INSTITUTE FOR DEFENSE ANALYSES

DATAWorks 2021: Characterizing Human-Machine Teaming Metrics for Test and Evaluation

Heather Wojton, Project Leader

Brian Vickers
Kristina Carter
David Sparrow
Leonard Wilkins
Caitlan Fealing

April 2021

Approved for Public Release

Distribution Unlimited

IDA Document NS D-21564

Log: H 2021-000050

INSTITUTE FOR DEFENSE ANALYSES
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

The Institute for Defense Analyses is a nonprofit corporation that operates three Federally Funded Research and Development Centers. Its mission is to answer the most challenging U.S. security and science policy questions with objective analysis, leveraging extraordinary scientific, technical, and analytic expertise.

About This Publication
This work was conducted by the Institute for Defense Analyses (IDA) under contract HQ0034-19-D-0001, Task BD-9-2299(90), "Test Science Applications," for the Office of the Director, Operational Test and Evaluation. The views, opinions, and findings should not be construed as representing the official position of either the Department of Defense or the sponsoring organization.

Acknowledgments
The IDA Technical Review Committee was chaired by Mr. Robert R. Soule and consisted of Dr. Allison Goodman, Dr. Daniel Porter, Dr. Justace Clutter, and Mr. Scott Shaw from the Operational Evaluation Division, Mr. Brian Williams from the Joint Advanced Warfighting Division, and Dr. Emily Fedele from the Science and Technology Division.

For more information:
Heather Wojton, Project Leader
hwojton@ida.org • 703-845-6811

Robert R. Soule, Director, Operational Evaluation Division
rsoule@ida.org • (703) 845-2482

Rigorous Analysis | Trusted Expertise | Service to the Nation

INSTITUTE FOR DEFENSE ANALYSES

IDA Document NS D-21564

DATAWorks 2021:
Characterizing Human-Machine Teaming Metrics for Test and Evaluation

Heather Wojton, Project Leader

Brian Vickers
Kristina Carter
David Sparrow
Leonard Wilkins
Caitlan Fealing

Executive Summary

The Department of Defense and partner organizations are developing advanced machine systems that will work with humans to accomplish missions. Given that these human-machine teams (HMTs) have never undergone test and evaluation (T&E), this briefing helps guide evaluators through the new challenges HMTs bring. It defines human-machine teaming, describes challenges in evaluating HMTs, and provides a framework for categorizing metrics important to the T&E of HMTs.

Human-machine teaming is broader than the simple act of an individual using a system to accomplish a task. It involves extensive interactions between humans and systems as they work together to achieve a collective goal. Given the highly collaborative nature of HMT, measuring the machine and the human is insufficient. We also need to measure the team itself, and those measures need to be mission relevant, quantitative, and objective.

Several unique challenges arise when evaluating HMTs. These include how to address opaque mental models and situations in which machines direct communications, self-task, or task the human. For example, consider a human-machine search-and-rescue team in which an autonomous drone flies overhead to locate survivors in a collapsed building, alerting a robot on the ground when it finds a survivor. The robot then pulls the survivor from the rubble and brings them to a human medic for treatment. How do you evaluate the drone's process for deciding where to search? Or how it communicates with the robot? What about the robot's responses to those communications? The medic's decisions about how to treat the survivors and in what order? How do the drone, robot, and medic cooperate and prioritize their efforts toward the most seriously injured survivors? How do they coordinate their other efforts? How do they handle the difficulties inherently associated with constantly changing circumstances? As is evident, team member interactions are key.

The framework outlines major categories of HMT evaluation, including capabilities (what abilities did the team have?), interactions (how did the team work together and coordinate their actions to achieve the goal?), and performance outcomes (what resulted from the team's efforts?). It emphasizes team metrics and the coordination of metrics across humans and machines. Thus, if you assess cognition for a human (i.e., attention and judgment), you

need to assess cognition for the machine (i.e., information-processing architecture and decision-making algorithms) as well.

The framework also provides a structure for identifying and selecting appropriate measures to evaluate team effectiveness. All such metrics are derived from prior scientific research.

