Ethically-aligned Experimentation and T&E in the Profession of Arms – Takeaways Pt B

A human centric view of complex military and safety critical intelligent and autonomous systems

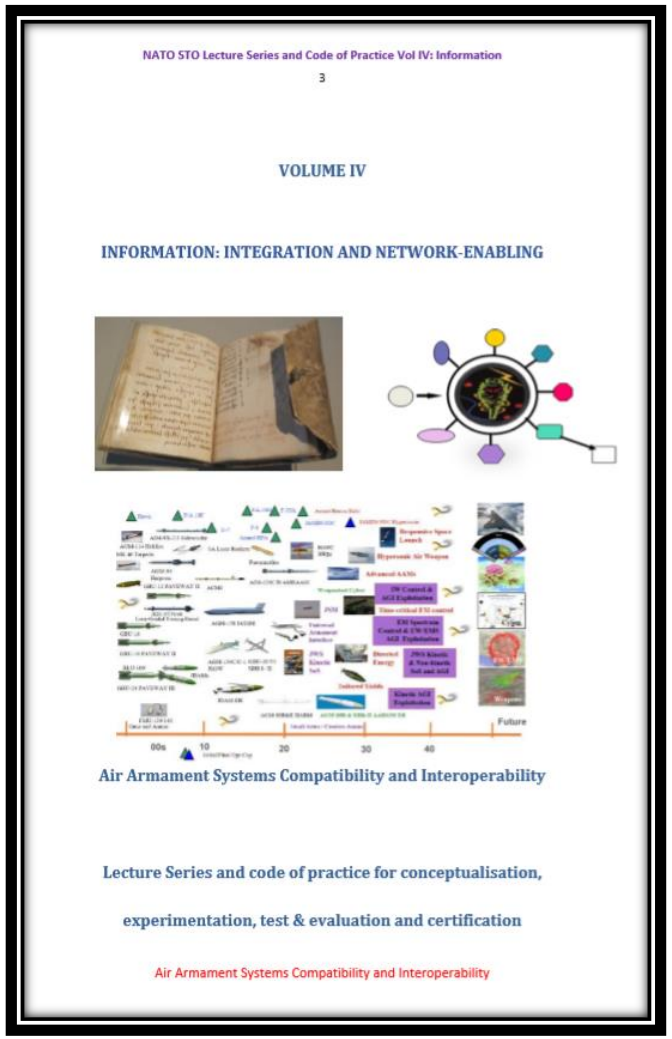
38th ITEA Symposium Tutorial 14 September 2021 0000 – 0400 (for me)

Dr Malcolm Tutty & Keith Joiner

Vice President ITEA Southern Cross Chapter & Senior Lecturer in T&E, UNSW

13-September-2021

NATO Air Armament Lecture Series



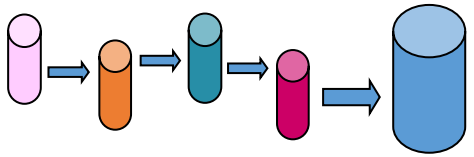
IV. Information: Integration and Network-enabling

Background Challenge 1 - Growing System Synthesis

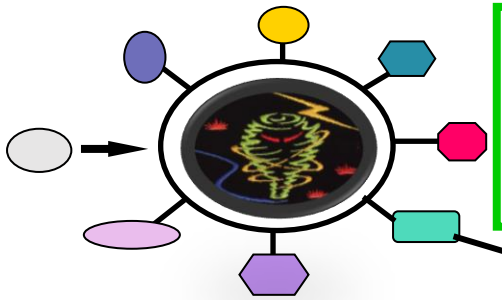
- Systems are becoming so synthesised or fused, complex and interdependent that they can, even without taking into account human agency, have *emergent properties* or exhibit *behaviours* that vary to an extent that is not easily predicted.
- The number of permutations of modern software-intensive systems make classical rigorous testing of them, all but impractical, (Cofer, 2015).
- Reliance on some modelling required, (Hecht, 2015).
- Continuous through-life monitoring required, (Normann, 2015).
- Situation challenges safety-critical assurance and mission system capabilities, (Tutty, 2016).

Systems – The Platform, Flight **&** Mission Simulators, Syst Eng Integ Lab, Part Task Trainers

SoS – Σ Platform **Type**(s) + Flight & Mission Sims + SE Integ Lab + Part Task Trnrs + Weapons + JWS + EW + LVC + TBMCS + Intel…



SoS/FoS - Small stovepipes to large stovepipes **–** **NO**



FoS - Σ SoS - Loosely coupled and tightly integrated **–** **YES**

Background Challenge 2 – Higher-Order Human Functions

- Software-intensive systems are enabling higher-order human-like functions such as strategies and decision-making not simply control.
- The difficulty in specifying what the system must do becomes harder.
- More crucial to include representative human agency and decision making to adapt the systems during development.
- Usability tests for meaningful human control (Roff & Moyes, 2016) are fundamental

Level of Autonomy	Mission Decision Transitions				
	Find	Fix/Track	Target	Engage	Abort
Manual	Human	Human	Human	Human	**Human**
Human-in-the-loop	Human or System	Human	Human	System	**Human or System**
Human-on-the-loop	Human selected ??	System	System	System	**Human or System**
Human-out-of-the-loop	**System**	**System**	**System**	**System**	**Human**

Background Challenge 3 – Cyber Threat Complexity & Information Dominance

- The threat to weapon systems has adapted as a result of the [western] push for information exploitation/dominance, Jordan et al (2016).
- Cooperative engagement has systems responding in defence of other systems.
- Such inter-connected systems enables exploiting the broader cyber-attack surfaces of such systems.
- Not just about malicious attacks but as part of multi-layered hybrid/hyper warfare, Allen (2017).
- Offensive cyber
- The Threat is more complex and probably more adaptive.

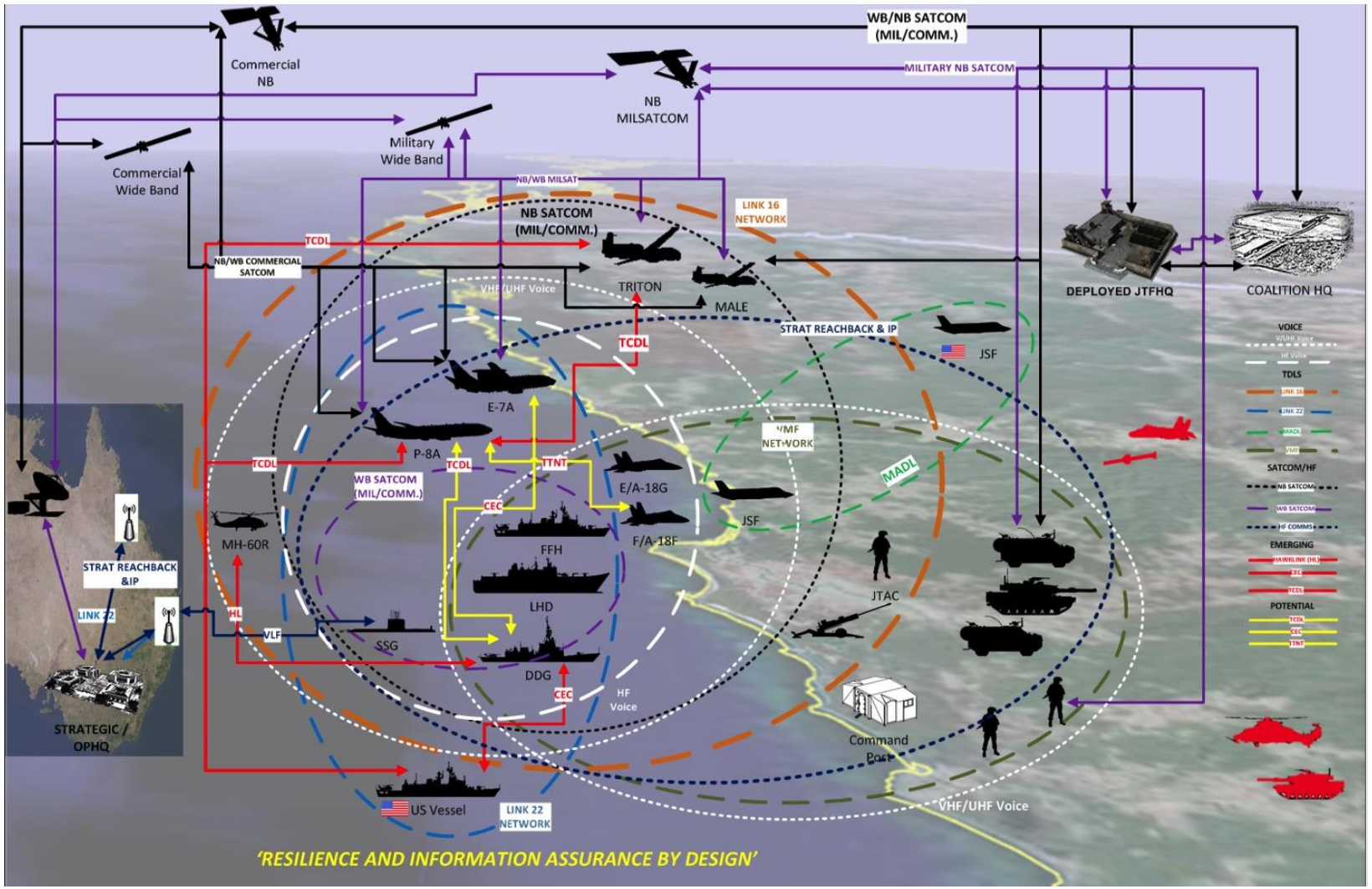
Challenge 4 - Requirements Stasis

- Requirements stasis during development & build of large complex systems, largely through an emphasis on project management cost & schedule achievement on contracts, & unfortunately including their processing & software.

- Such a requirements stasis soon creates an alternative reality that is too far out of alignment with the contemporary family-of-systems & strategic/operational reality into which that complex new system must go into service.

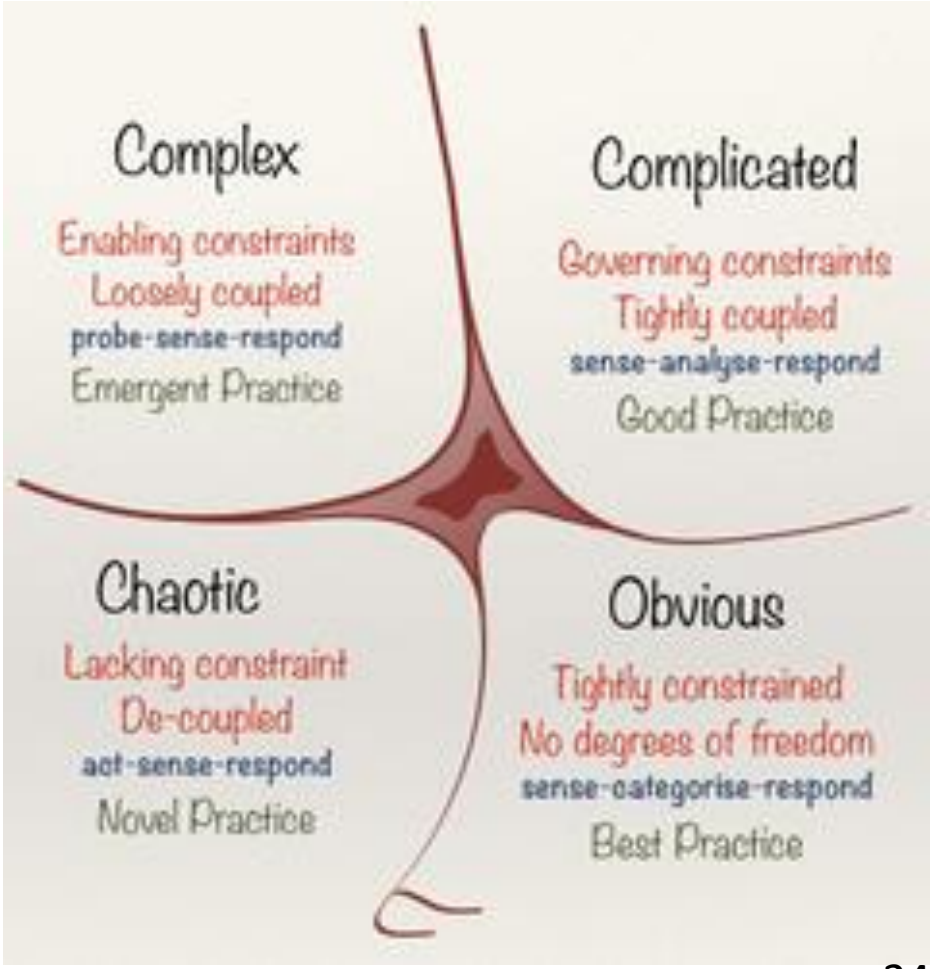
COMPLEXITY, INTERCONNECTEDNESS, AND VULNERABILITY

INCREASE LETHALITY, RESILIENCE AND RAPID ADAPTIBILITY



Humans fight wars, not drones or robots.

IV.1: Information and Warfighting Family-of-System-of-Systems under control.



240px-Cynefin_as_of_1st_June_2014

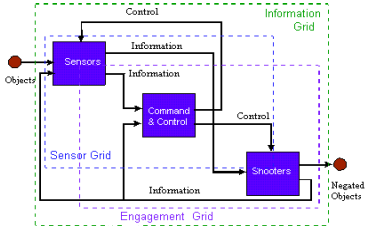
Simple, Complicated and Complex Systems 101

- Simple (obvious) Systems – graduates, COTS/MOTS – straight-forward systems and potentially Systems of Systems with linear outputs given inputs.

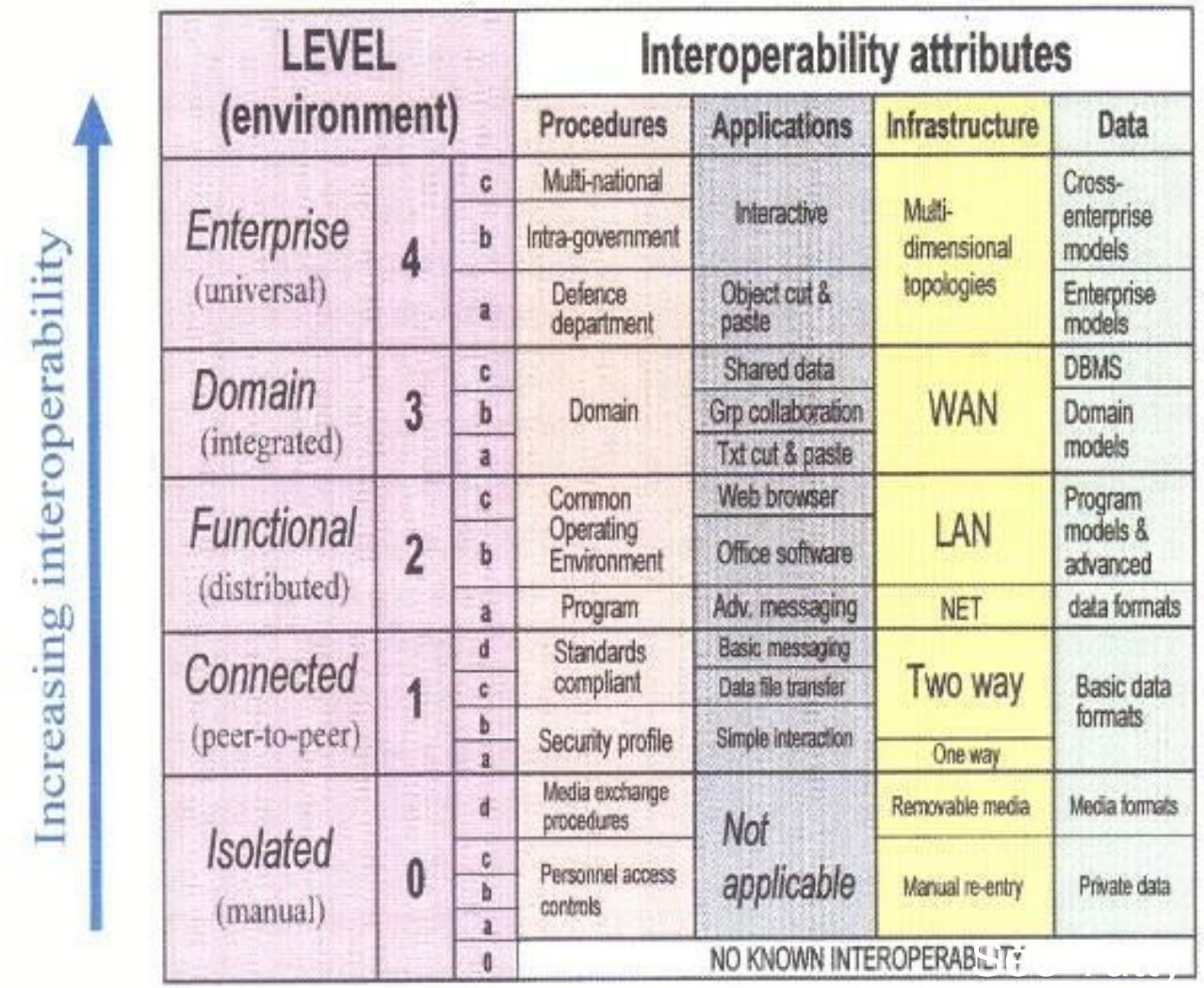
- Complicated Systems – interesting, non-trivial, linear systems and typically SoS that need team(s) of competent systems engineers, testers and users to confirm emergent properties meet the cited need that are usually time invariate/stable and are under organisational control.
  Traditional Systems Engineering & T&E approaches cope well.

- Complex & Complex, Adaptive Systems – non-trivial, potentially non-linear systems and SoS that may be time varying / unstable / 'learn' and do not have unity in organisational arrangements.
  Traditional SE and T&E do not cope with time dependencies and adaptability.

