

# Life-long assurance of Meaningful Human-Autonomy Teaming

Never Stand Still

Capability Systems Centre, School of Engineering and Information Technology



Ms Kate Yaxley



Dr Keith Joiner, CSC

- Meaningful Human Control
- Cognitive agents and trust
- Human-Autonomy Team (HAT)
- Meaningful-HAT
- Readiness levels
- Building and Testing
- Conclusions

# Artificial intelligence

## Narrow AI

- Task specific

## General AI

- Smart Autonomous System (SAS)
- Cognitive
- Allowed to choose

# Meaningful Human Control

- K1 Predictable, reliable and transparent technology
- K2 Accurate information for the user of the outcome sought, operation and function of technology, and the context of use
- K3 Timely human action and a potential for timely intervention
- K4 Accountability to a certain standard

H. Roff and R. Moyes, "Meaningful Human Control, Artificial Intelligence and Autonomous Weapons," in *Briefing paper prepared for the Informal Meeting of Experts on Lethal Autonomous Weapons Systems*, Geneva, UN Convention on Certain Conventional Weapons, 2016, pp. 1-6.

# Cognitive-Cyber Physical Systems



*Perform T&E  
to assure  
K1, K2*

*Trust warriors  
to assure  
K3, K4*

Peacock, K 2019, 20190307raaf8206953\_001.jpg, Defence Image Gallery  
<<https://images.defence.gov.au/assets/archives/5003-All-Defence-Imagery/>>

# Trust between cognitive agents

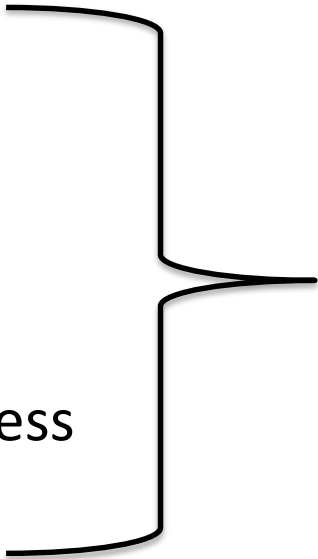
- Risk relationship between truster and trustee
- Allows for sharing of ideas, tasks and outcomes
- Can be shaped by environmental, cultural or learned experiences
- Transparency and communication help to build trust between agents

# Trusted autonomy

Free to choose a truster and trustee relationship

Must also consider

- Reliability
- Performance
- Error type
- Transparency
- Shared awareness
- Shared intent



Performance appraisal of  
Human Teams

# Human-Autonomy Teaming (HAT)

- Operator interaction
- Human-autonomy communication
- Transparency

# Meaningful-Human- Autonomy Teaming (M-HAT)

- Meaningful interaction (learning and reward)
- Human-autonomy communication
- Transparency
- Trust

M. Matessa, K.-P. Vu, T. Z. Strybel, V. Battiste, T. Schnell and M. Cover, "Using Distributed Simulation to Investigate Human-Autonomy Teaming," *Human Interface and the Management of Information. Information in Applications and Services*, vol. 10905, pp. 541-550, 2018.

# Technology Readiness Level

1	Research
2	Analytic study
3	Individual components
4	Early integration
5	Laboratory integration
6	Laboratory or Modelling and Simulation demonstration
7	Demonstration in operational environment
8	Developmental Test and Evaluation
9	Operational use or operational testing

U.S. Government, Technology Readiness Assessment Guide: Best Practices for Evaluating the Readiness of Technology for Use in Acquisition Programs and Projects, Washington: GAO, 2016.



# SAS Education Level

A(a) Remembering	Recall game-play, recite an instruction
B(b) Understanding	Explain process using own words, explain movements
C(c) Applying	Create new moves, create a procedure for others to follow
D(d) Analysing	Troubleshoot, identify tasks for training, conduct a test
E(e) Evaluating	Select the most suitable player, justify a process
F(f) Creating	Integrate information from multiple sources to create a solution, network with others

D. R. Krathwohl, "A Revision of Bloom's Taxonomy: An Overview," *Theory Into Practice*, vol. 41, no. 4, pp. 212-218, 2002.

		SEL					
		A (a)	B (b)	C (c)	D (d)	E (e)	F (f)
TRL	1	SAS Readiness Level (SRL)					
	2						
	3						
	4						
	5						
	6						
	7						
	8						
	9						

# SAS Readiness Level

- Allows for the classification of a SAS system in terms of technological and educational readiness
- Ensures SAS used for teams have the technical and cognitive ability to build effective relationships, contribute and learn

# AlphaGo

TRL 7 – Demonstration in operational environment

SEL A – Recall game-play

## **SRL: 7.Abc**

SEL b – explain process using own words (machine language to teach AlphaGo Zero)

SEL c – create new moves, yet no procedure to follow

# Building M-HAT

- Team building between humans and technology
- Reward trustworthiness
- Transparency in actions and communication

# Testing SAS

- Complete in phases
  - Simulation
  - HWIL
  - Operating environments
- Iterative education
- Iterative technology development
- Iterative training

C. Robbins and M. R. Steffen, "The Future of Autonomous Ground and Surface Systems Testing," *The ITEA Journal of Test and Evaluation*, vol. 39, pp. 82-85, 2018.

# Testing Team

Supervisor

WAI Agent

Educator

Operator

Mechanic

SAS  
Trainers

SAS

Peer(s)

Bystander(s)

M-HAT

J. Scholtz, "Theory and evaluation of human robot interactions,"  
in *Proceedings of the 36th Annual Hawaii International  
Conference on System, Sciences, 2003*, Big Island, 2003.