First, look at the abilities of the human and of the machine, as any of these may underlie failures in teaming. Assess such things as the human's training and experience, psychological traits, physical abilities, attitudes, cognitive resources, and mental workload or fatigue. Consider factors related to the machine's cognitive structure and hardware components, such as its programmed mission knowledge, its operating system and other software, and its physical sensors and the platform on which it is built.

Next, examine critical areas that may underlie interaction failures. These include the machine's situational awareness, resource allocation, and resource use in different situations. For instance, how much power a machine requires to use its sensors to find new survivors can affect whether the machine is available to assist with other team needs. These critical areas also include the human's perspectives and decision-making processes. For instance, a human's understanding of the situation can affect how they behave in that situation and whether they trust the machines they're working with.

Finally, consider potential vulnerabilities. What threats may prevent the team from accomplishing its goal? What consequences are there should the team fail? What additional problems may cascade from this failure? It's important to identify any problems so they can be alleviated or resolved going forward.

Ultimately, this briefing conveys two important takeaways for the T&E community:

- Challenges are different when evaluating an HMT than when evaluating humans using tools or systems. A team's humans and machines (called agents) must pursue the same goal, affect the current problem state, and coordinate action amongst themselves; these interactive factors expose the team to new vulnerabilities and more points of failure.
- Mission outcomes alone cannot be used to identify potential system vulnerabilities. Interactions among agents increase the problem space being evaluated.

Contents

DATAWorks 2021:

Characterizing Human-Machine Teaming Metrics for Test and Evaluation

Heather Wojton, Dave Sparrow, Brian Vickers,
Kristina Carter, Jay Wilkins, and Caitlan Fealing

April 2021

Institute for Defense Analyses
4850 Mark Center Drive ● Alexandria, Virginia 22311-1882

Assessing human-machine teams must involve more than outcome-based testing

This brief's framework extends past mission accomplishment metrics to include metrics for evaluating suitability and effectiveness

The framework:

- Identifies concepts critical to effective teaming
- Emphasizes the importance of interaction – an element that is not analyzed currently
- Provides a structure for identifying and selecting appropriate measures to evaluate team effectiveness

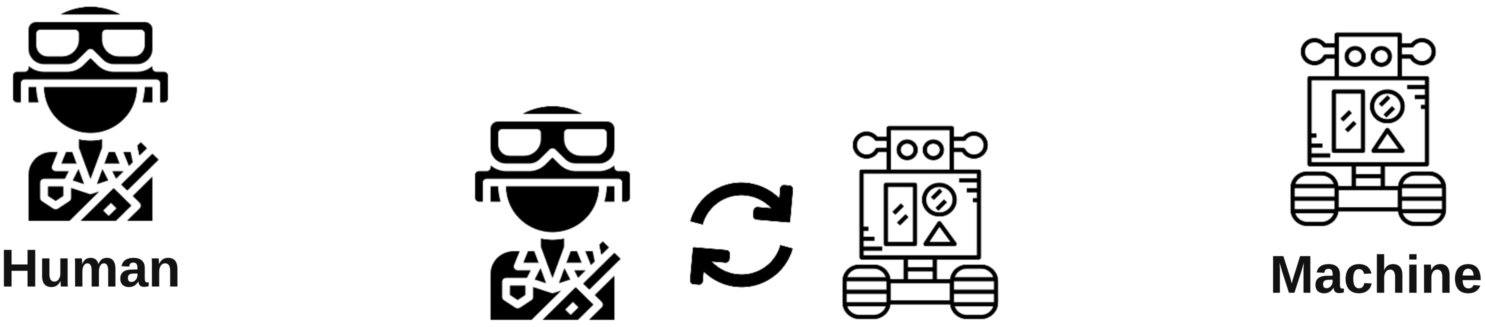
This framework allows us to understand whether a team is effective ***in general***, not just effective during the observed task.

The DoD is investing in human-machine teaming systems and concepts

Our experience is with human teams and humans using tools. Human-machine teams are different.