- Chaos …

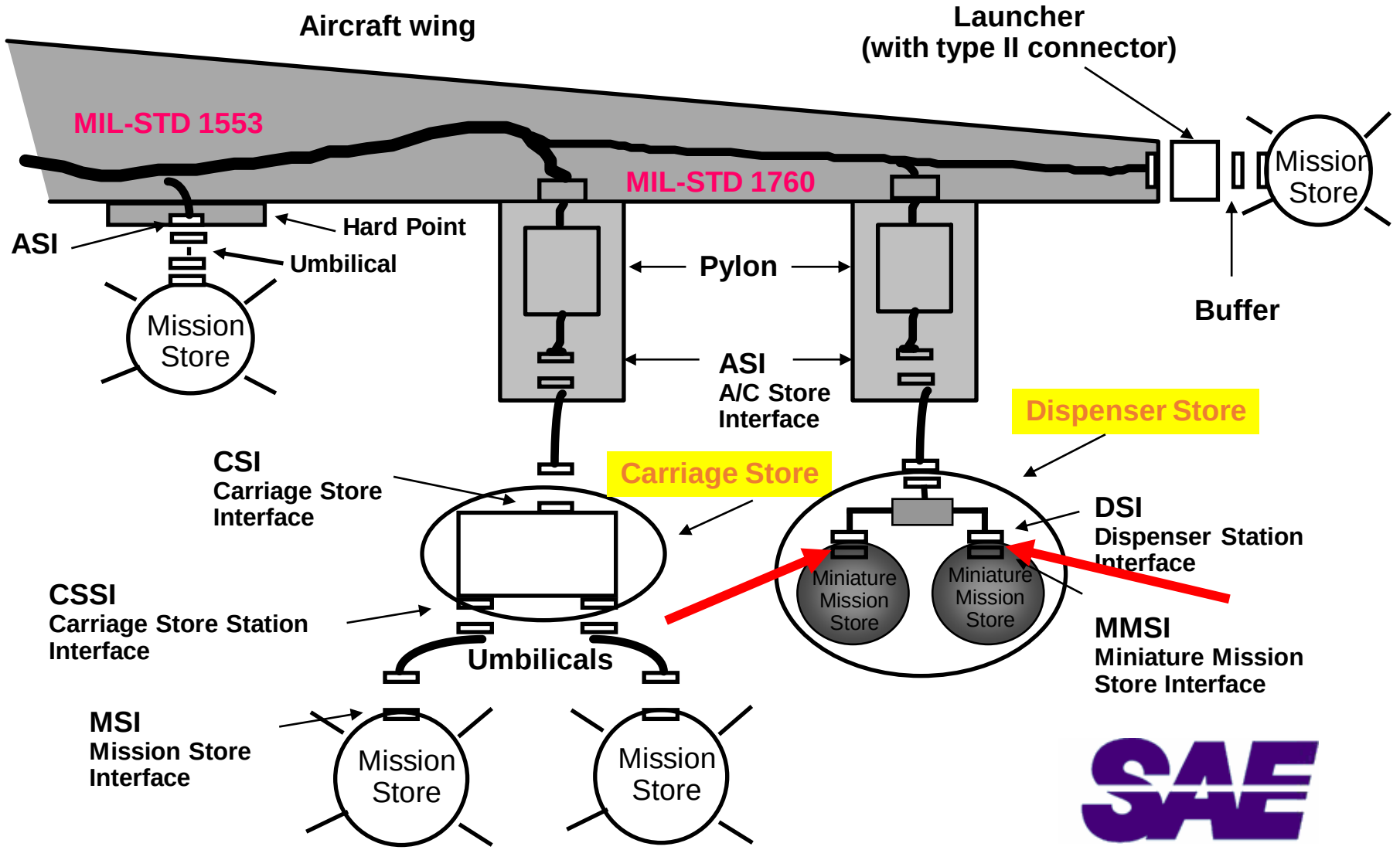


NEO: network-enabled operations via LISI



LEVEL (environment)			Interoperability attributes			
			Procedures	Applications	Infrastructure	Data
Enterprise (universal)	4	c	Multi-national	Interactive	Multi-dimensional topologies	Cross-enterprise models
		b	Intra-government			
		a	Defence department	Object cut & paste		Enterprise models
Domain (integrated)	3	c	Domain	Shared data	WAN	DBMS
		b		Grp collaboration		Domain models
		a		Txt cut & paste		
Functional (distributed)	2	c	Common Operating Environment	Web browser	LAN	Program models & advanced
		b		Office software		
		a	Program	Adv. messaging	NET	data formats
Connected (peer-to-peer)	1	d	Standards compliant	Basic messaging	Two way	Basic data formats
		c		Data file transfer		
		b	Security profile	Simple interaction		
		a			One way	
Isolated (manual)	0	d	Media exchange procedures	*Not applicable*	Removable media	Media formats
		c	Personnel access controls		Manual re-entry	Private data
		b				
		a				
		0	NO KNOWN INTEROPERABILITY			

Plug & Play Weapons: Stores Universal Armament Interface (UAI)



AI today and 'existential threats'

Generating advantage and the intellectual edge

Penn State – Human Cognitive Functions

(1) **enhanced memory –** machine learning algorithms,
(2) **attention and search -** deep longer-term importance but lacks shorter-term context;
(3) **comprehension and expression -** significantly improved understanding of information,
(4) **planning and executing activities -** AI-extenders that could develop models of action, testing and comparing various activities against known and projected enemy capabilities; and
(5) **Metacognition –** what individual or system knows about its own cognition, or cognition in general: literally 'thinking about thinking'.

Ryan, Gen Mick, *Extending the intellectual edge with artificial intelligence,* Australian Journal of Defence and Strategic Studies | Vol. 1 No.1 pp 23-40

- **Machine-learning** is the ability of a machine to learn new abilities without having been explicitly programmed by it's designer(s).

- **ANI - Artificial Narrow Intelligence:** machine intelligence that *equals or exceeds* human intelligence *for specific tasks:* IBM's Deep Blue (Chess), Google's AlphaGo (go), High-Frequency Trading Algorithms, or any automatic systems performing beyond human reach (Google Translate; spam filters; the guidance systems of missiles and point-defense systems cannons etc.).
- **AGI - Artificial General Intelligence** ("strong AI"): machine intelligence meeting the *full range* of human performance across any task.
- **ASI - Artificial Superintelligence**: machine intelligence that *exceeds human intelligence* across any task. Machines becoming self-aware …

History of ICT/Cyber/AI and weapons: operational views and perspectives

- **Many of the first applications of what is now called AI since 1956 were in ballistics, fire control systems, analog weapons seekers and warhead effectiveness models!**

- **The "3rd AI Spring" (1990 – 2010) shifted focus to the commercial world with 'big money' being injected into 'Big Data' and 'Deep Learning' since 2011 …**

- **Safety critical and mission affected operations.**

> 2030 ICT / Cyber / ASI

- **Mission critical – safety affected operations.**

> 2010 ICT / Cyber / ANI
>2020 ICT / Cyber / Strong AGI

- **Mission affected/advisory – 'non-safety critical' operations.**

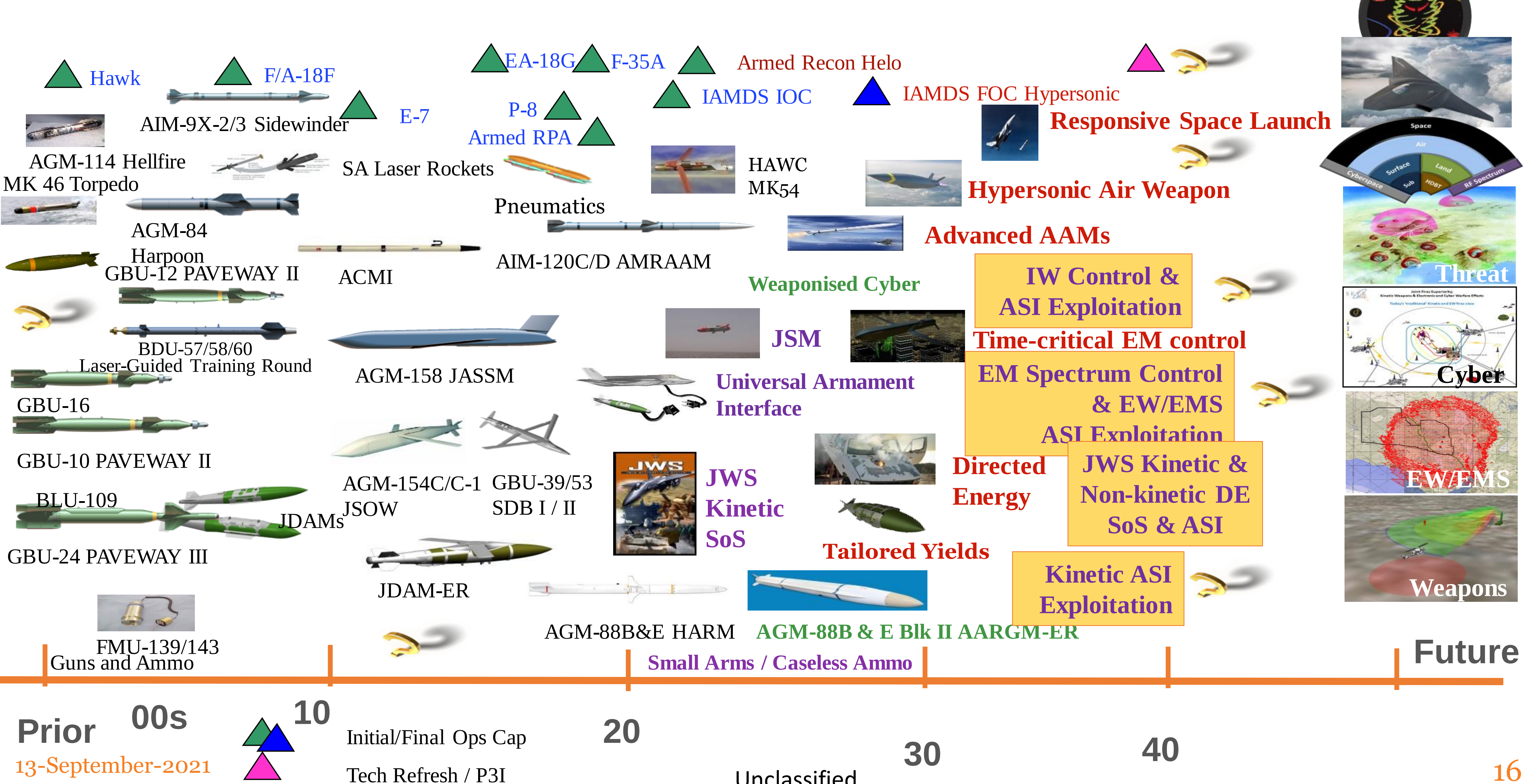
< 2000 ICT/Cyber/ Narrow AI

EW and IW is about …

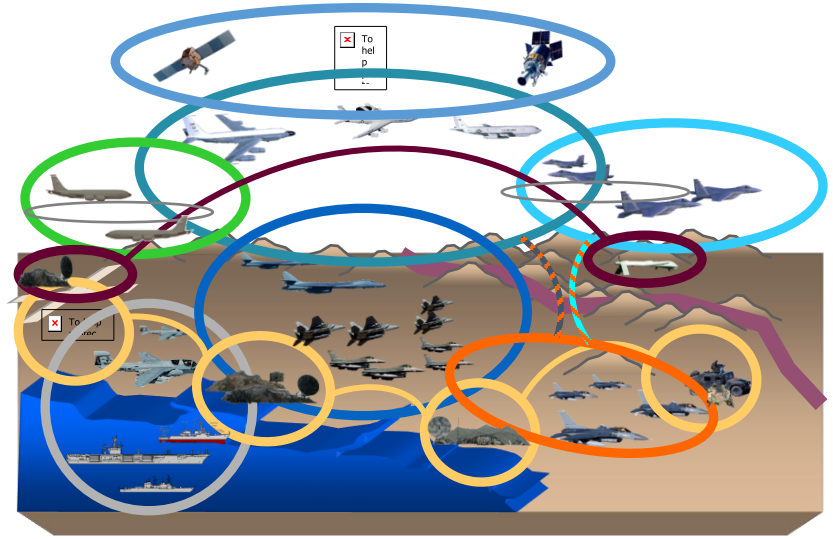
- Warfighting … not *Physics*
- Effects … not *Systems*
- Using mediums … not *Owning* mediums
- Organisation … not organisations
- Synergy … not *Segregation*
- Integration … Not just *Synchronization*
- Preserving national treasure … not *Being* a national treasure
- What's important … not *Who's* important
- The Right Force… not *Equal Shares* of the force
- Mavens perspective on EW: Air Domain is the Medium and the Information Domain has the "Protocols and Waveforms" …

EW is not just using the power range equation for the frequency

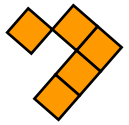
Future Armament Families of SoS

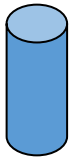


A Family of System of Systems

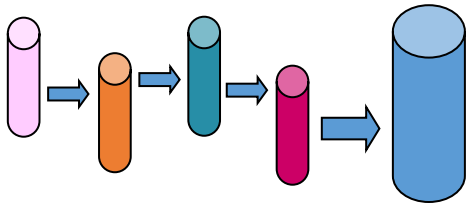


Systems – The Platform, Flight **&** Mission Sim, SE Integ Lab, PT Trnrs

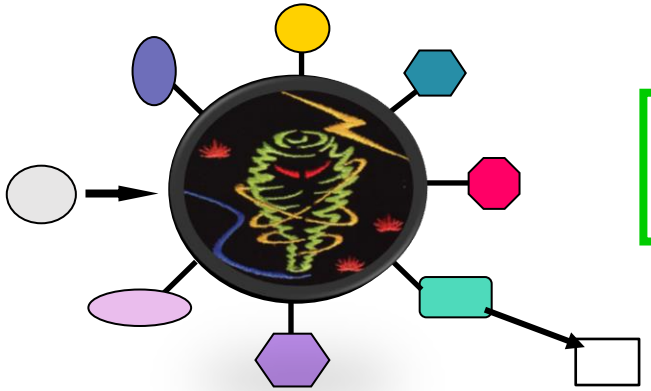




SoS – Σ Platform **Type**(s) + Flight & Mission Sims + SE Integ Lab + PT Trnrs + Weapons + EW + LVC + TBMCS + Falcon View + Intell…



SoS/FoS - Small stovepipes to large stovepipes **– NO**

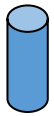


FoS - Σ SoS - Loosely coupled and tightly integrated **– YES**

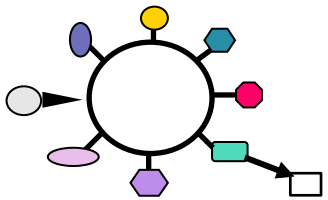
Definitions - Systems, SoS & Family of SoS

- **Systems** An integrated composite of **People**, **Products** and **Processes** that provide a **capability to satisfy a stated need or objective**.

- **System of Systems** A SoS results when independent and useful systems are integrated into a larger system that delivers unique capabilities resulting from a series of acquisition actions and ***typically has no one single management entity***.
  SoS types: *Virtual – Collaborative – Acknowledged – Directed.*



- **FoS - Family of SoSs** A FoS results when **independent and useful SoSs are integrated into a larger system that delivers unique capabilities**.
  To achieve joint mission capabilities, FoS should be considered to be made up of SoS.



- ***TRL / CPL.*** Technology Readiness Level **/ Capability Preparedness Level**.

IV.2 Warfighting Preparedness

0	No identified capability, basic principles conceptualised and studies initiated. Capability development being proposed.
1	Basic principle observed and reported. Studies or initial investigations undertaken. Capability development initiated.
2	Technology concept and/or application formulated. Potential applications have been identified.
3	Analytical and experimental critical function and/or characteristic proof of concept. R&D has been initiated, work towards validating the concept done. Capability is Managed.
4	Component and/or breadboard validation in lab environment. The basic elements of the Capability (SoS/Major system/product) have been integrated to show they will work.
5	Capability SoS/Major systems and/or components/breadboard validation in relevant environment. A higher fidelity validation of the Capability in a realistic environment. Capability is Defined.
6	**Capability SoS model or prototype demonstration in a relevant / realistic environment.**
7	**Capability SoS prototype demonstration in operational environment with representative personnel and C2 iaw Conops.** 'Production' can now commence. Capability is being Quantifiably Managed.
8	Actual Capability SoS completed and mission qualified through test and demonstration with operational personnel and C2. **Actual Capability SoS has been successfully tested, qualified and certified iaw Conops.**
9	Actual Capability SoS proven through successful mission operations with operational personnel and logistics support. **Actual Capability SoS has been successfully fielded.**
10	Actual Capability SoS has been found to be operationally effective, suitable and sustainable in successful real-world network- enabled operations with other identified SoS as a FoS. Capability is being Optimised.

