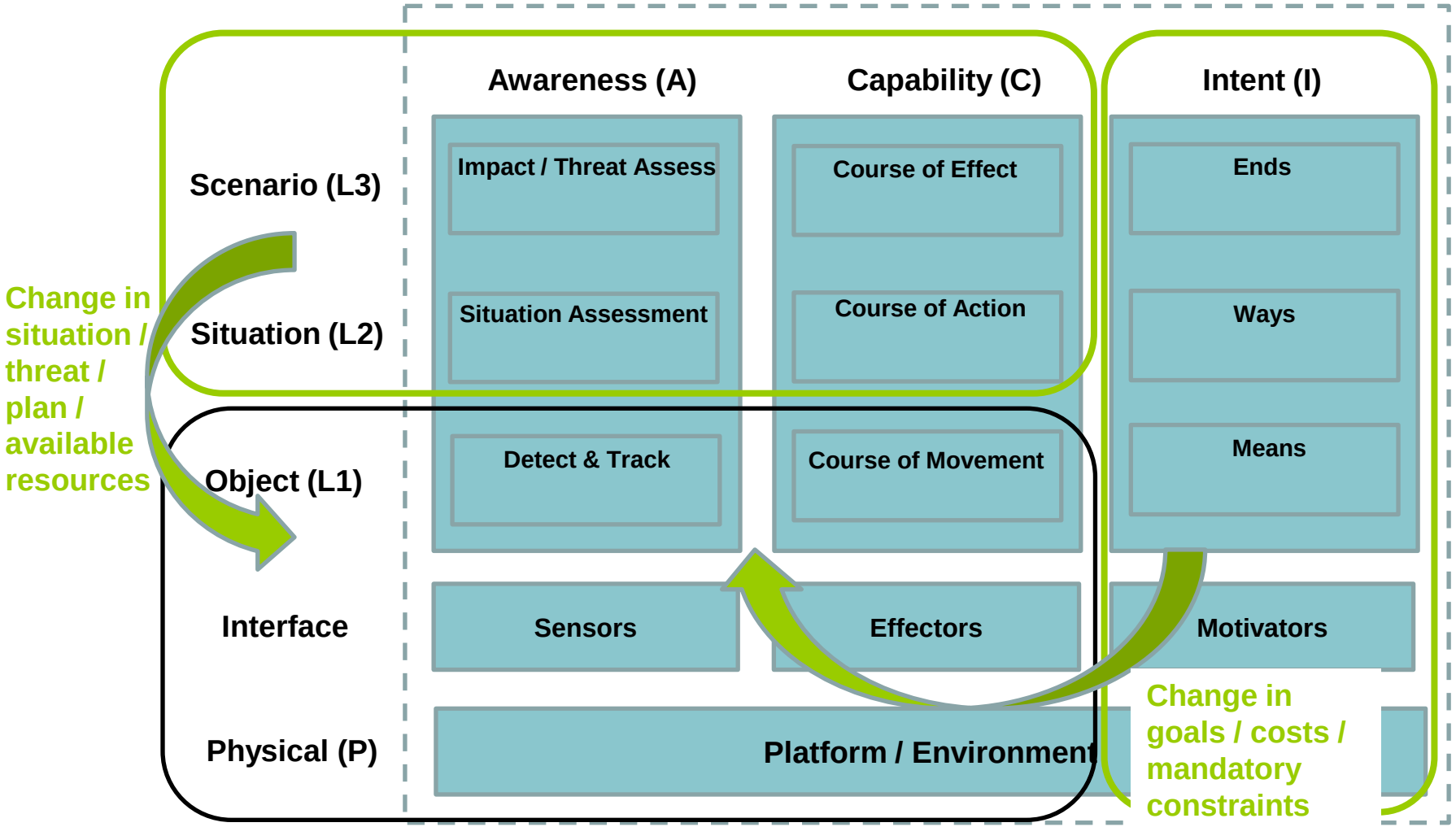
Machine Cognition

- Thinking Fast *and* Slow (at the same time)
  - System I – fast, feeling, parallel, inexplicable
    - *A machine* for jumping to conclusions, a story teller, confidence
    - Usually right, but often doesn't know when it's wrong - bias
  - System II – slow, logical, sequential, reasoned
    - A little bit of self control (but don't need to do much of it!)
    - Induction (generalise), Abduction (explain), Deduction (predict)
- Metacognition and Reorganisation (learning)
  - Metacognitive Strategies
  - Extreme Programming (human-guided and self-guided)
  - Creativity
- Action Trinity
  - Information Fusion (awareness)
  - Resource Management (capability)
  - Goal Management (intent)