Bruemmer, Marble, & Dudenhoeffer, 2002; Laird, Ranganath, & Gershman, 2019; Pellerin, 2015

Team interaction adds a new degree of unpredictability that requires new approaches



Human

Team

Machine

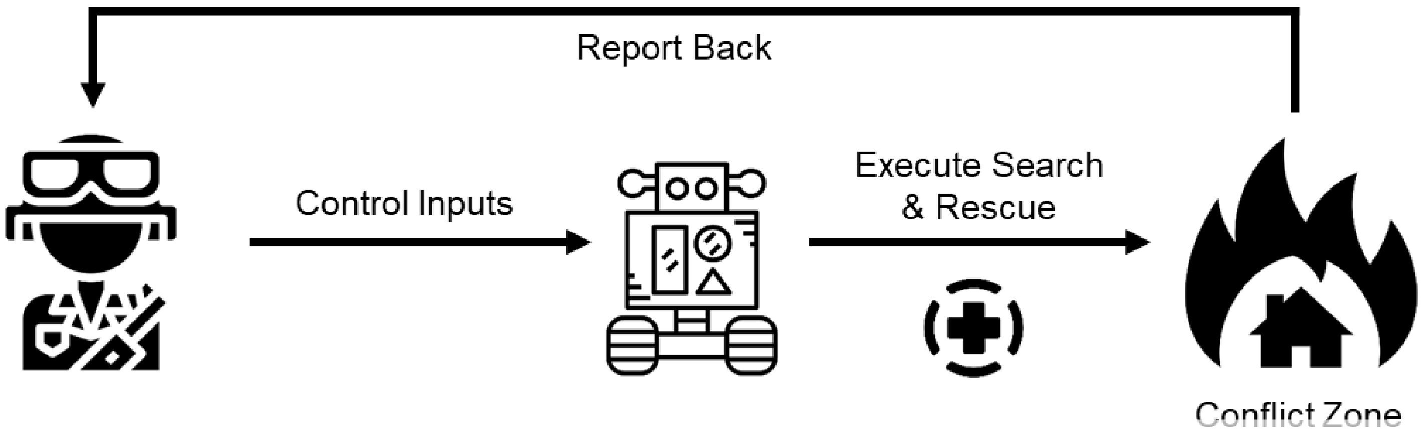
In a teaming context, measuring the machine and the operator alone is insufficient; interaction between the human and machine increases the problem space

Test and evaluation (T&E) measures of the agents and the team should be mission relevant, quantitative, and (if possible) objective