Yin - Yang: Danger - Confidence

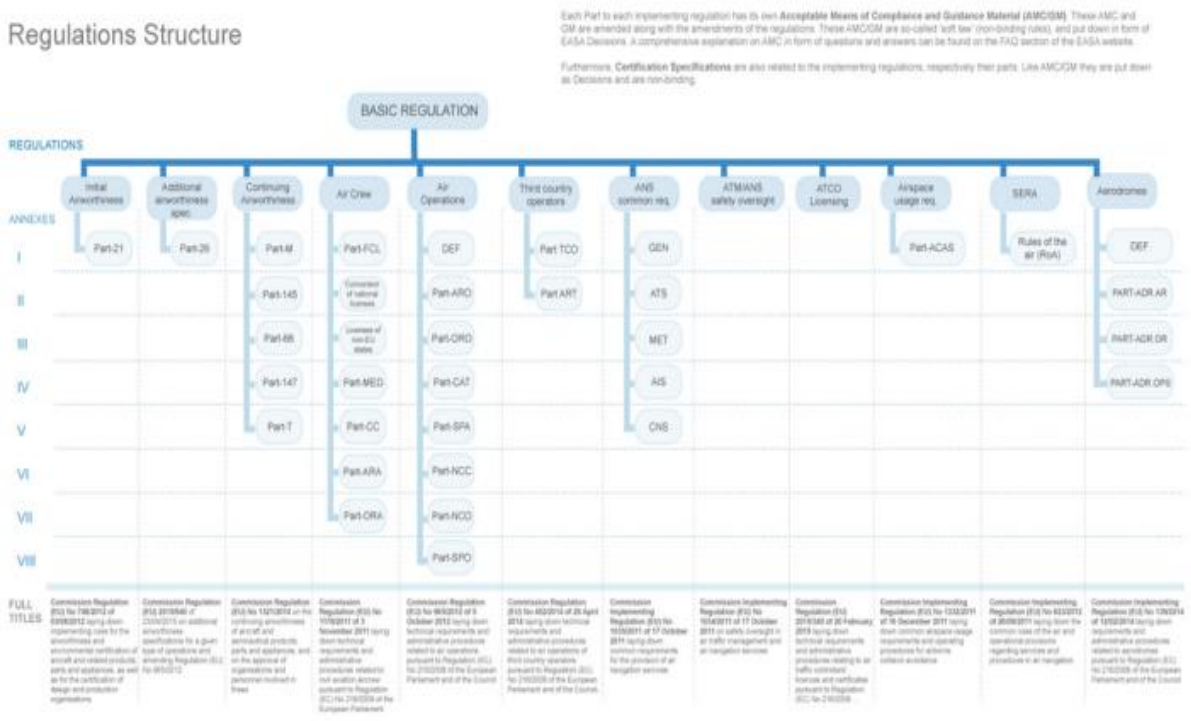
Yin - Danger		Risk Level	Yang - Confidence γ
0 1	Operations Only iaw SPINS	Extreme	Very, very, very ‘Bad’
2 3	High Risk Operations (needs Waiver)	High	‘Bad’
4 5	Test and Operations (needs Test Plan)	Medium	Some ‘goodness’, there are ‘Issues’, some potentially ‘Bad’.
6 7	Mission Essential Personnel Only (needs Test / Exercise Plan)	Low	‘All goodness’, ‘On Target’, ‘Under control’.
8 9 10	Very Low: Public Risk	As Low As	Robust Solution that meets and exceeds the Need identified that is being IV&Ved & Optimised

Airworthiness: EASA and EMAR

European Defence Agency (EDA – 27 European Union militaries – not NATO) and European Aviation Safety Agency (EASA – EU's CASA/FAA)

EUROPEAN MILITARY AIRWORTHINESS REGULATOR (EMAR): standardised vehicle certification criteria and process for EU, UK, US, Canada and Australia …

TYPE CERTIFICATION BASIS: establishes the tailorable airworthiness certification criteria (aligned with EASA legislated requirements) to be used in the determination of airworthiness of all [EDA military] manned and unmanned, fixed and rotary wing air vehicle systems. Cites STANAG 7068.



EASA European Aviation Safety Agency

EUROPEAN MILITARY AIRWORTHINESS CERTIFICATION CRITERIA- EMACC

EUROPEAN MILITARY AIRWORTHINESS CERTIFICATION CRITERIA- EMACC

EUROPEAN DEFENCE AGENCY

MILITARY AIRWORTHINESS AUTHORITIES FORUM

Ethically Aligned Design …

- As ***A/IS*** become pervasive, we need ***societal and policy guidelines*** in order for such ***systems to remain human-centric***, serving humanity's values and ethical principles.
- ✓ These systems have to behave in a way that is ***beneficial to people*** beyond reaching functional goals and technical problems.
- ✓ Allow for an ***elevated level of trust between people and technology*** that is needed for its fruitful, pervasive use in our daily lives.

- **Key Principles:**
  - **Human Safety & Rights – Prioritising Well-being – Accountability - Transparency – A/IS Technology Misuse and Awareness of It**
  - **Under no circumstances is it morally permissible to use AWS without meaningful human control, and this should be prohibited.**
- **Future Technology Concerns**

- ***Reframing Autonomous Weapons***
- Autonomous systems designed to cause physical harm have additional ethical dimensions as compared to both traditional weapons and/or autonomous systems not designed to cause harm. These ethical dimensions include, at least, the following:

❖ Ensuring ***meaningful human control*** of weapons systems

❖ Designing automated weapons with audit trails to help guarantee accountability and control

❖ Including adaptive and learning systems that can explain their reasoning and decisions to human operators in a transparent and understandable way

❖ Training responsible human operators of autonomous systems who are clearly identifiable

❖ Achieving behaviour of autonomous functions that is predictable to their operators

❖ Ensuring that the creators of these technologies understanding the implications of their work

❖ Developing professional ethical codes to appropriately address the development of autonomous systems intended to cause harm

Version 2 - For Public Discussion IEEE Advancing Technology for Humanity

- ***Safety and Beneficence of Alleged Artificial General Intelligence (AGI) and Artificial Superintelligence (ASI)***
- ***Affective Computing***
- ***Mixed Reality***

Ethically Aligned Design …

- **Implementing Norms: Issue 3: Failures will occur.** Because designers cannot anticipate all possible operating conditions and potential failures of A/IS, multiple additional strategies to mitigate the chance and magnitude of harm must be in place.

- **Safety and Beneficence of AI**
- 1. Contribute to research on concrete problems in AI safety …
- 2.Work to ensure that A/IS are transparent, i.e., that their internal reasoning processes can be understood by human operators.
- 3. Work to build safe and secure infrastructure and environments for development, testing, and deployment of powerful A/IS.
- 4. Work to ensure that A/IS "fail gracefully" (e.g., shutdown safely) in the face of adversarial inputs, out-of-distribution errors, unexpected rapid capability gain, and other large context changes.
- 5. Ensure that A/IS are corrigible, and assist (do not resist) the operators in shutting down and modifying the system.
- 6. Explore methods for making A/IS capable of learning complex behaviours and goals from human feedback and examples.
- 7. Build extensive knowledge layers and automated reasoning into systems to expand their contextual awareness and common sense so undesirable side effects can be determined and averted dynamically.

- ***Autonomous systems designed to cause physical harm have additional ethical dimensions as compared to both traditional weapons and autonomous systems not designed to cause harm.***
- Recommends that technical organizations … should ensure that there is ***meaningful human control*** of weapons systems:
- • That automated weapons have audit trails to help guarantee accountability and control.
- • That adaptive and learning systems can explain their reasoning and decisions to human operators in transparent and understandable ways.
- • That there be responsible human operators of autonomous systems who are clearly identifiable.
- • That the behaviour of autonomous functions should be predictable to their operators.
- • That those creating these technologies understand the implications of their work.
- • That professional ethical codes are developed to appropriately address the development of autonomous systems and autonomous systems intended to cause harm.

Ethically Aligned Design …

- **Issue 6: There are multiple ways in which accountability for the actions of AWS can be compromised.**
- **Background.** Weapons may not have transparency, auditability, verification, or validation in their design or use.
- Various loci of accountability include those for commanders, and operators.
- Ideally all procurers, suppliers, and users of weapons systems components have accountability for their part of every weapons system, potential incorporation in future systems, and expected and potential users.
- **Candidate Recommendations**
- • Designers should follow best practices in terms of design process, which entails clearly defined responsibilities ...
- • Systems and components should be designed to deter the easy modification of the overall weapon to operate in fully autonomous mode.
- • Further exploration of black box recording of data logs, as well as cryptographic, blockchain, and other technical methods for tracing access and authorization of weapons targeting and release is needed.
- • System engineers must work to the same high standards and regulations of security for AWS design from a cybersecurity perspective than they would for any other work. Weapons systems ought to be designed with cybersecurity in mind...
- • ***Procurement authority***: only contract with contractors who have proper legal and security processes; carry out Article 36 reviews at all major steps in the procurement process; maintain database of design, tests, and review evidence.
- • ***Contractors***: ensure design meets relevant engineering and defense standards for military products; deliver evidence for Article 36 reviews using, but not restricted to, design reviews and simulation models; provide evidence requested by user for setting ROE; ensure design has clear criteria for decisions made by their product.
- • ***Acceptance body***: have validation and test plans for behaviour of actual system produced; test weapons systems in a number of representative scenarios; have plans to ensure upgrades are reviewed against IHL criteria such as Article 36.
- • ***User/military commanders***: only operate weapons systems with meaningful human control …
- • Weapons systems must have *default modes of operation* agreed with campaign planners before operation commences.
- • Ensure as many aspects of weapons systems as possible are designed with *fail-safe behaviours*.
- • *Ensure clear embedded lines of accountability* in the design, deployment, and operation of weapons.
- • *Trusted user authentication logs* and audit trail logs are necessary, in conjunction with meaningful human control.
- • ***Tamper-proof the equipment used to store authorization signals*** and base this on open, auditable designs.
- • The hardware that implements the human-in-the-loop should not be physically distinct from operational hardware.

Armament FoS safety criteria

- **I.1.5.0 Safety.** Installations shall provide maximum protection against inadvertent release as a result of either human error, carelessness, or the material failure of components of the firing control system. Operationally categorised as:
  - **A – safety and mission critical operations.**
  - **B – mission critical – safety affected operations.**
  - **C – mission affected/advisory – 'non-safety critical' operations.**
  - ***D – Prototype System / SoS mission affected/advisory – not yet 'safety critical'***
  - ***E – Uncategorised [and very troubling]!***

Armament safety criteria

(STANAG 7068 App B Pt I.1.5)

- I.5.1 **Ground safety device.** The store release system shall be equipped with a positive safety device or devices to preclude functioning, dropping, launching, or ejecting of suspended stores or activation of ejector devices when the aircraft is on the ground even if the release or actuation system is energised.
- I.5.2 **Erroneous switch selection and single component failure.** The control of store stations shall be such that no single operation on the part of any crew member will result in the inadvertent release or function of a store. ***No single component failure in the function or release control system shall result in the inadvertent function or release of a store.***
- I.5.3 **Safetying.** Parts which may cause a hazardous condition by working loose in service shall be safetied or shall have other approved locking means applied.

Armament FoS Operational safety criteria

- **A – Safety and mission critical operations.**
- The output or function directly affects the immediate performance of the vehicle (eg an aircraft flight control) and personnel during 'Danger Close' operations.
- Use of 'Master Arm' and 'Kill' Flight Termination System Switches used.
- 10E-6 / 16.5 db margins are to be used. Independent verification of functions normal.

- **B – Mission critical – safety affected operations.**
- The output or function is relied upon by the crew for the safe operation of the vehicle and operations. *The failure of an Operational Category B system could possibly result in injury to personnel or damage to property if crew are unaware of the problem and fail to execute normal recovery procedures.*
- Category most likely for targeting information assurance and for which weapons and fire control system safety measures are designed to.
- Fully engineered with Configuration Mgt / Quality Assurance and 'Interventionist' Maintenance policies typical.
- 10E-3 / 6 db margins are to be used.

Armament FoS Operational Safety criteria

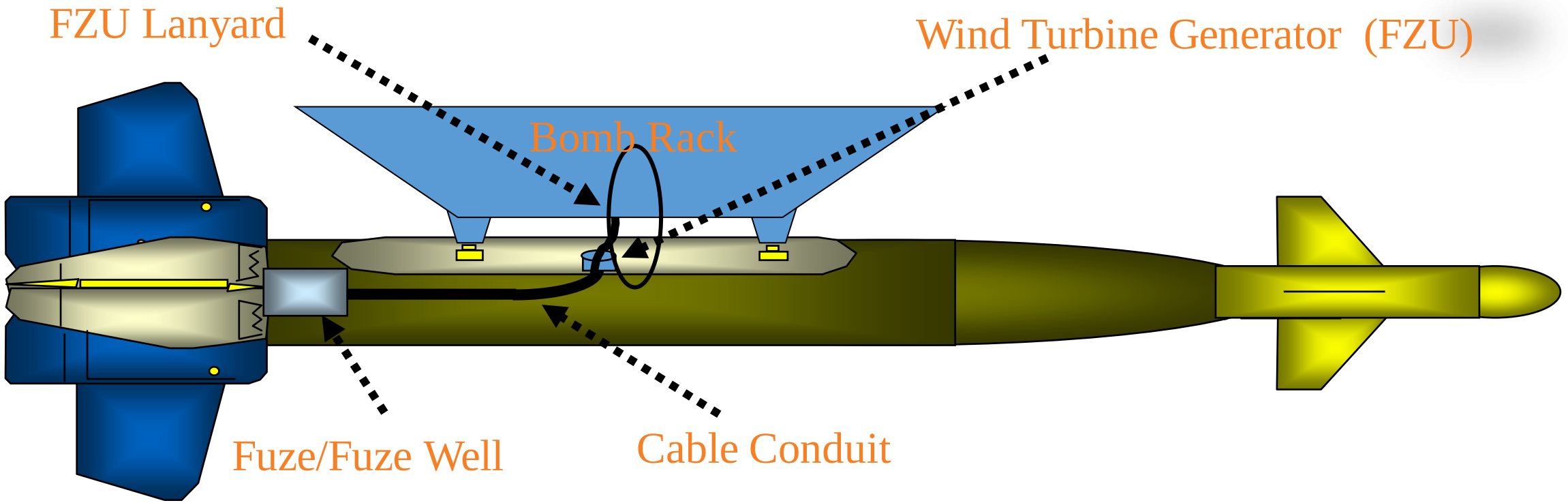
- **C – mission affected/advisory – ‘non-safety critical’ operations.**
- Fully ‘Engineered Systems’ with CM and QA.
- The output or function is used by the crew for advisory or non-safety purposes only.
- *The failure of a Operational Category C system would not be expected to result in injury to personnel or damage to property.*
- ‘On condition maintenance’ may still be rampant in ICT segments and configuration management of software is ab initio.

- ***D* - Prototype System / SoS that is mission affected / advisory - not yet ‘safety critical’.**
- Such systems may not have been fully engineered across the armament FoS without certification/qualification and properly engineered and full configuration management in place ‘as yet’. ‘On condition maintenance’ may be rampant in hardware and ICT segments and software is from a laboratory or ground test environment.

- ***E – Uncategorised!***
- Interesting category!

- Based on ADF Operational Aviation criterion and Aircrew proficiency categorisation.

Safe & Arm: air to surface



RULE: Two independent environments indicating proper launch

(1) Release from the airplane (lanyard pull on FZU)

(2) Windstream velocity great enough to turn FZU turbine

- Fuze arms after getting FZU-generated power

Kill Chain (F2T2EA)

Find
- Detect existence of target
- Intel Collection
- Area/Volume Search with Sensors

Fix
- Focus sensors to generate reliable detections
- Locate in space and time
- Identify

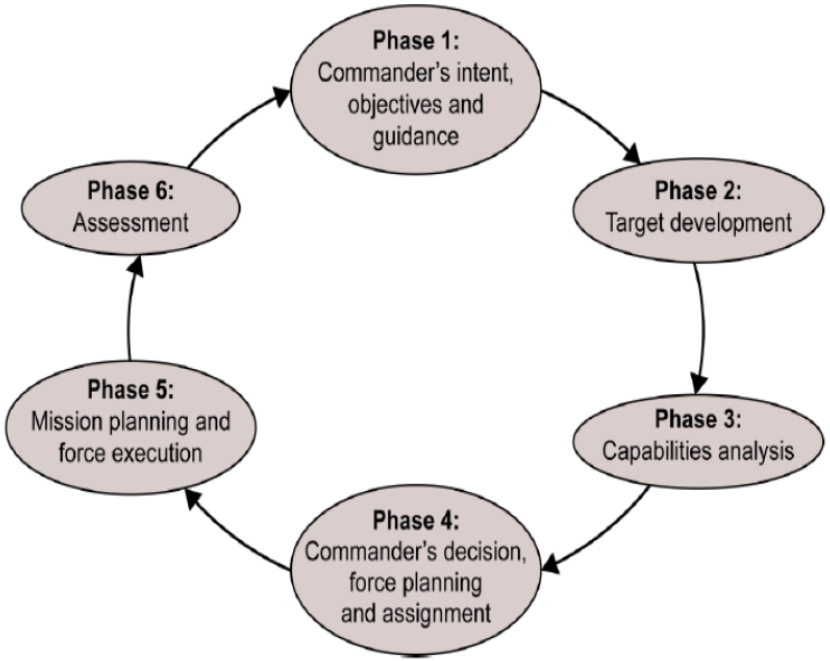
Track
- Hold tracking sensor on target
- Calculate accurate target position and motion
- Determine available time to engage

Target
- Select attack vector and weaponeer target
- Determine whether to engage
- Clear target area, deconflict, assess engagement risk

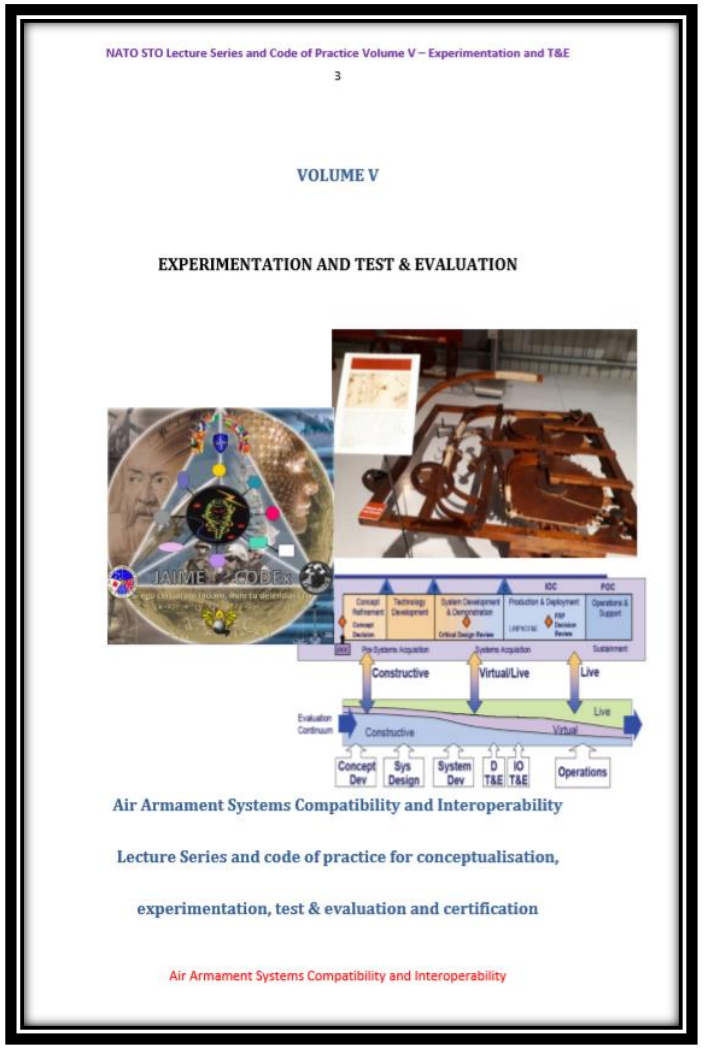
Engage
- Order Engagement
- Launch attack through chosen vector
- Provide track update, and monitor/manage attacker

Assess
- Assess target damage
- Assess resulting Effect
- Decide whether to Re-attack

Commercially-led data analysis technologies and simulation capabilities offer the potential to provide significant operational decision support



NATO Air Armament Lecture Series



V. Experimentation and T&E

Traditionally, T&E is the process by which a system is

compared against technical or operational criteria through ***test***

and the results are

evaluated to assess performance against agreed criteria (including design, performance, supportability) to

determine the system's fitness for purpose.