# Testing Team

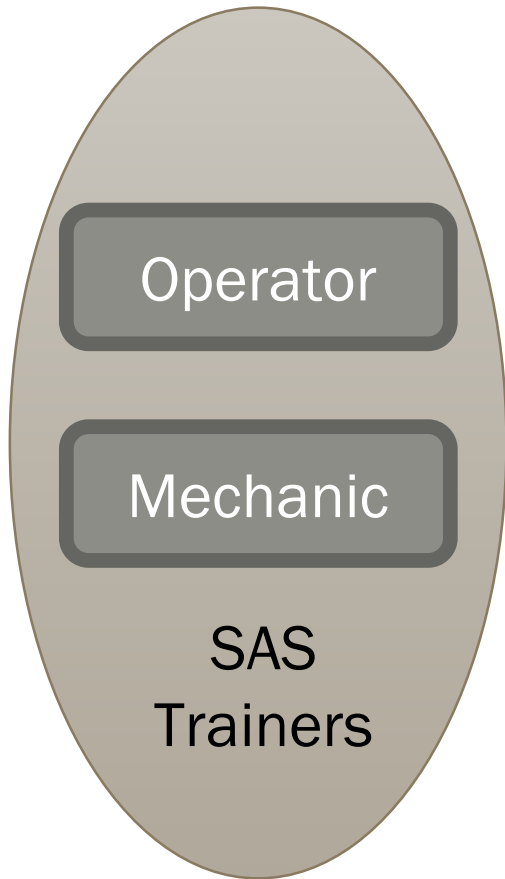
Supervisor

WAI Agent

Educator

- Supervisor
  - *As per HRI*
  - *Human agent*
- WAI Agent
  - *Performs same role as supervisor*
  - *SAS, but not M-HAT*
  - *Interrogates and assesses*
- Educator
  - *Assists with assessing SEL*
- Testing Team responsible for assuring
  - *K1*
  - *K2*
  - *K3*
  - *K4*





- Operator
  - Software assurance
- Mechanic
  - Hardware assurance
- SAS Trainers responsible for assuring
  - K1
  - K2
  - K4



## WAI Agent

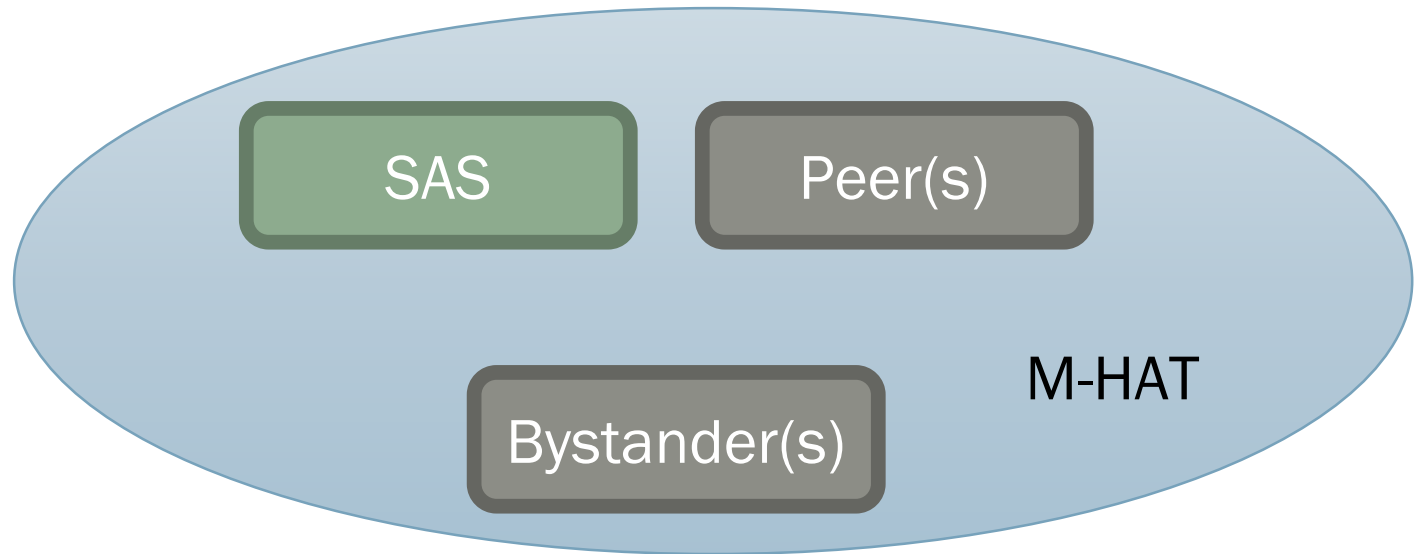
- K1
- K2

## Peer

- K3
- K4

## Bystander

- K3
- K4



# Conclusions

- Teaming will exist both at the ground and supervision level
- HRI concept to assure MHC
- Requires understanding of both technological and cognitive levels to assure M-HAT
- Using SRL, M-HAT and MHC, situational awareness of teams and team dynamics may be developed

# QUESTIONS

	Known to SAS	Not Known to SAS
Known to WAI & M-HAT	Open area/Arena <b>Low Risk</b>	Blind spot <b>Low-Moderate Risk</b>
Not known to WAI & M-HAT	Hidden area/Façade <b>Moderate risk</b>	Unknown <b>High Risk</b>