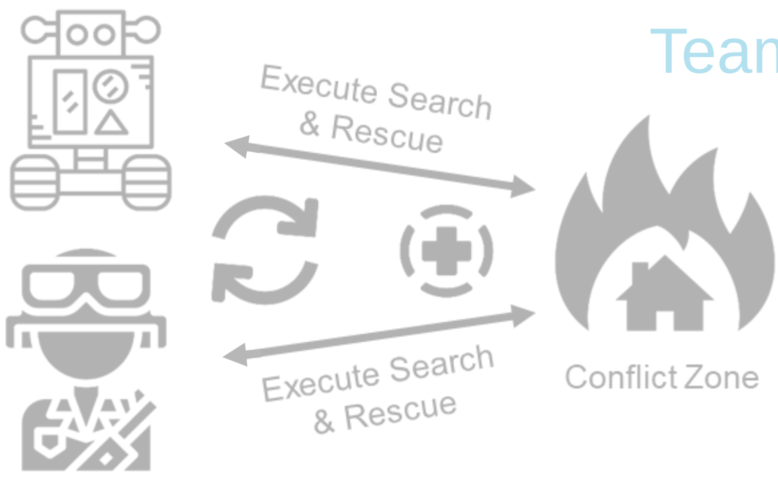
Damacharla et al., 2018

Currently, measuring humans using tools characterizes performance well

Tool



Teammate



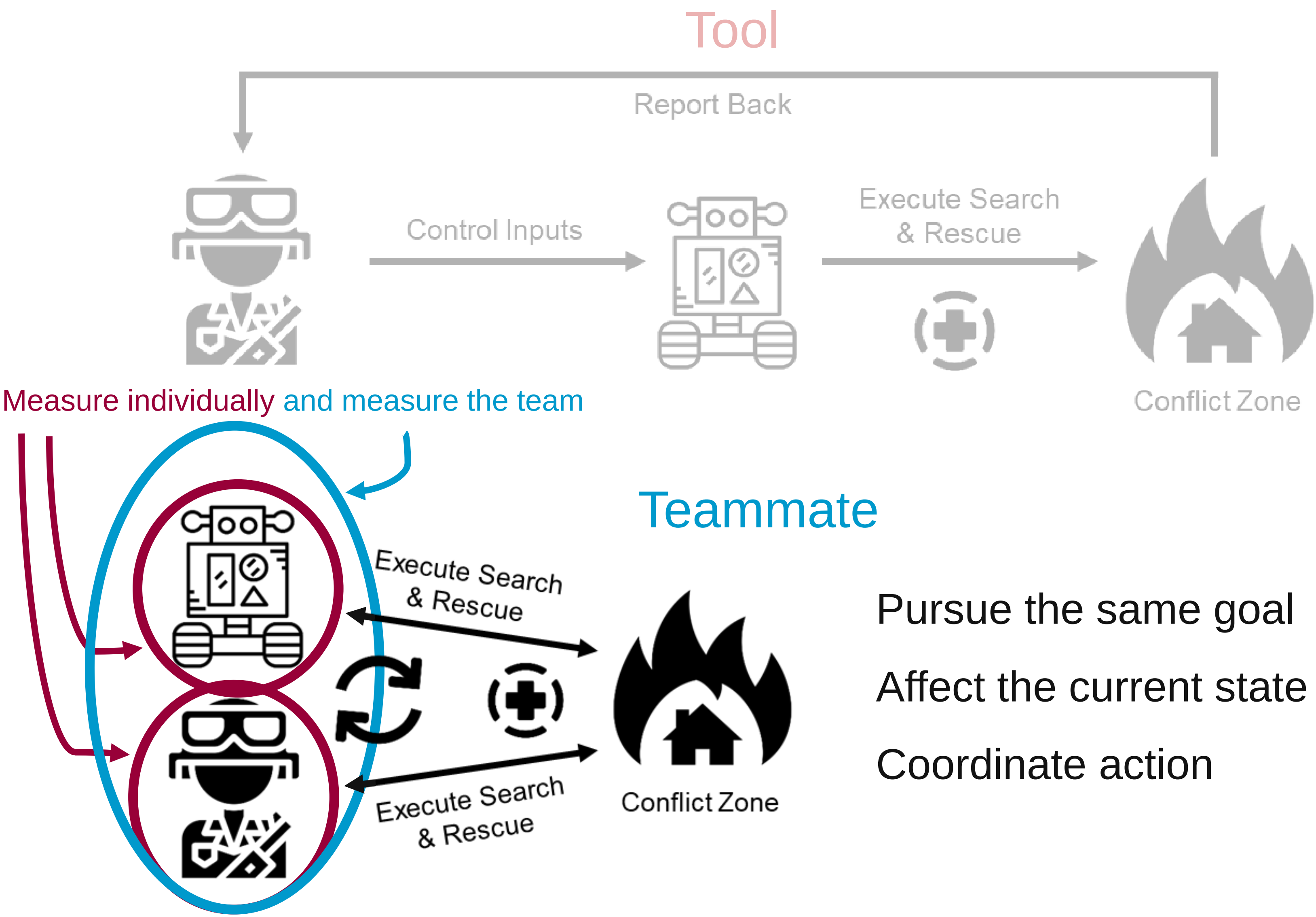
Pursue the same goal

Affect the current state

Coordinate action

Bruemmer, Marble, & Dudenhoeffer, 2002; Laird, Ranganath, & Gershman, 2019; Pellerin, 2015

Teammates, working in parallel or synergy with one another, need to have more information to characterize performance



Pursue the same goal

Affect the current state

Coordinate action

Bruemmer, Marble, & Dudenhoeffer, 2002; Laird, Ranganath, & Gershman, 2019; Pellerin, 2015