Test & Experimentation

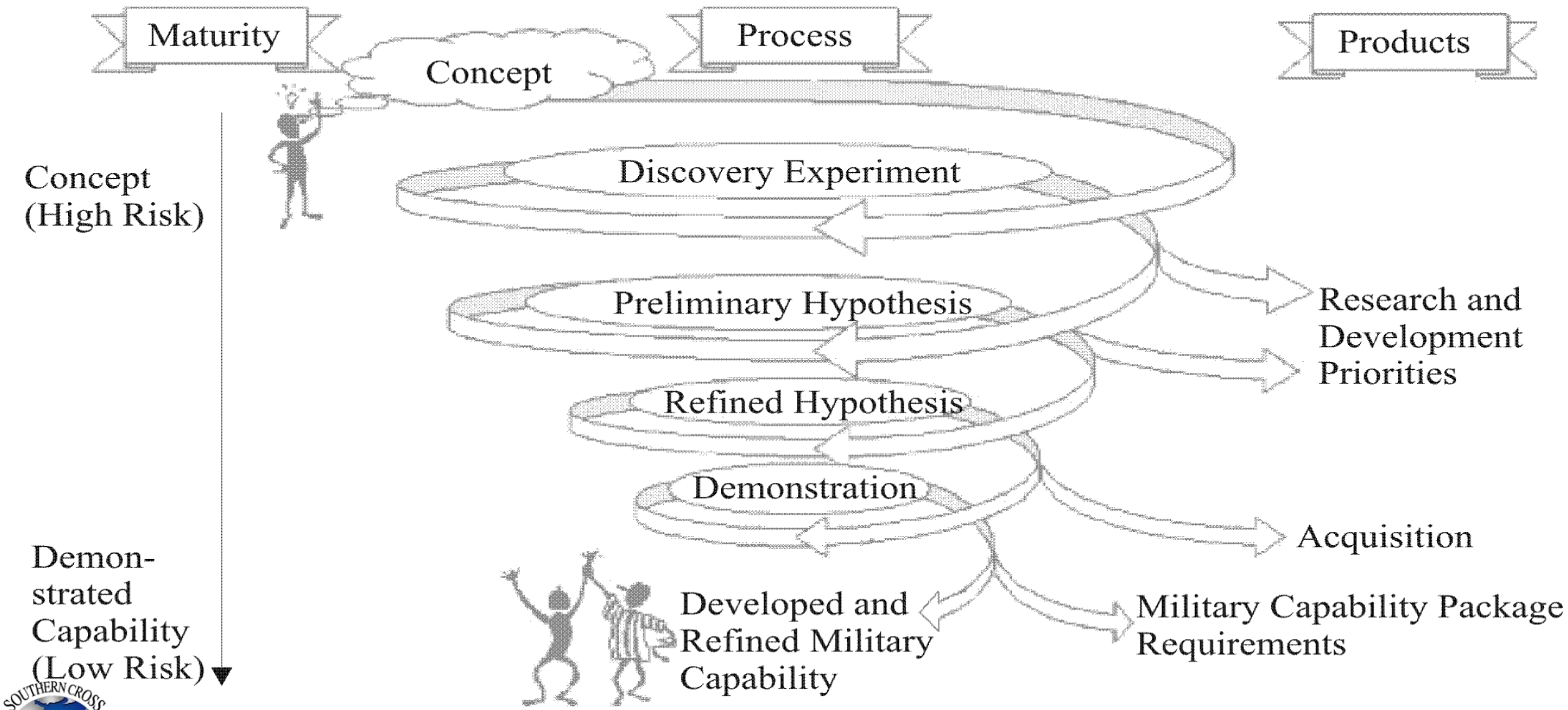
Sorting Through Terminology

A = New Sensor
B = Detections

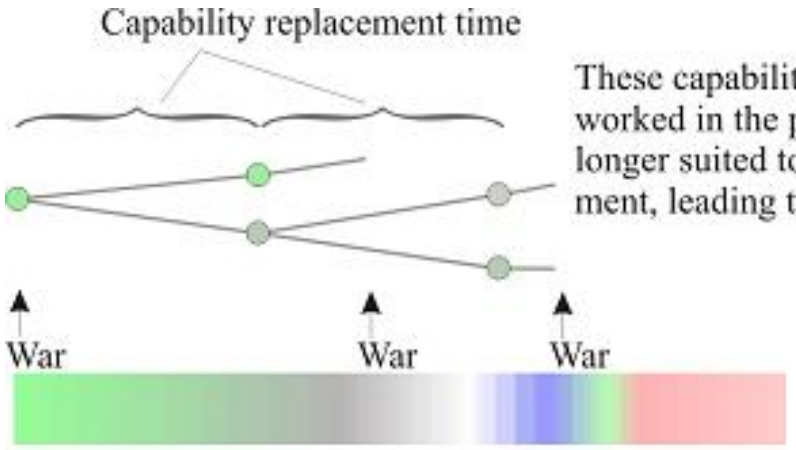
Event	Goal Stimulating Event	Purpose of Event
Training	**Practice on A to get B.**	Operation to **assist** entity in acquiring ability to do A.
Demonstration	**Show how A works to produce B.**	Operation to **show/explain** how A works.
Test	**Determine if A works** (produces B). •How effective is A? •Can operator/unit do A?	Operation to **confirm the quality of A.**
Experiment	**Determine if A solves B.** •Is A related to B? •How much does A affect B? •Did something else produce B?	Operation to **discover a causal relationship between B and something else, A.**

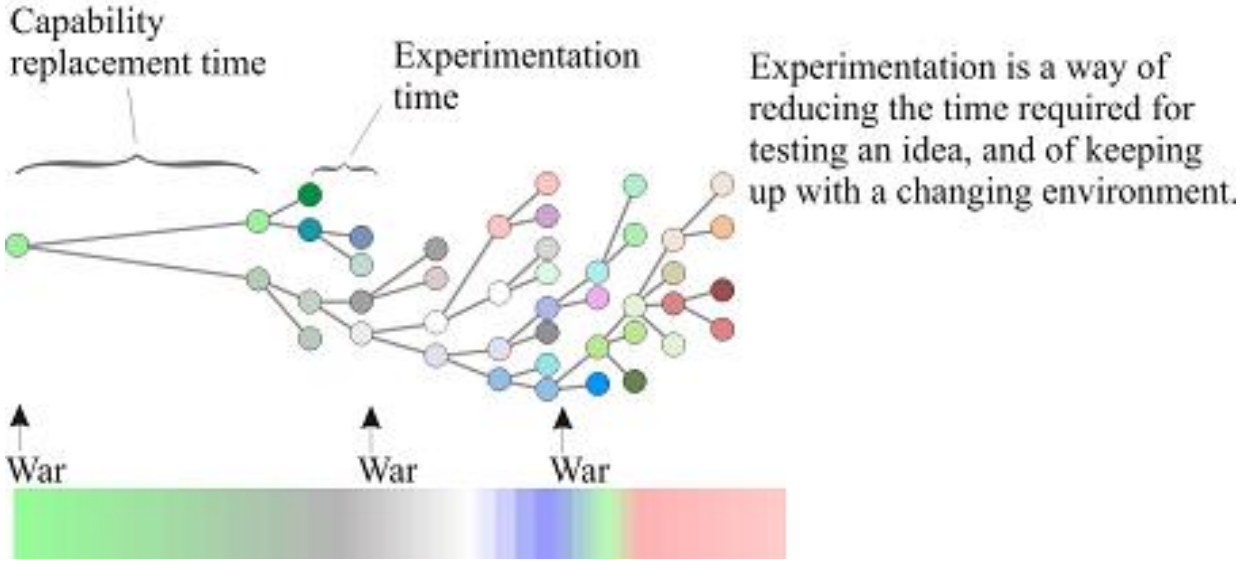
Figure 43 Comparison: terminology for training, demonstration, tests and experimentation

Why Experiment: the Big Picture



Why Experiment (Yang) & Igor's View (Yin)





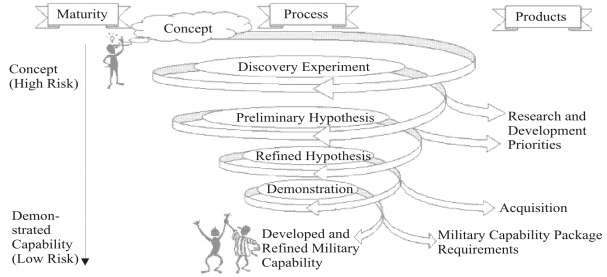
At that time [in 1909] the chief engineer was almost always the chief test pilot as well.

That had the fortunate result of eliminating poor engineering early in aviation.

Igor Sikorsky

Knight, Dr., 2004, *An Evolutionary Framework for Experimental Innovation*, ADF Experimentation Symposium, DSTO, June 2004

STANAG 7068 Ed 3 +



- TTCP GUIDe to Experimentation history
- MIL-HDBK-1763 aircraft stores compatibility history
- Armament Systems Compatibility
- Consistent view of framework for use of previous work, for compatibility/safety and Analogy/Similarity & Risk Mgt
- Aim – armament SoS integrated mission environment:
  - Test 100 Series M&S
  - Test 200 Series – Laboratory Ground Tests
  - Test 300 – Open Air Ground Tests
  - Test 400 Series – Open Air / Flight / Sub-surface Tests
  - Test 500 Series – In-Service & Joint Operations continuously monitoring effectiveness and performance!

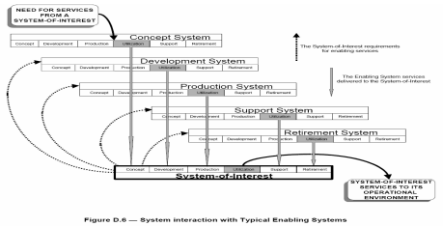
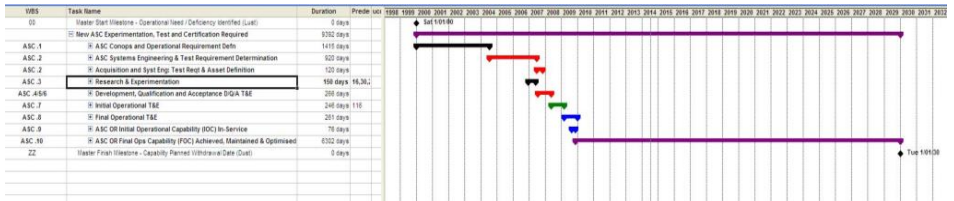
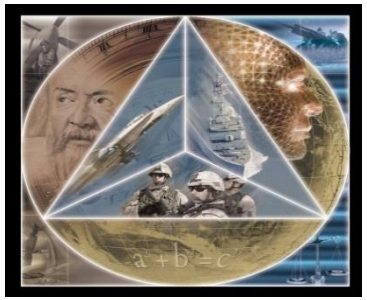


Figure D.6 — System interaction with Typical Enabling Systems



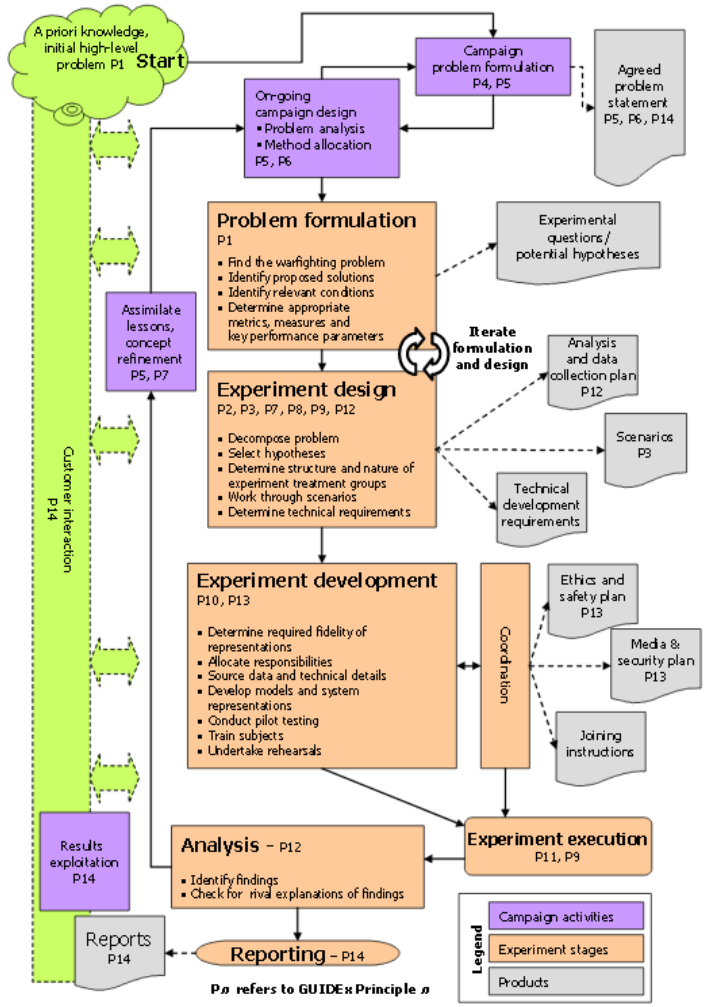
TTCP GUIDEx

- TTCP – NATO STO Action Group
- Refinement, focus, extrapolation
- Internal and External Validity
  - Identifying Cause and Effect
  - Relating to the Real World
- Break-it
  - Stressful Scenarios
  - Capable and Intelligent Enemy
- Communicate
- Work being used by NATO STO M&S Group



TTCP GUIDEx

- Robust experimentation methods from the sciences can be adapted and applied to military experimentation and will provide the basis for advancements in military effectiveness in the transformation process.
- 14 Principles & good experimental design practices to counter the '21 threats'.
- Who should read the GUIDEx:
  - Those who ask force capability questions and act on the answers.
  - Those who decide how the force capability question is to be addressed and what methods are to be used.
  - Those who design, execute, and interpret defense warfighting experiments.
  - Those engaged in Operational Test & Evaluation (OT&E).
  - All those for whom experimentation matters!



Test or Experimentation

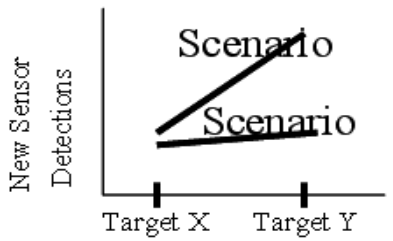
Test or Experiment?