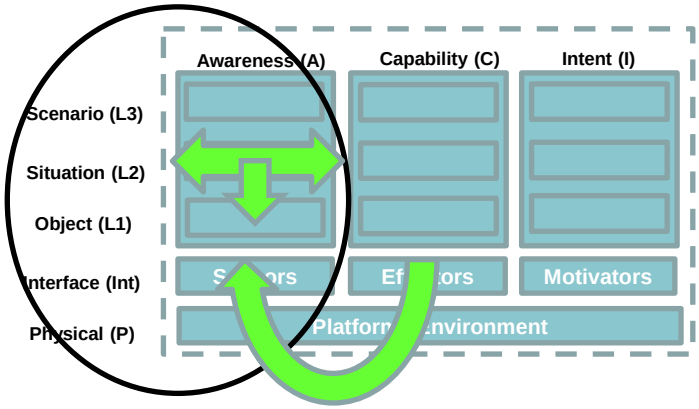
Machine Cognition

Abstraction levels

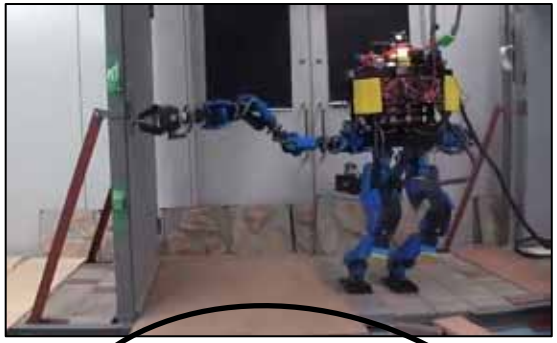
Extended from DSTO Fusion for Situation Awareness Initiative Model (Lambert).
"Blueprint for C2" (Scholz et al) Fusion 2012 (3 papers).

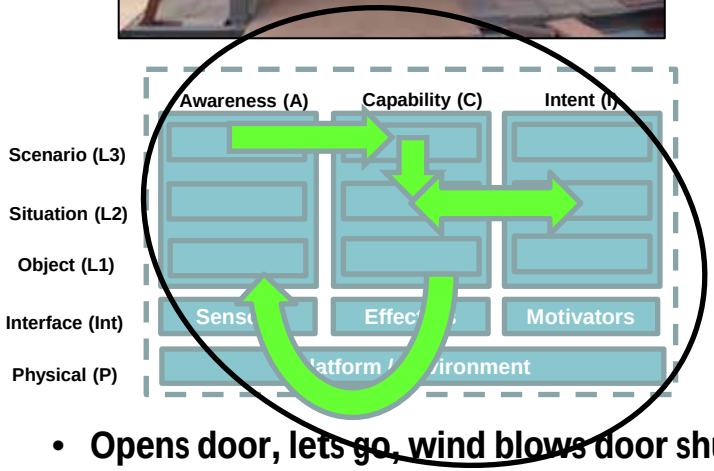


Example Solutions



- **Light conditions changed**
- **L2 Awareness (Situation Assessment)** determines context (water on the road)
- **L1 Awareness (Detect & Track)** lowers the Particle Filter threshold setting
- **Execution**: Vehicle safely drives on





- **Opens door, lets go, wind blows door shut**
- **L3 Scenario Awareness** projects that opening door again will result in repeat consequence
- **L3 Capability (Effect)** hypothesises force needed to hold door open
- **L2 Capability (COA)** proposes use of robot's foot to hold door open (change in routine)
- **L2 Intent (Ways)** to let go of door modified to not let go until foot is in place.
- **Execution**: Robot passes through

Conclusion

- Defence applications of autonomous systems under managed conditions have reached a limit of diminishing returns
- The ability to deal with *uncertainty* is the fundamental limitation to large scale future deployable systems in Defence
- There is hope if we redress the choice of research problem
  - We can't expect a different result by doing more of the same
- A range of cognitive abilities indicated will be needed to achieve individual machine plasticity, with extension to plasticity of human-machine and social interaction