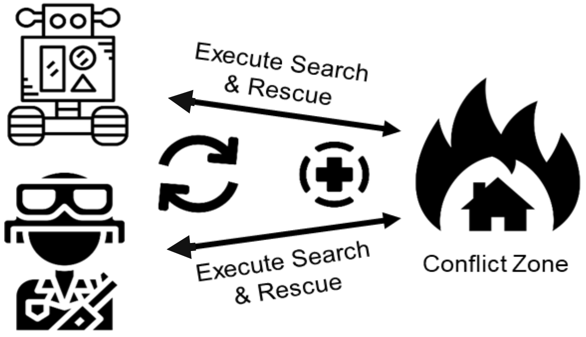
Human and machine collaboration must meet all three conditions to be considered teams

Machines as teammates:

Pursue the same goal

Affect the current state

Coordinate action

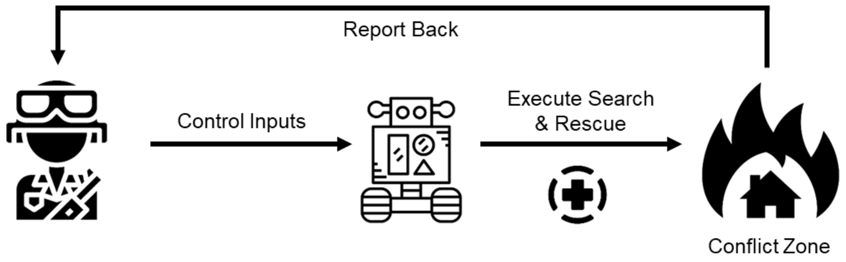


Machines used as tools:

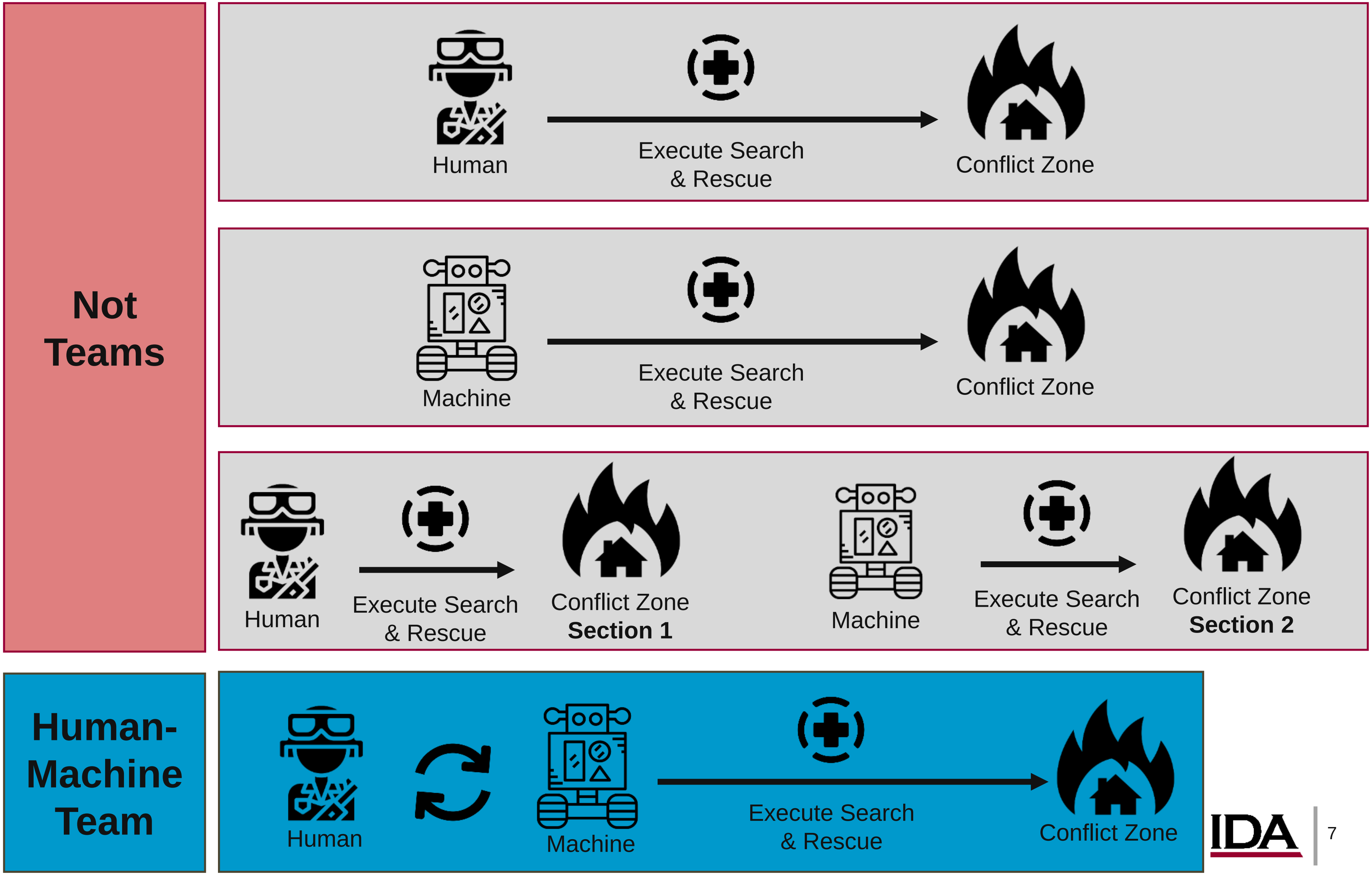
Handle inputs, not goals

Require direct instruction for action

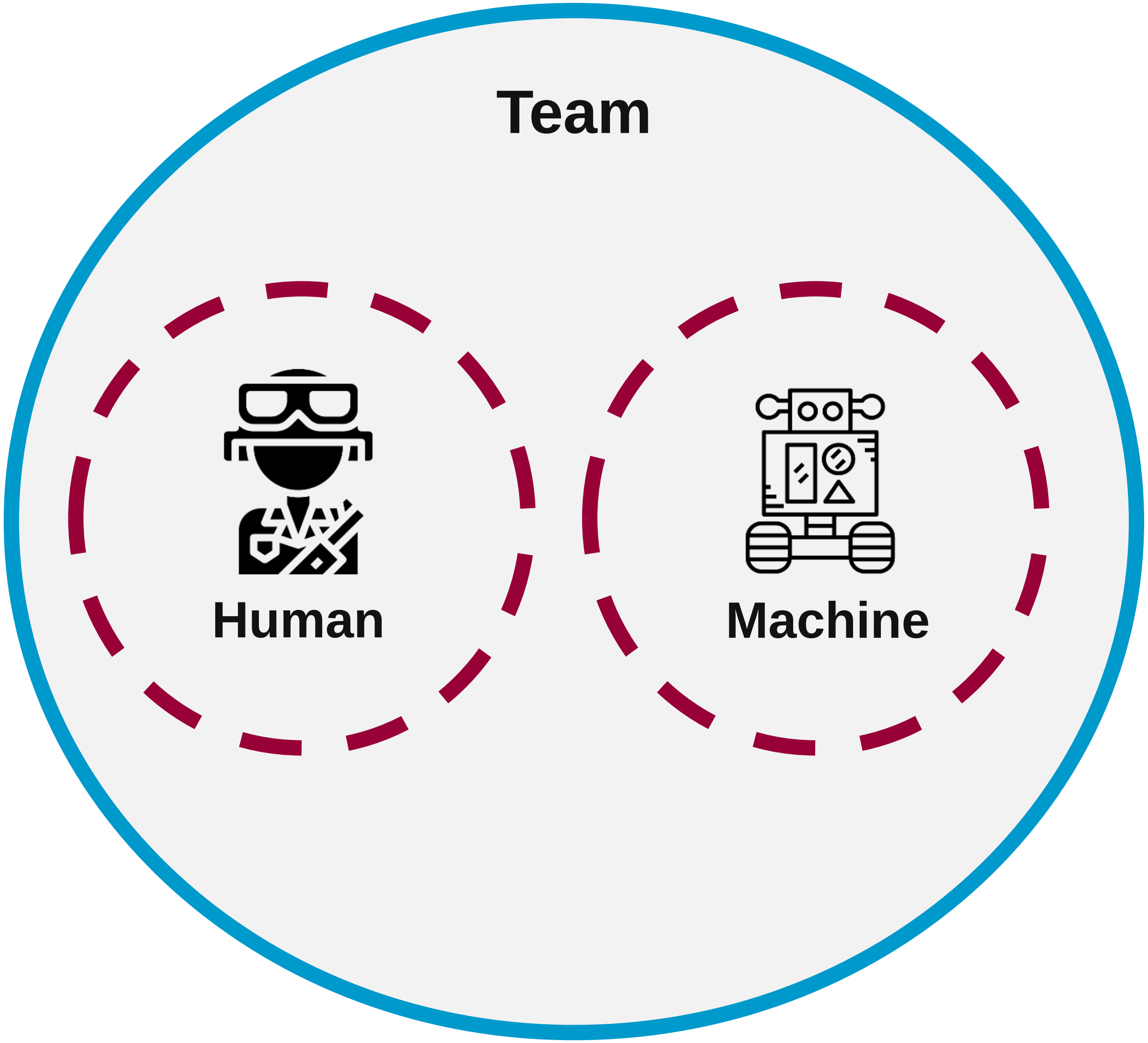
Only complete assigned functions



Humans and machines can support missions in different ways, but human-machine teams are special cases