Different questions ,
but similar design and execution
(can answer both questions in one design)

New Sensor A

Number of Targets Detected B

	Target X	Target Y
Scenario	12 11	16 15
Scenario	11 10	12 11



Test

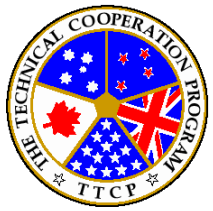
Issue: **Is the New Sensor effective?**
Criteria: Must detect more than 14 targets
Answer: **Partially Effective--only against Target Y in Scenario 1**

Experiment

Issue: **Does the type of target affect the ability of the New Sensor to detect targets?**
Hypothesis: **If** Scenario 1, **then** the New Sensor will detect more targets.
Answer: **Target Y impacts Sensor capability in Scenario 1 only.**

Figure 44 Contrasting tests and experiments

Test or Experiment



Test during experiment

Types of system representations in defense experiments

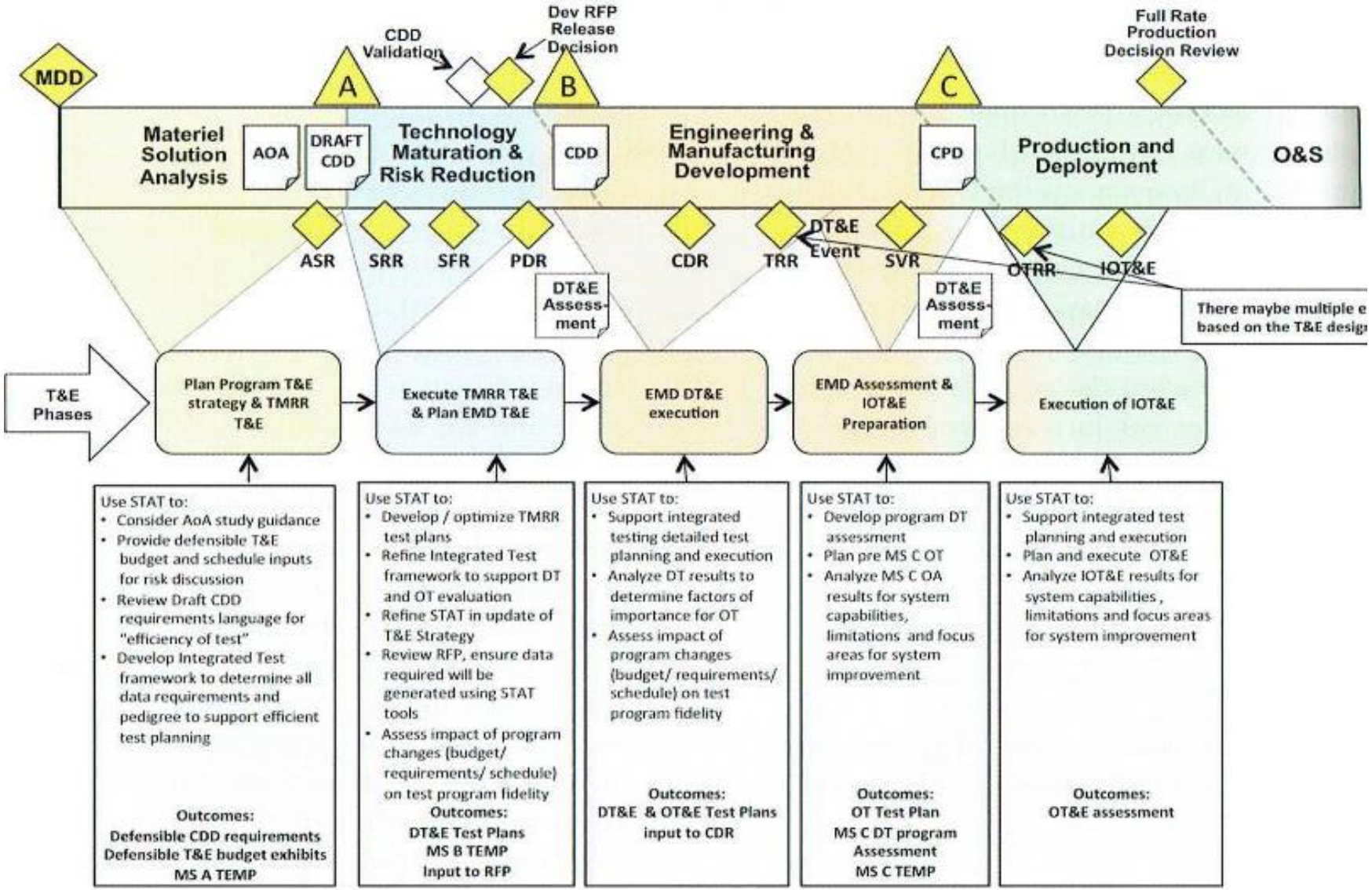
	Wargame Emulated	Constructive Simulated	Virtual Simulator	Field Prototype
Ability to assess system under usual "Test Conditions"				
•Under specified conditions	+	+	+	o
•Quantitative outcome data	-	++	++	+
•Sufficient diagnostic data	-	+	++	o
•Sufficient repetitions	-	+++	+	-
Ability to assess system characteristics?				
•contribute to mission success	✓	✓	✓	✓
•SW modules		✓	✓	✓
•interfaces/interoperability			✓	✓
•Functionality				✓
•Reliability				✓

Limited to same scenario and conditions in experiment

Legend: o N/A, - not possible, + potential, ++ more, ++ even more potential

Figure 45 Can one test during an experiment?

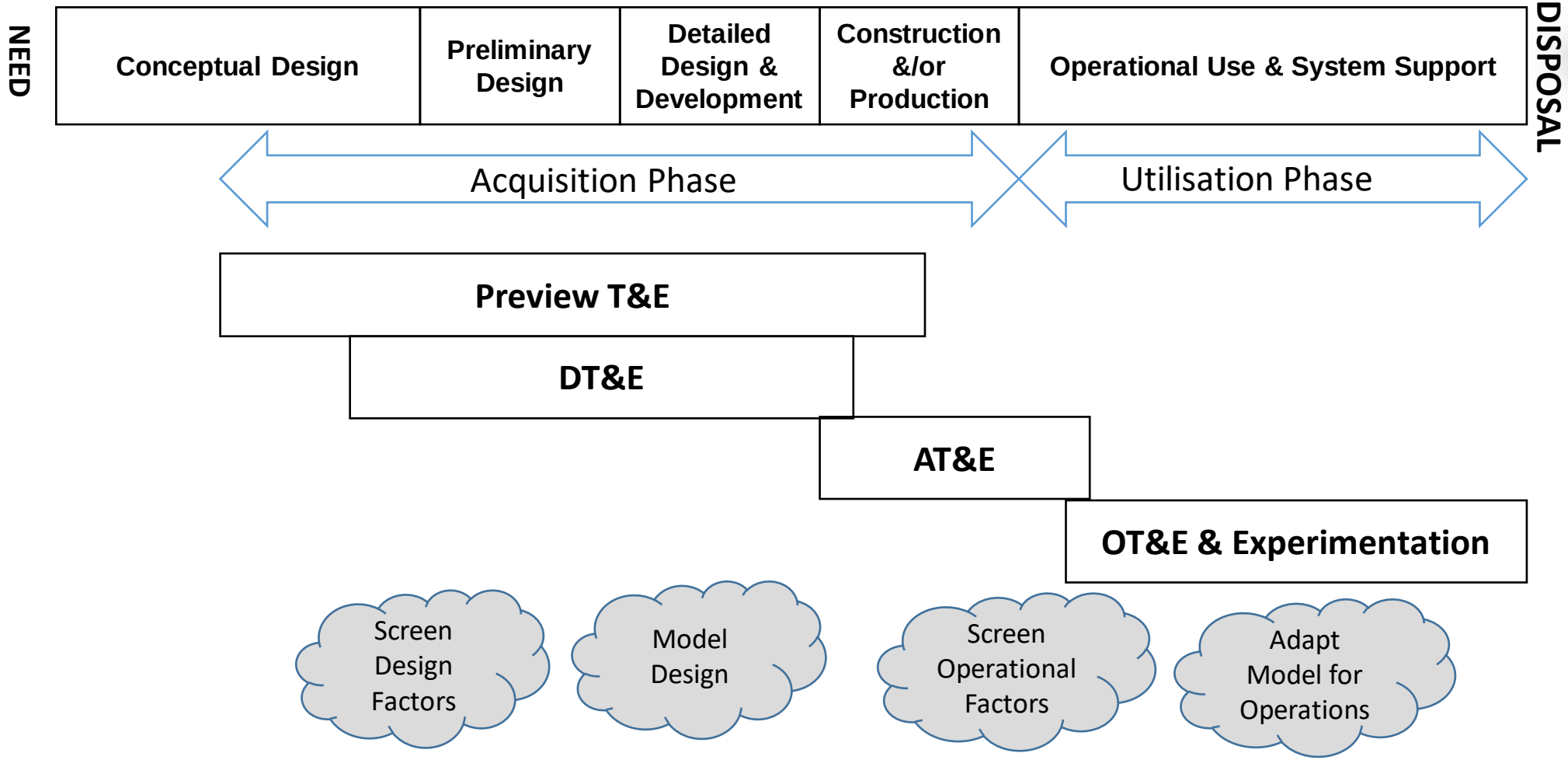
US System Life Cycle (simplified)



Murphy et. al. (2015)
ITEA Journal

Via Joiner (2018)

System Life Cycle

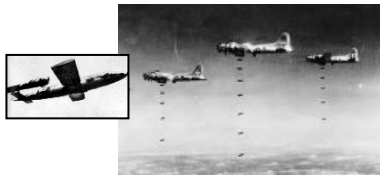


Adapted from Joiner (2018)

Epochs of war and Air Armament - The Capability Transformation Story continues

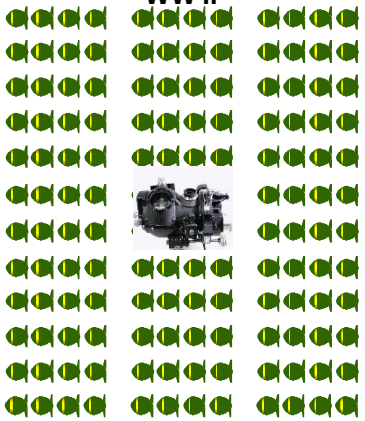
Epoch 0 / I

Epoch II



1500 Bomber sorties
9000 bombs 250lbs

3300 ft CEP
One 60' x 100' Area Target
WW II

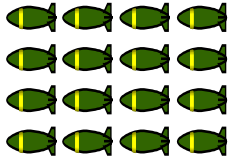


Per Area Target

Epoch III

30 Fighter / Attack / Helo sorties
176 bombs 500lbs

400 ft CEP
One Target
Yom Kippur / Vietnam



Per Target

Epoch IV

1 F-111/F-117 sortie
4/2 bombs 2000lbs
10 ft CEP

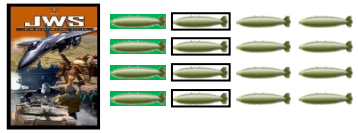
Four/Two Targets per Sortie
Gulf War I

Targets per Pass

Epoch V

1 F-22 sortie
8 bombs 500lbs
5 ft CEP
8 Targets per Pass

4 EA / 4 Cyber
All Weather

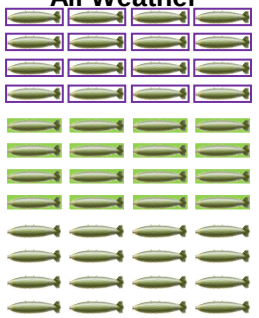


Targets

Epoch VI

Enterprise

1 Swarm sortie
Cyber/B-21/RPA/DE
16 250lbs & 16 Bots
1 ft CEP
16 Targets per Pass
512+ Cyber / 128+ EA
All Weather



Targets

Network-enabled

Accuracy and precision

Accuracy: Circular Error Probable – "50% of Bombs"

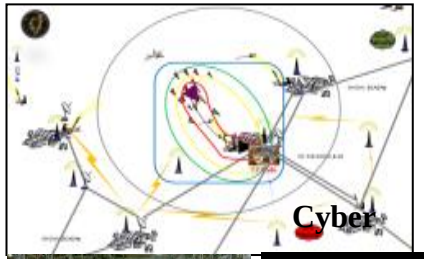
'Revolutionary' Technologies
Hand-held to Aircraft EO dispensing
Impact Fuzes / LOS to Norden Bomb Sight
First Guided Within Visual/Radar AI Seekers

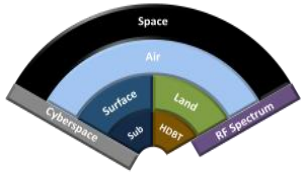
'Revolutionary' Technologies
Analog IR/RF SA AI Seekers / ISR
Mechanical/Prox Fuze Options with
SAFE ARM / MIL-STD-1763

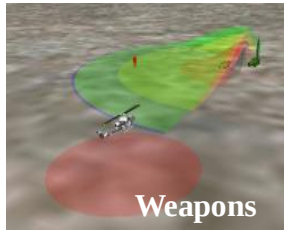
'Revolutionary' Technologies
Digital Avionic Systems / Laser SA GW /
MIL-STD-1760 / EW/EA+ GPS Aided /
Electronic Fuzes / Low Observables

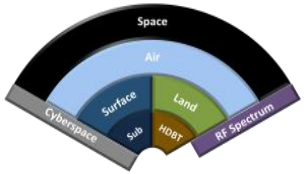
'Revolutionary' Technologies
Info Age Network enabled SoS / Cyber
and Enterprise / Link 16/MADL/ ANI/AGI
Synthetical: Live Virtual Constructive

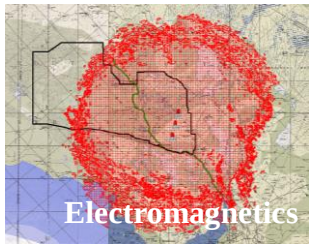
Armament 6.0 +

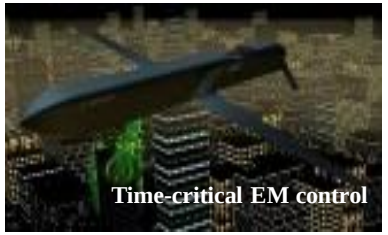












Armament Families of SoS include:
Weapons, directed energy, electronic and information warfare and their intelligent/autonomous systems.

- Information Assurance (Defence) and Offensive SoS Control & ANI/AGI RoSI Exploitation
- Kinetic weapons SoS Control and ANI/AGI RoSI Exploitation
- AI
- Kinetic & Non-kinetic DE SoS Control & ANI/AGI RoSI Exploitation
- EM Spectrum & EW/EMS SoS Control & ANI/AGI RoSI Exploitation

'Armament 6.0+'
Threat and Effects Assessment, Synchronisation and Reporting:

Joint Armament FoS Integrated Mission Environment: Experimentation, Rehearsal, Test, Training, Operations, Replay & Reporting

Future

- Sensors & C2
- Open Mission Systems / Universal Armament Interfaces

Hypothesis: Probability α and Confidence β

		True State of Nature	
		H_0	H_1
Conclusion Drawn	H_0	Conclusion is correct	Conclusion results in a Type II error
	H_1	Conclusion results in a Type I error	Conclusion is correct

- Defendant Innocent
- Capability is Good

- Guilty Defendant let go free
- False Negative - No Alarms but should have
- Failing to accept Good Capability
- β

- Defendant Innocent but found Guilty
- False Positive / Alarms
- Accepting Bad Capability as Good
- α = 0.1, 0.05, 0.01 depending on criticality

- Defendant is Guilty
- Capability is Bad

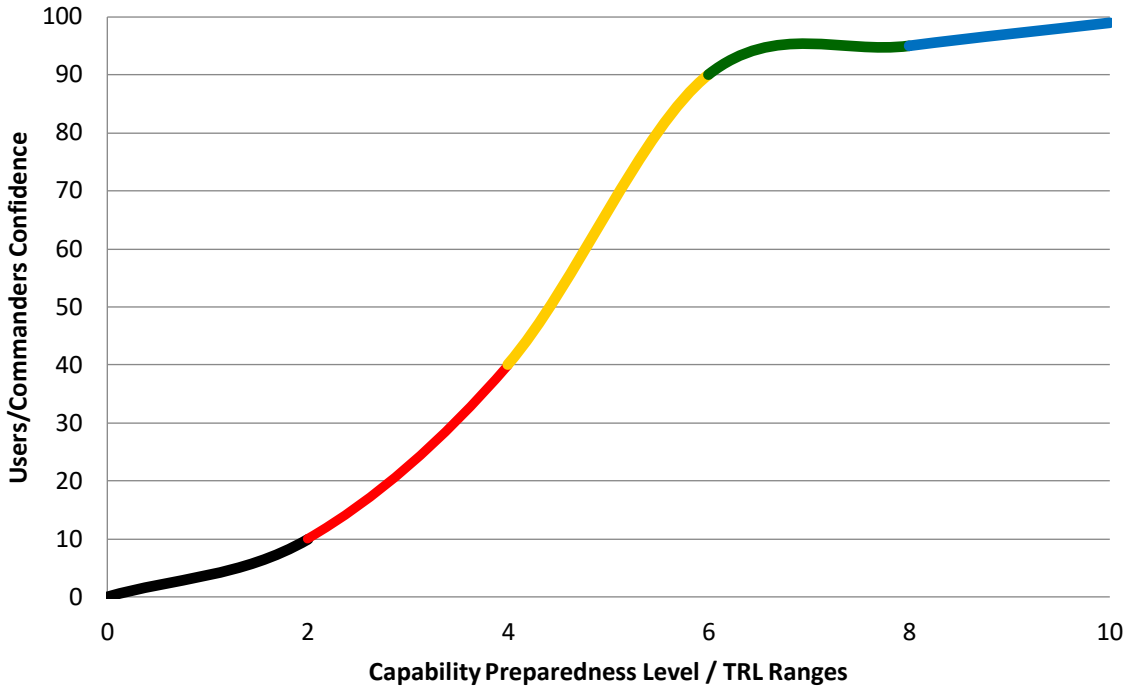
‘Black Swans’ and ‘Fat-tails’

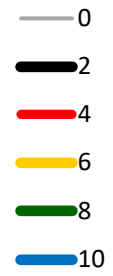
Black Swans are those highly improbable circumstances outside of previous experience – in Europe there were only White Swans …

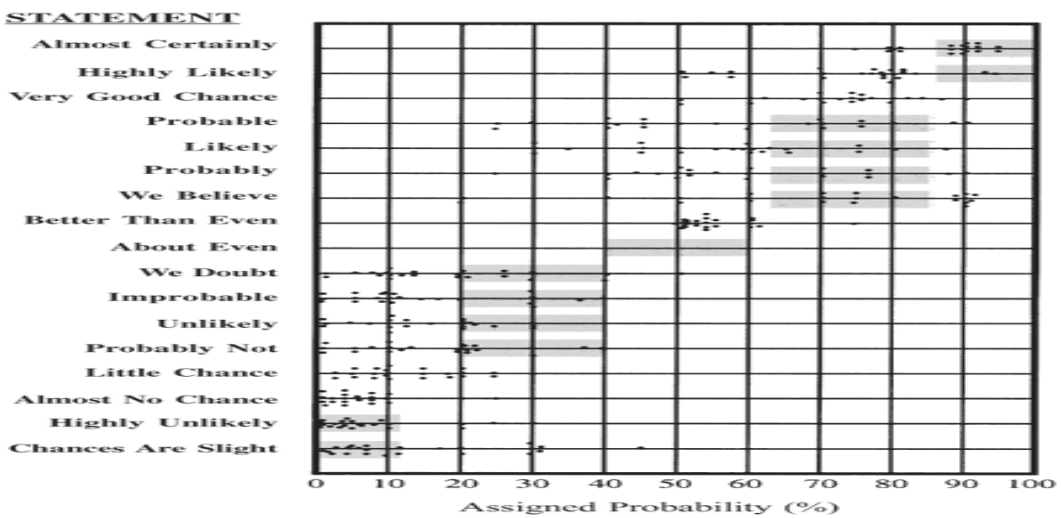
- **Extremistan** – where the total is conceivably impacted by a single observation.
- **Mediocristan** – where the mediocre are king – grounded by the Gaussian Bell curve where no single observation can meaningfully affect the aggregate!
- **Asymmetry in reward on those who prevent and those who cure**
  - **Brooklyn Tony vs Dr John** – 99 flips of a coin are heads what is the probability for the next flip ?
  - **Checking Assumptions / Confirmation Error**
  - **Fallacy of silent evidence** – we do not see the full historical story
  - **Fooled by Randomness** – confusion between luck and randomness / ‘Fat Tails’ of the ‘small-world’
  - **Future Blindness** – natural inability to see properties of the future
  - **Lottery Ticket Fallacy** – naive analogy – they are not scalable
  - **Narrative Fallacy / Retrospective Distortion / Round-trio Fallacy / Scandal of Prediction**
  - **Scorn of the Abstract** – The death of one child is a tragedy; the death of a million is a statistic
  - **Statistical Regress Argument**

Taleb, Prof N, 2007, *The Black Swan – The impact of the Highly Improbable*, Penguin Books

User and Commander Confidence Levels







The Growing Challenge

Distributed Test:

- Linking geographically separated LVC sites and capabilities together in a distributed environment to support T&E of a system or SOS in a Joint Air/Land/Maritime/Space/Cyberspace integrated environment

Use When:

- Requirement to exchange data w/in your system or w/in SoS
- Address SoS interoperability issues early in the acq process
- Lack adequate numbers of systems under test for live testing
- Lack adequate numbers/resources for supporting systems, C4ISR assets, etc.
- Lack adequate threat types, fidelity, and density in realistic numbers at realistic ranges for live testing

NOT ROUTINELY DONE IN THEATRE

U.S. DoD I3 assurance Initiatives

- Initiative 1 - Augmenting Operational Exercises with Formal Experimentation
- Initiative 2 - Integration System Program Offices and New Certifications

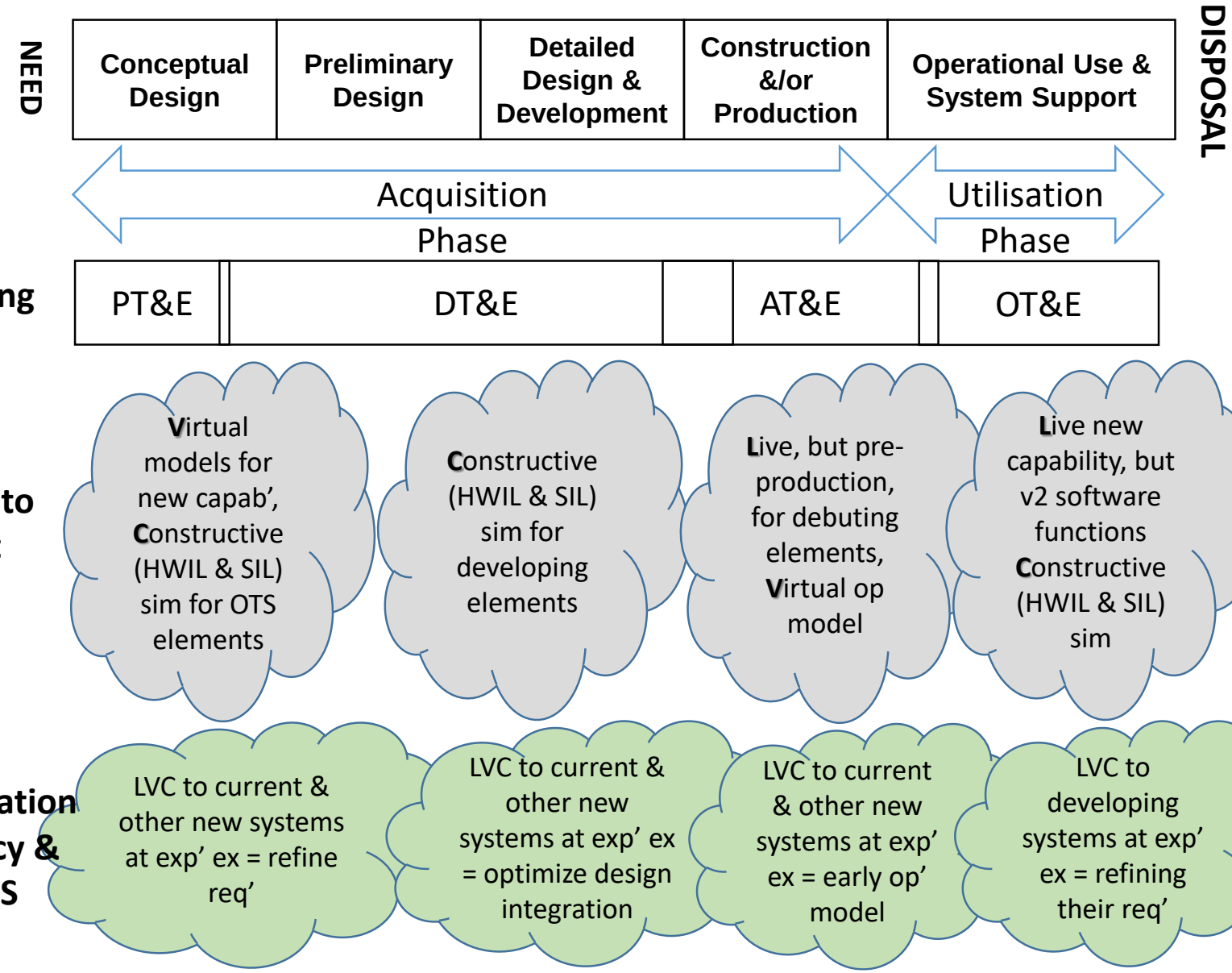
- Initiative 3 - Enhanced T&E Regime – Earlier, Evidence-based Rigour and
- Innovation: Test Smart not Test Often!

- Initiative 4 - T&E Network Infrastructure

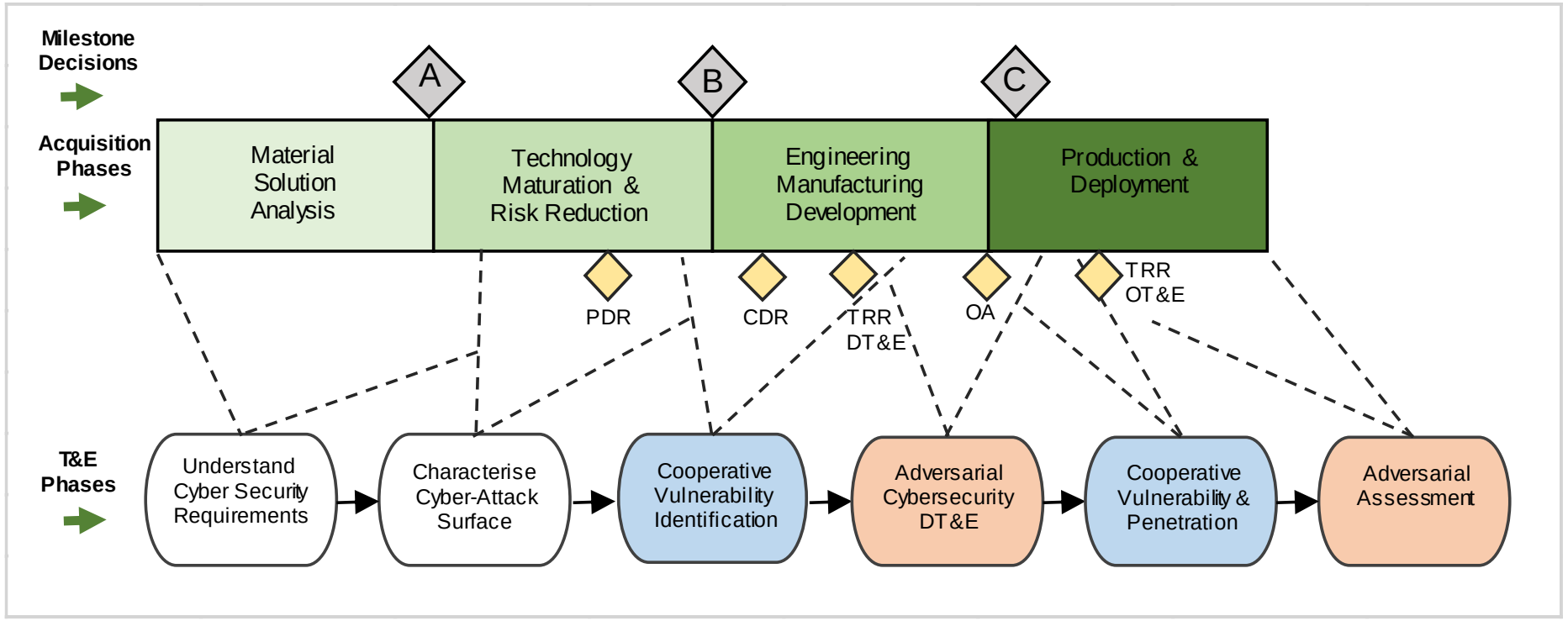
- Initiative 5 - Cybersecurity Protection Plans and T&E

- Initiative 6 - Permeating these U.S. Initiatives into Industry

How US initiative works



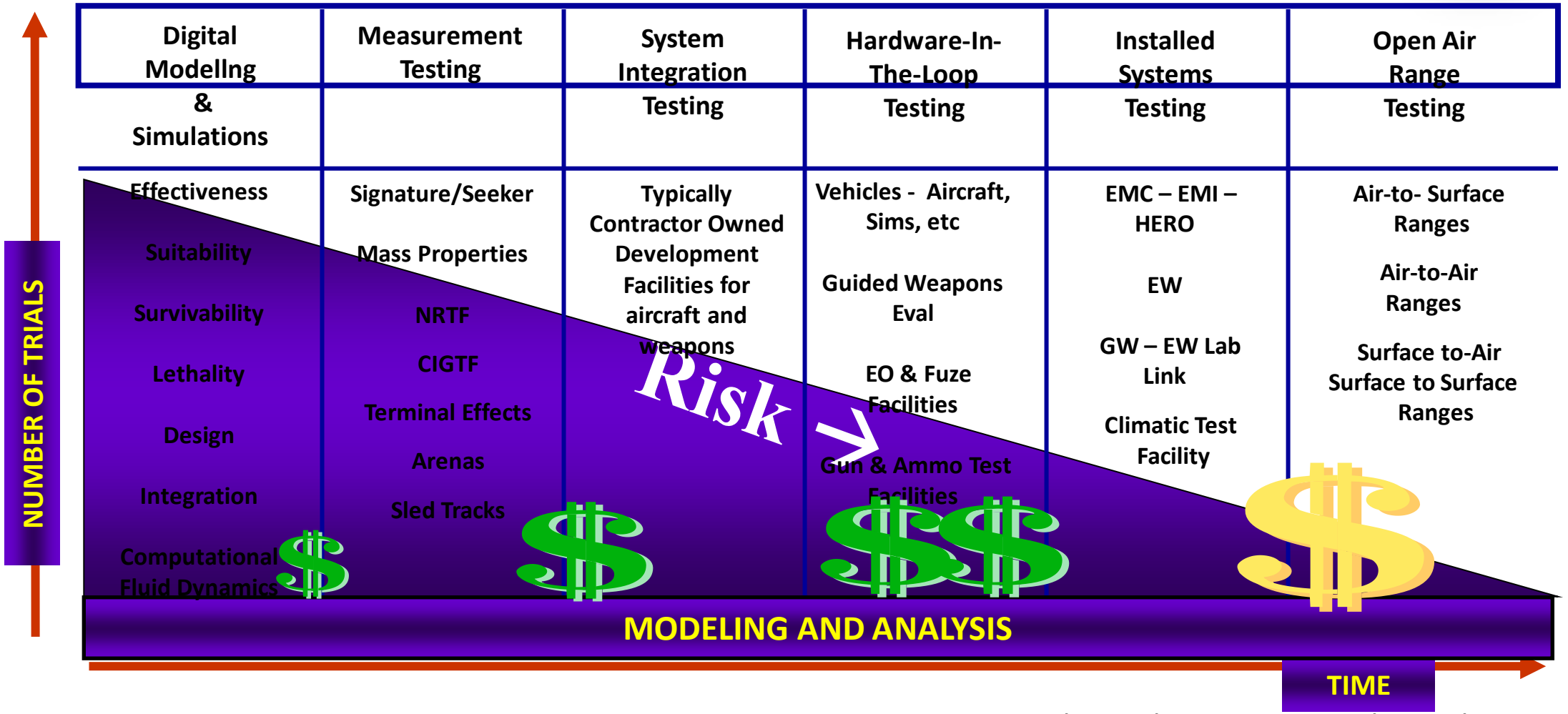
U.S. Cyber T&E Steps on U.S. DoD Lifecycles



Only works by using JIOR (underpinned by JMETC etc) & the NCR

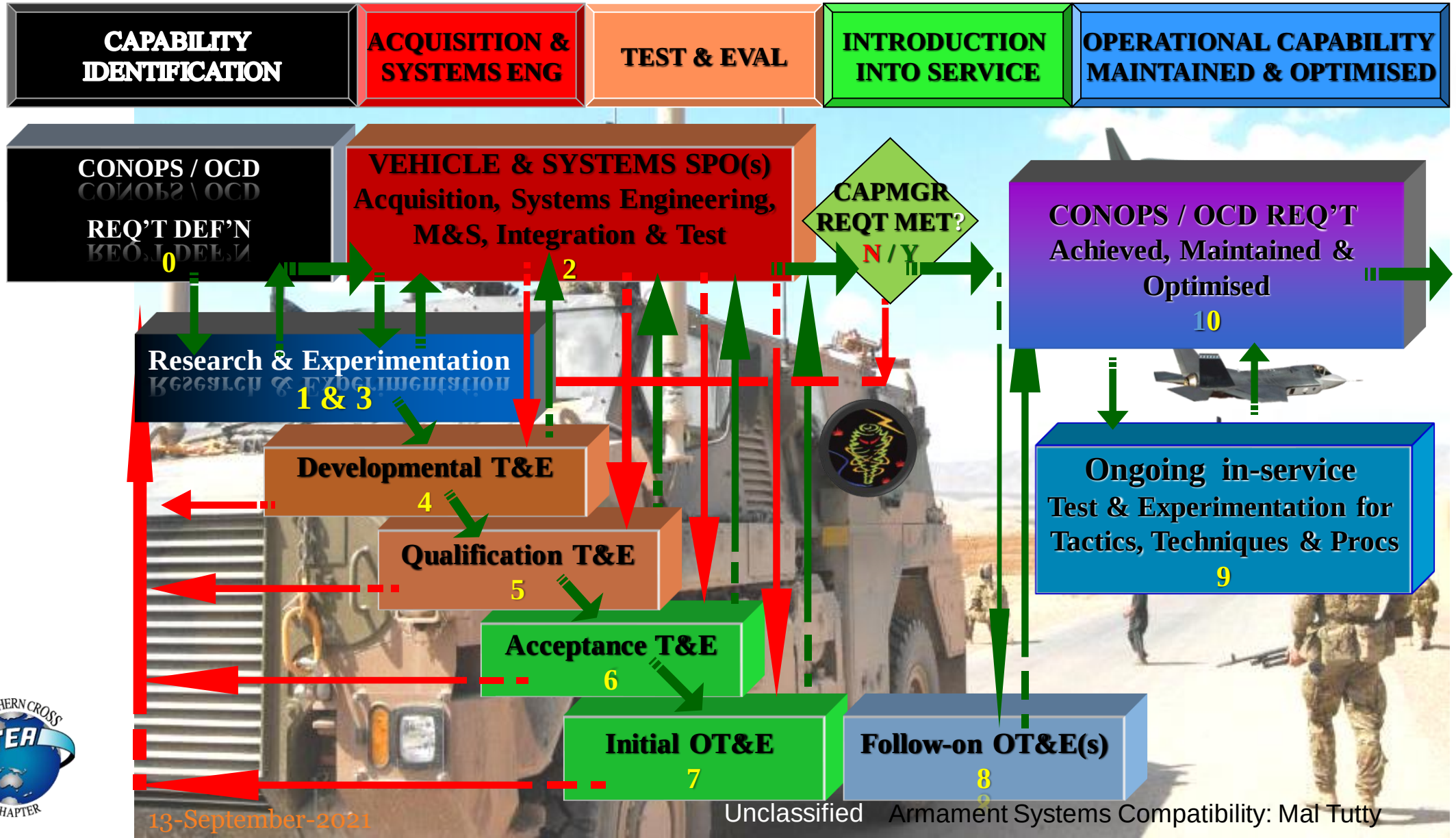
From Brown et al. (2015)

Disciplined Test Approach for kinetic & non-kinetic effects



Developed from Tutty USAF TPS and FWS presentations (1990) and Arnold (2010)

V.1 Innovative Warfighting Experimentation and T&E Mindset



Initiative 2 - Integration System Program Offices and New Certifications

Organisational alignment to System-Of-Systems view of the world

Portfolios, Programs & Projects (P3O) with I3 assurance accountabilities

Certifications like: tactical data links, cybersecurity, joint fires (JTAC) etc

•**System of systems** is a collection of task-oriented or dedicated **systems** that pool their resources and capabilities together to create a new, more complex **system** which offers more functionality and performance than simply the sum of the constituent **systems**.