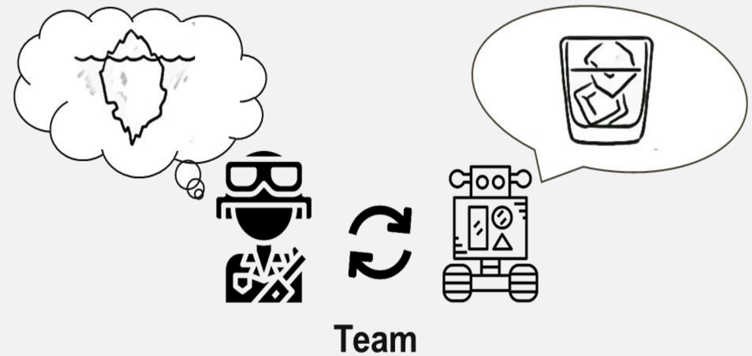
You cannot predict team performance from your knowledge of the individual team agents



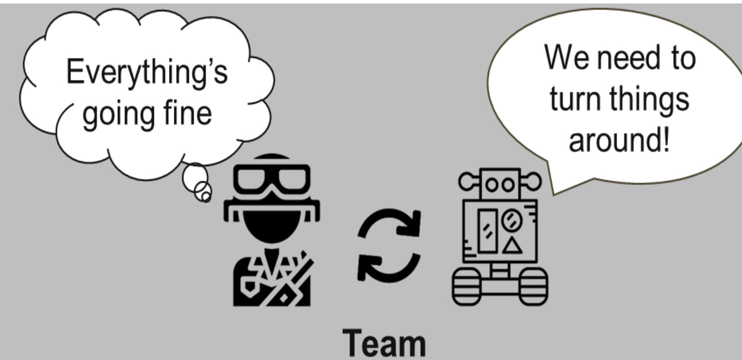
Bruemmer, Marble, & Dudenhoeffer, 2002; Laird, Ranganath, & Gershman, 2019; Pellerin, 2015