Initiative 3 - Enhanced T&E Regime

- ***Earlier, Evidence-based Rigour and Innovation: Test Smart not Test Often!***

- Use of mandatory test measures under-pinned by rigorous & highly efficient new test design & test analysis techniques, has removed much of the scope in the U.S. DoD for '*decision by conjecture and influence*' or what is also commonly called '*paper-based analyses*'.
- Where decision-making still occurs without testing (to include modelling on VV&A models), the '*name & shame*' of independent annual reports to Congress by Director OT&E calls such practices out to Congress to help end them.
- No such legal or *name & shame* processes exist within the Australian DoD to call out acquisition practices that are not based on experimentation, test & accredited modelling, leaving it to the Parliament (Australian Senate, 2012 & 2016; Australian Parliament, 2016), ANAO (2002, 2013, 2016).

Capability Preparedness Levels:

Operations Criticality Levels & Mission Confidence γ

		A	B	C	γ
0	No identified capability, basic principles conceptualised and studies initiated. Capability development being proposed.	1σ	50	20	50
1	Basic principle observed and reported. Studies or initial investigations undertaken. Capability development initiated.				
2	Technology concept and/or application formulated. Potential applications have been identified.	2σ	1σ	50	70
3	Analytical and experimental critical function and/or characteristic proof of concept. R&D has been initiated, work towards validating the concept done. Capability is Managed.				
4	Component and/or breadboard validation in lab environment. Basic elements of the Capability (SoS/Major system/product) have been integrated to show they will work.	3σ	2σ	1σ	90
5	Capability SoS/Major systems and/or components/breadboard validation in relevant environment. A higher fidelity validation of the Capability in a realistic environment. Capability is Defined.				
6	**Capability SoS model or prototype demonstration in a relevant / realistic environment.**	4σ	3σ	2σ	95
7	**Capability SoS prototype demonstration in operational environment with representative personnel and C2 iaw Conops.** 'Production' can now commence. Capability is being Quantifiably Managed.				
8	Actual Capability SoS completed and mission qualified through test and demonstration with operational personnel and C2. **Actual Capability SoS has been successfully tested, qualified and certified iaw Conops.**	5σ	4σ	3σ	99
9	Actual Capability SoS proven through successful mission operations with operational personnel and logistics support. **Actual Capability SoS has been successfully fielded.**	6σ	> 4σ	> 3σ	
10	Actual Capability SoS has been found to be operationally effective, suitable and sustainable in successful real-world network- enabled operations with other identified SoS as a FoS. Capability is being Optimised.				

Evidenced based Capabilities: Combinatorial ET&E

Complex Adaptive – Force Level Capability Integrated Armament Mission Environment

-Seeking understanding and adaptability of FoS Emergent properties at Buyers Risk!

Complex & adaptive - Family of SoS

- Lots of unknowns and unknowables
- SoS - Mission Level PICs & 'ilities' proven (sic)
- "FoS" 'ilities' are rare but WITH "Mission Expectations" -> adaptability?
- Safety critical output/function – autonomy
- SOA / OSA predominate – but not common
- Continuous measurement of COI/MOS / E / P

Complex Acquisitions of FoS

- ✓ Synchronisation
- ✓ Proven Config/Role/Environment (CRE) for EACH SoS in the FoS
- ✓ "Rub Points" & hence all 'Key Interfaces' are identified
- ✓ 'PICS' Critical

V&V & Implications

- ❖ QA
- ❖ Structured Experimentation/T&E of Systems for Joint Force Level End-Effects
- ❖ Exp/T&E of SoS boundaries - Sellers Risk
- ❖ Exp/T&E with FoS continuously - Buyers Risk

Complicated - Bounded 'Systems Engineered' SoS / Capabilities

- Pick and maintain a model with agreed standards to measure and control system, or any SoS emergent properties, at Sellers Risk

Complicated - Systems of Systems

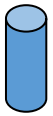
- Few unknowns unknowns
- Bounded 'Things' with controlled connectivity
- Engagement 'ilities' & PICs specified & known but no "prescriptive" Mission Level COI/MOS / E / P
- Mission Critical – safe operations
- Users need Training / TTPs developed on the new and/or unique Tools

Typical Complicated Acquisitions

- ✓ Minimal coordination
- ✓ Config/Role/Environment (CRE)
- ✓ 'Isolated' with few connections
- ✓ Sub-systems can be simply acquired
- ✓ 'PICS' Important

V&V & Implications

- ❖ QA
- ❖ Structured Exp/T&E to verify agreed 'ilities', Training & TTPs
- ❖ SoS – essentially Sellers Risk



Simple - 'Traditional' Engineered Systems

- Mature COTS & MOTS 'Heaven / Nirvana' at Buyers Risk

Straightforward 'Simple' Standalone Systems

- No Unknowns and No Unknowables!
- We're buying 'someones' 'working' system
- A Collection of Managed Things (ie. 5 apples & 3 bananas)
- No 'ilities' thought to be 'needed'
- No dedicated training needed (ie known by Users)
- Mission affected/advisory – non-safety critical
- Metaphysics Rules – & all characteristics 'known'

Simple Acquisitions

- ✓ Rare coordination
- ✓ Buy each 'thing' 'OTS'
- ✓ CIOG & DSG space critical in Australia
- ✓ What are 'PICS'?

V&V & Implications

- ❖ QA checks 'things' from Contractor
- ❖ Buyer must be aware of any/all cross dependencies with COTS/MOTS support and updates
- ❖ Systems – essentially Buyers Risk

Test & Experimentation

Sorting Through Terminology

A = New Sensor
B = Detections

Event	Goal Stimulating Event	Purpose of Event
Training	**Practice on A to get B.**	Operation to **assist** entity in acquiring ability to do A.
Demonstration	**Show how A works to produce B.**	Operation to **show/explain** how A works.
Test	**Determine if A works** (produces B). •How effective is A? •Can operator/unit do A?	Operation to **confirm the quality of A.**
Experiment	**Determine if A solves B.** •Is A related to B? •How much does A affect B? •Did something else produce B?	Operation to **discover a causal relationship between B and something else, A.**

Figure 43 Comparison: terminology for training, demonstration, tests and experimentation

Test during Experimentation

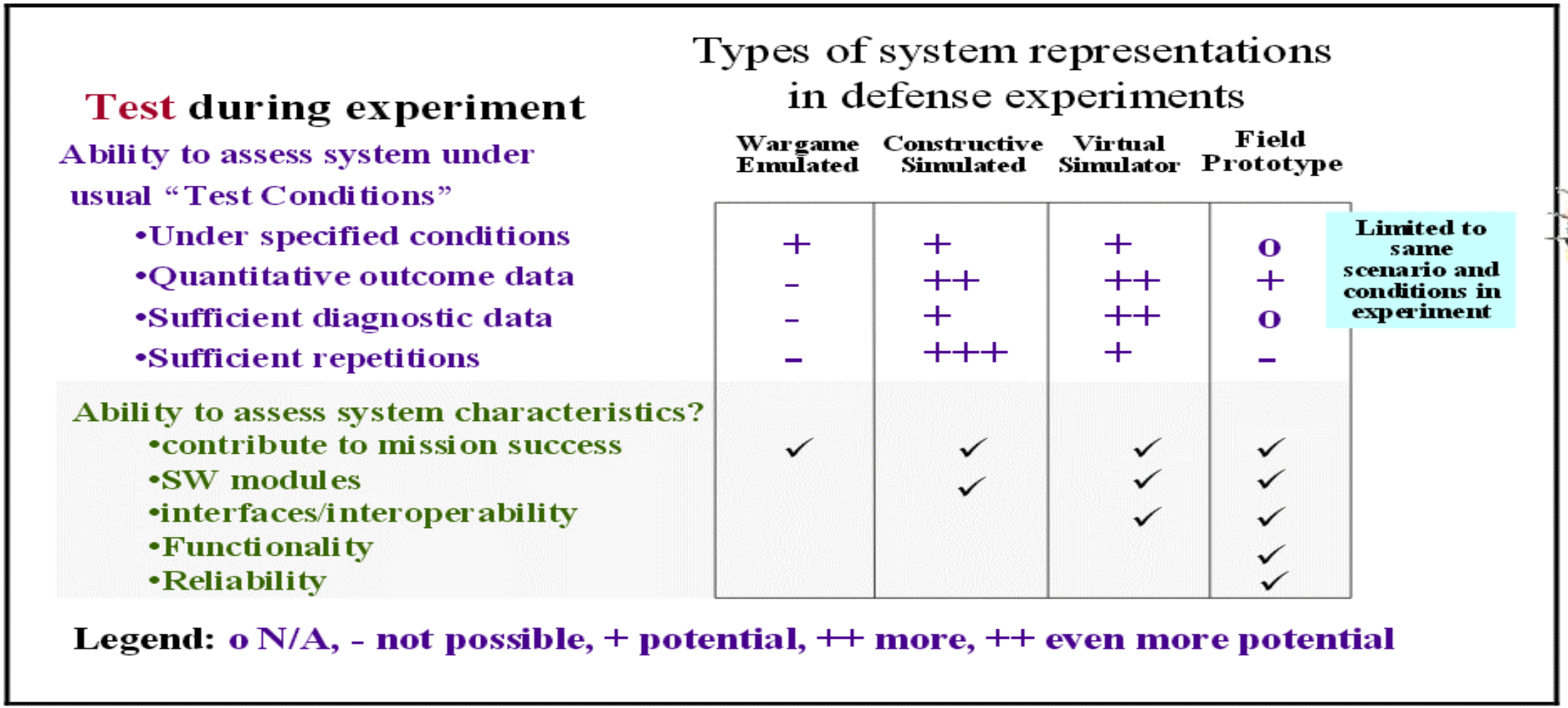
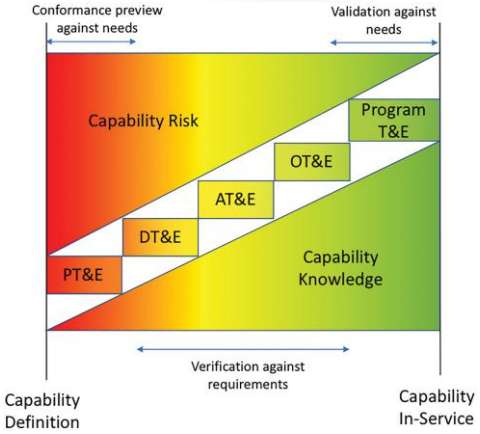


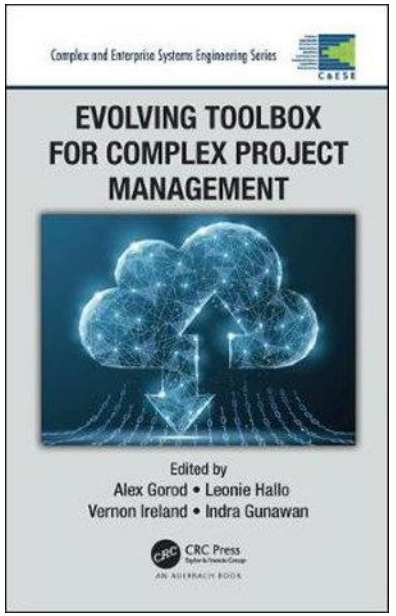
Figure 45 Can one test during an experiment?

T&E - influence design/project risk or fait accompli

T&E Toolset, Joiner, Efatmaneshnik and Tutty, *Evolving Toolbox for Complex Project Management,* Chapter 16, Edited by Gorod, Hallo, reland, and Gunawan, CRC Press, 2020.



- ✓ Strategy for system adaptability and changeability - functions are designed in the software not hardware, and early usability test demonstration of software functionality, modelled if necessary, is key.
- ✓ ***The development of combinatorial test design and multifactor, multi-response test design techniques and their spread into industry has been phenomenal, with very significant improvements in testing rigour and efficiency.***
- ✓ Confidence limits fundamentally improve the board-level decision-making (Rucker, 2014) because they inform all stakeholders of risks in proceeding at each project gate with the sufficiency or adequacy of testing so far.
- ✓ A flexible P3O gating process should simply fund a short examination phase as part of the conclusion of tendering which involves *offer-definition* activities and Preview T&E. PMOs are incentivised in gate reviews to test early and often so as not to inherit an operational burden.
- ✓ ***All chances of the ideal three to five iterations for optimum efficiency are lost when users are not involved.***
- ✓ **Very** representative system and software integration labs are ***quintessential to maintaining complex systems with resilience***.
- ✓ Many allegedly complex project acquisitions have mixed maturity of systems and deliver into complex interconnected legacy system-of-systems and yet do so without modelling and simulation for the new systems or a representative test environment for the interconnected legacy system-of-systems. That is poor planning not true complexity.
- ✓ **The most insightful examination of the factors causing surprises in major Defence projects was project over-optimism leading repeatedly to a belief that capabilities were more mature than they turned out to be and that pre-contractual preview T&E was fundamental to de-risking.**

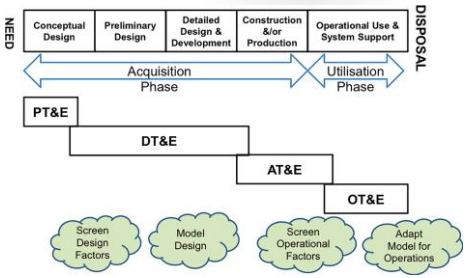


- ✓ **Preview T&E. T**hat DT&E conducted to evaluate the feasibility and performances of alternative capability options and identify risk areas prior to a final decision to acquire.

T&E - influence design/project risk or fait accompli

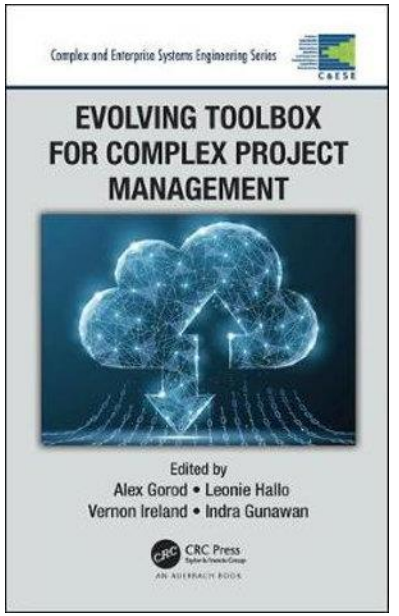
T&E Toolset, Joiner, Efatmaneshnik and Tutty, *Evolving Toolbox for Complex Project Management,* Chapter 16, Edited by Gorod, Hallo, reland, and Gunawan, CRC Press, 2020.

Interconnectedness and System-of-Systems Initiatives in today's complexity challenges:



- Creating more representative environments through federating SILs and HWILs to enable regular testing against new threats, new roles, new environments or developing capabilities.
- Creating annualised experimentation exercises with a battle rhythm of testing for I3 assurance using the infrastructure above. Such exercises are at the portfolio or program level
- Adopting use of six-sigma test design and analysis methodologies in acquisition policy, project milestone approval gates and acquisition staff competencies, so as to achieve the additional rigour and efficiency of these methods, especially in more complicated highfactor, highly connected (interdependent) environments.
- Adopting cybersecurity T&E in acquisition policy, infrastructure and staff competency targets
- Assigning dedicated acquisition staff, closely aligned with P3O staff, focused on the infrastructure, competencies and other wherewithal to give I3 assurance.

The threat of greater interconnectivity and information flow is countered by greater interconnectivity of the representative test systems. Similarly, the cyber-threat activity is countered or met with greater cyber-resilience activity in new activities and strict battle rhythms for these to occur.



Malicious use of cyberspace and therefore the need for cyber-resilience

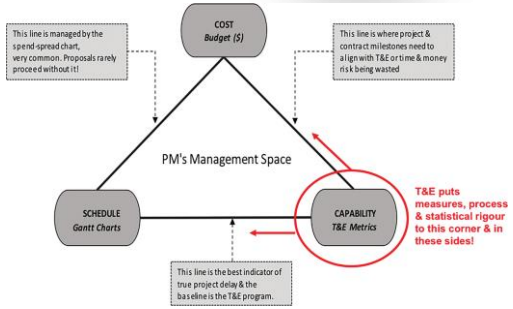
- Requires heavy reinvestment in *in-house* test capabilities, much of which has to necessarily be federated.

T&E - influence design/project risk or fait accompli

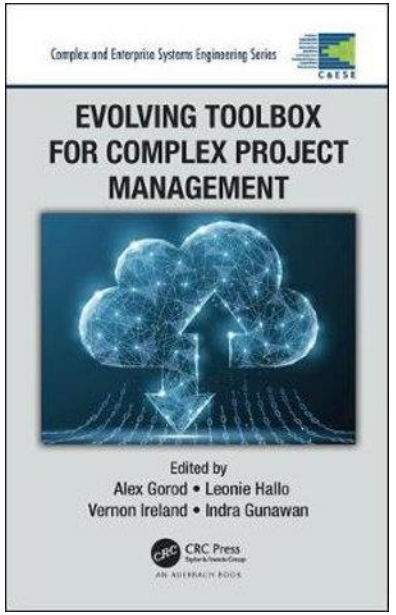
T&E Toolset, Joiner, Efatmaneshnik and Tutty, *Evolving Toolbox for Complex Project Management,* Chapter 16, Edited by Gorod, Hallo, reland, and Gunawan, CRC Press, 2020.

Autonomy and AI.

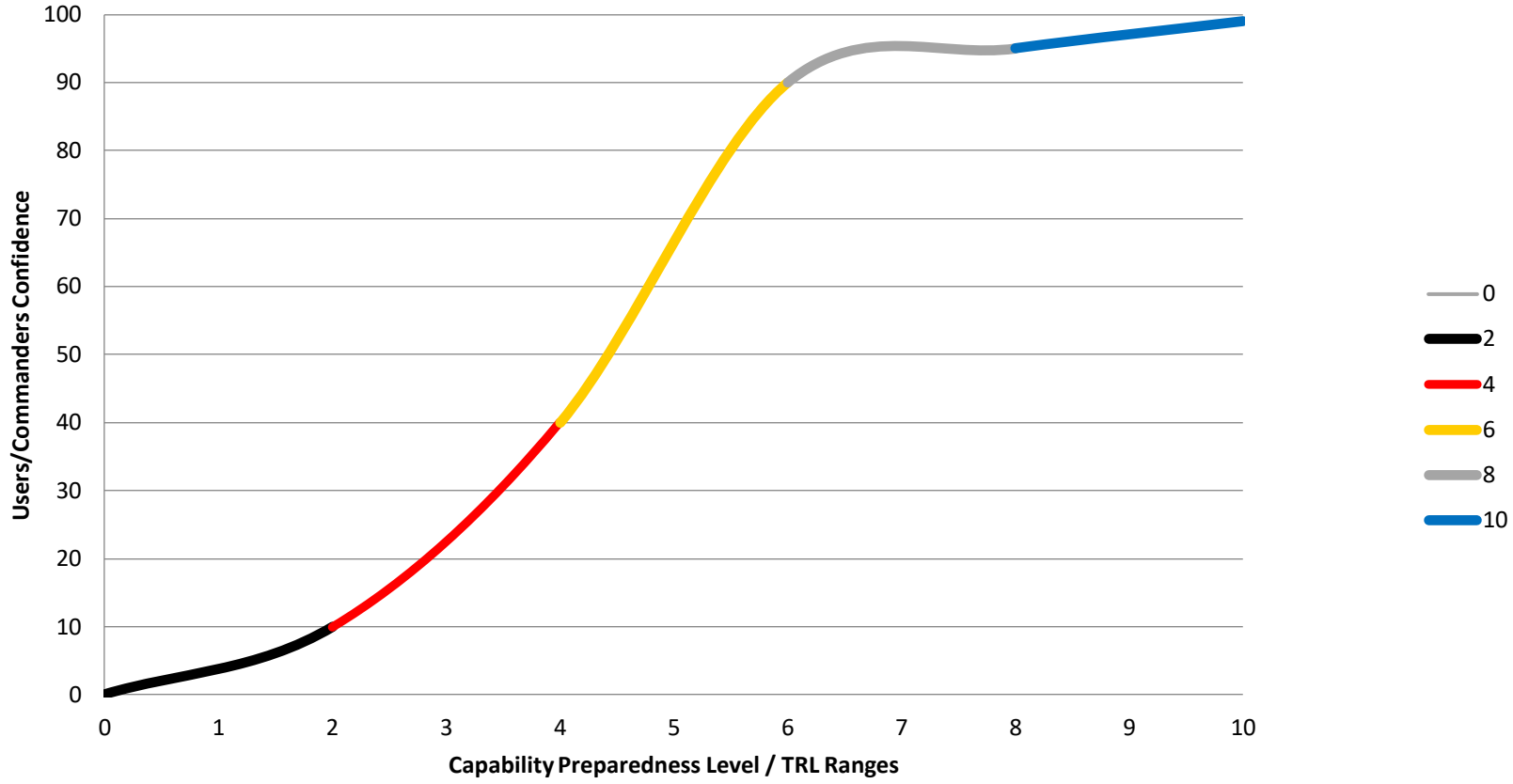
- It may be counter-intuitive to some project managers ***that the most difficult of usability T&E will be for the most autonomous of systems that have less and less users***.
- Contrary to perceptions, ***meaningful human control of such systems requires early and more difficult usability T&E with representative responsible humans, structured around measures that are essentially ones of trust.***



- Other T&E areas of interest for further development are as follows:
  - development of tools that facilitate distributed T&E
  - T&E of human-system interfaces for the positive active control of autonomous and Intelligent systems particularly those with swarm capabilities
  - T&E of non-functional requirements
  - more real-time T&E tools and reporting methods that are useful to key decision makers
  - most importantly T&E of resiliency issues seamlessly embedded in operational families-of-system-of-systems with useful AI assistants meaningfully aiding timely human decision-making, governance and agency.



User and Commander Confidence Levels



Testing early & often starting from M&S equals:

- Resource stability
- Momentum
- Declining risk & growing confidence.

Complex systems: decision making

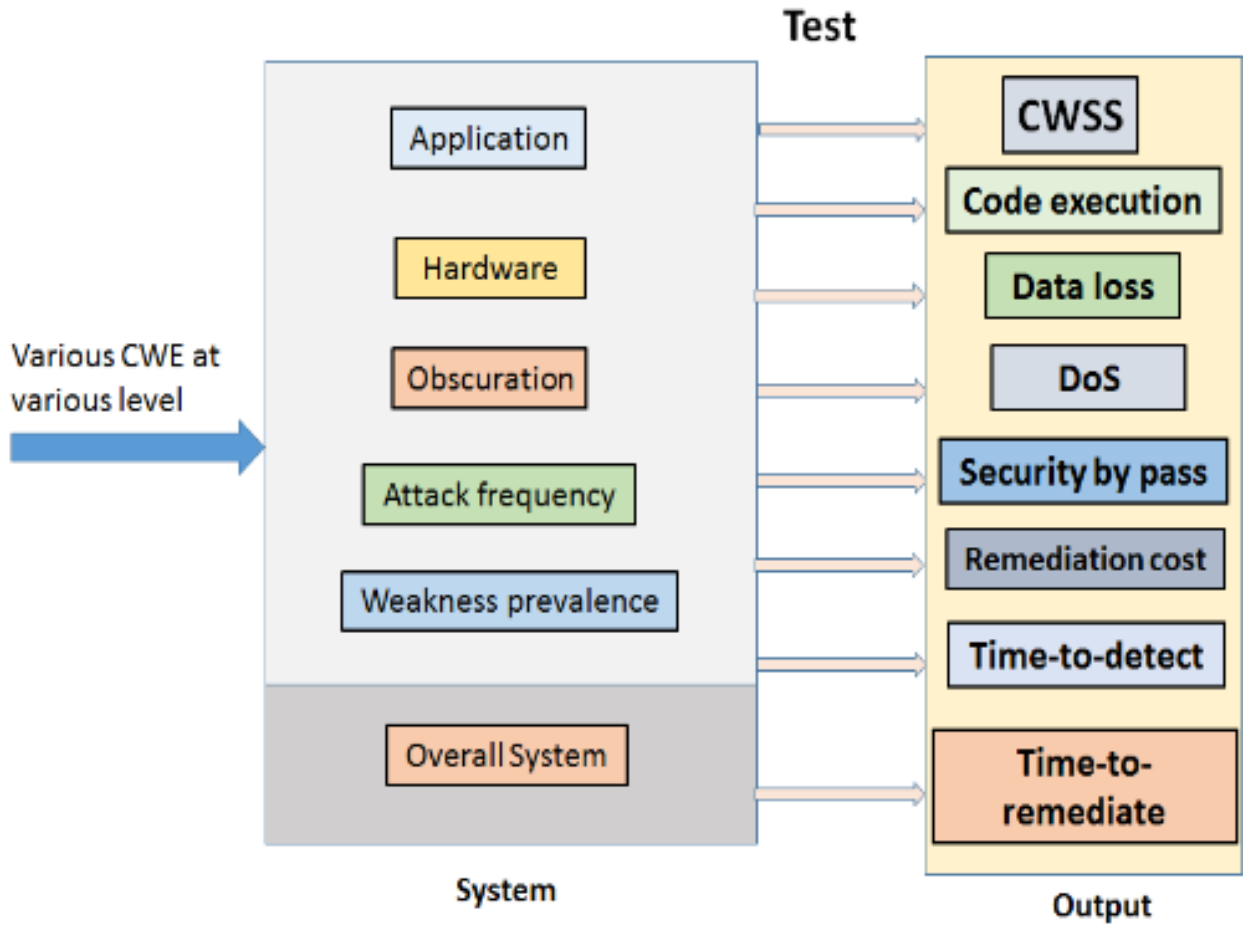
Yin - Danger		Risk Level	Yang - Confidence γ
0 1	Operations Only iaw SPINS	Extreme	Very, very, very 'Bad'
2 3	High Risk Operations (needs Waiver)	High	'Bad'
4 5	Test and Operations (needs Test Plan)	Medium	Some 'goodness', there are 'Issues', some potentially 'Bad'.
6 7	Mission Essential Personnel Only (needs Test / Exercise Plan)	Low	'All goodness', 'On Target', 'Under control'.
8 9 10	Very Low: Public Risk	As Low As	Robust Solution that meets and exceeds the Need identified that is being IV&Ved & Optimised

Initiative 5 - Cybersecurity Protection Plans & T&E

- US Presidential Directive, 2008.
- U.S. began cybersecurity reform with representative operational T&E, at the 'right' of the lifecycle, was fundamental to the DoD understanding the threat consequences and risks properly and then investing in the infrastructure, acquisition and T&E staff competencies, developmental design and then the subsequent two phases of 'shift-left' and 'fully integrated'.
- Clear & comprehensive *Cybersecurity T&E Guide* available on-line.
- Deeper into the U.S. lifecycle there are Cyber Security Assessment & Advisory Teams.
- Cybersecurity is required in the TEMP, including: (1) architecture, (2) operational environment, (3) evaluation structure, (4) authority to operate, & (5) time & resources for the key cybersecurity T&E steps.
- Cybersecurity content required in the *Operational Test Plan*.

Picture from: http://federalnewsradio.com/wp-content/uploads/2016/02/Cybersecurity-Insights2.jpg

Cybersecurity T&E needs advanced test design (HTT) for combinations



Example DOE model for CWE screen

- Incorporates key US DoD test outputs
- Consistent use of CWSS

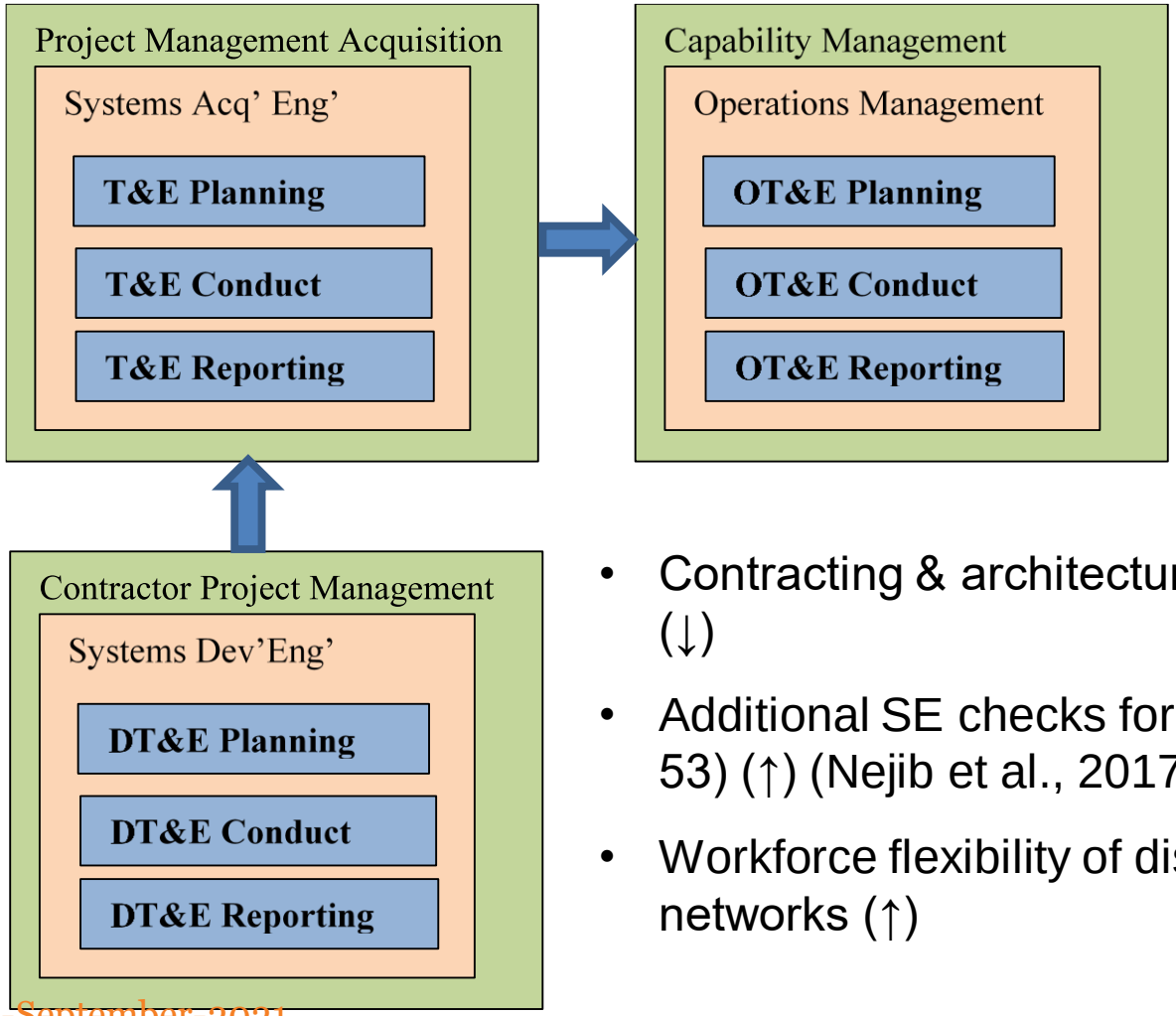
Lbl	Factor Name	No. Lvls	Level 1	Level 2	Level 3	Lev
A	Application	2	Active	Not Active		
B	Hardware	2	Type 1	Type 2		
C	Obscuration	3	Low	Moderate	High	
D	Attack Frequency	3	Rarely	Sometimes	Often	
E	Weakness Prevalence	3	Rarely	Sometimes	Often	
F	CWE Category	5	Type 1	Type 2	Type 3	Type 4	Type 5

Example simple HTT factor table

- Full factorial 540 runs
- 2-ways in rdExpert is 18 runs
- 3-way rigour is 54 runs

Initiative 6 - Permeating these U.S. Initiatives into Industry

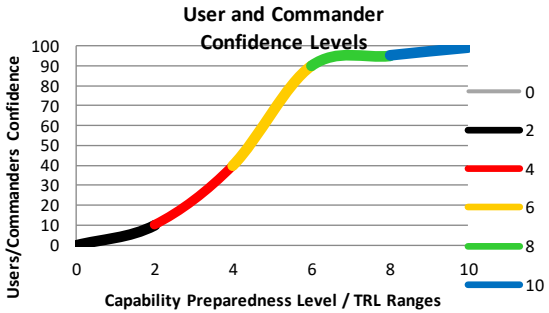
- The DoD is like a stage director & owner, but Defence industry (contractors) do the work. Consider the considerable impact of these U.S. initiatives on contractors:



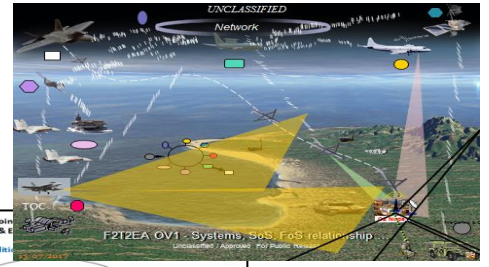
- Competency of industry testers (↑)
- Modelling & simulation skills (↑)
- Proprietary protections with pervasive LVC connectivity (↓)

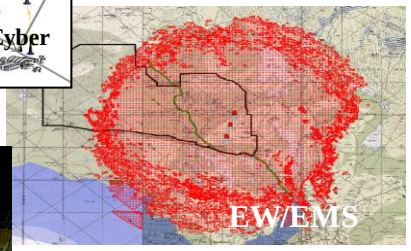
- Contracting & architectural control due Gov't I3 T&E (↓)
- Additional SE checks for cybersecurity tests (circa 53) (↑) (Nejib et al., 2017)
- Workforce flexibility of distributed T&E support via networks (↑)

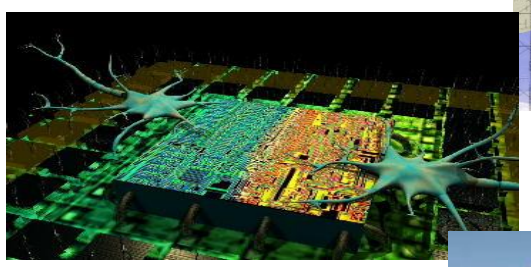
Ethically Aligned ET&E in the PoA: Summary and Recommendations



- **Confidence in Physical Safety and Effects on Threats/Targets**
- **Confidence in operational capabilities with complex adaptive systems — right application / understand risk in joint operations**
- **Confidence in Human Abilities – Cognitive with Tactics, Techniques and Procedures and the Social C2 – CONOPs**
- **Confidence in Information Advantage – Cybersecurity**
- **You and HITL / HOTL / HOOTL – LAWS self-defence against machines**
- **Live, Virtual and Constructive Integrated Mission Environment focused on human agency and decision making**



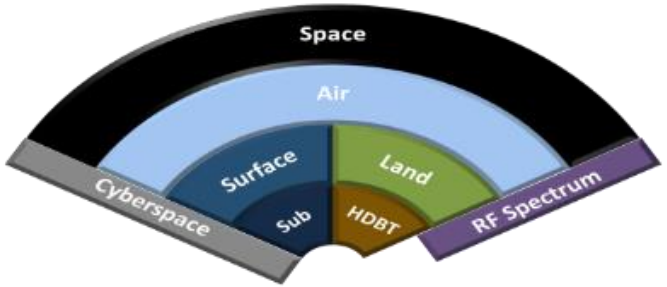




COMPLEXITY, INTERCONNECTEDNESS, AND VULNERABILITY

INCREASE LETHALITY, RESILIENCE AND RAPID ADAPTIBILITY

Humans fight wars, not drones or robots.



Comments?