The team interaction is the new, complicating factor

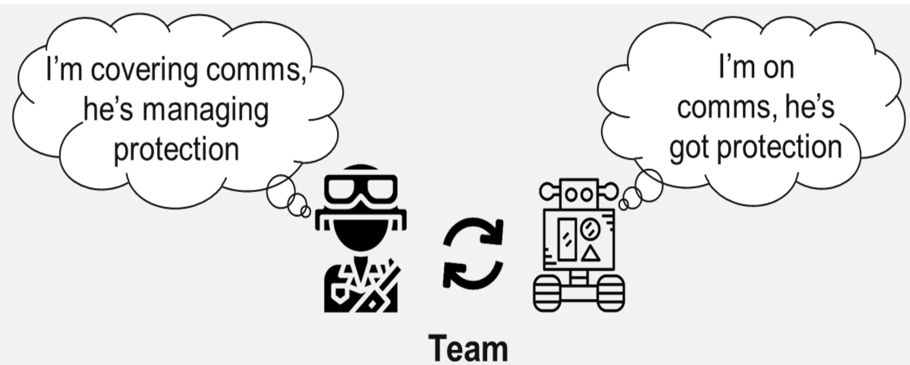
Agents' world or mental models may be opaque



Machines may direct communications within the team



Machines may self task or (implicitly) task the human

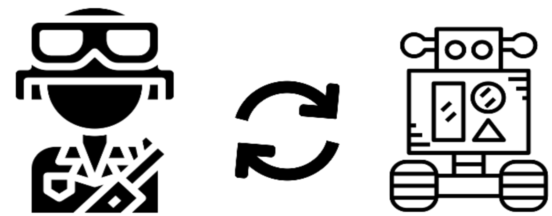


For further discussion, see Laird, Ranganath, & Gershman, 2019.

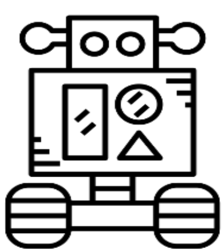
Testing mission outcomes alone will not capture team features such as potential vulnerabilities

Bruemmer, Marble, & Dudenhoeffer, 2002; Laird, Ranganath, & Gershman, 2019; Pellerin, 2015

Human-machine teams require a new focus on the team as the unit of analysis



RS47: Rescuing disaster victims



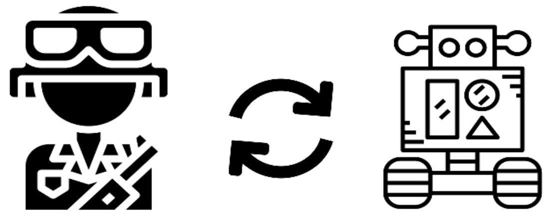
Locating victims
and transporting victims

Location information;
status report

Status report;
transport request

Attending to victims

The team interaction is a new, critical factor



Capability

Team abilities and capacities that are relevant to a particular application environment and the specific applications (i.e., missions) the team must perform

Interaction

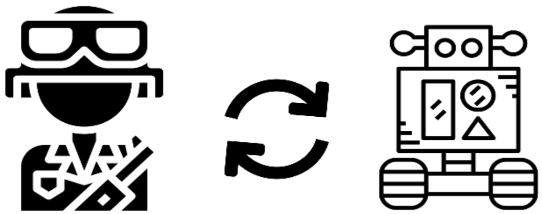
How team members engage in coordination, cooperation, and efficient goal pursuit during execution

Performance

Assessment of decisions, results, and subsequent effects generated by or attributable to team action

Interaction and performance definitions from APA Dictionary of Psychology, 2007

These concepts reflect Joint Force Command doctrine describing HOW you use MEANS to achieve ENDS



Capability

MEANS: Abilities of the team to get closer to the goal state. What the team can do coming into T&E.

Interaction

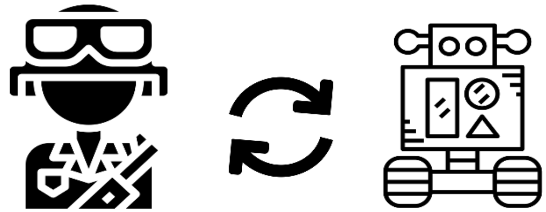
HOW: Engagement processes between members, with the environment and the mission. Things that change during T&E.

Performance

ENDS: Outcomes that characterize mission accomplishment. Large amount of MOEs, MOSs, etc.

Acronyms: MOE – Measure of Effectiveness; MOS – Measure of Success; T&E – Test and Evaluation

The concepts we care about at the team level relate to the team interactions and outcomes



Capability

Communication structure

Structure-based interaction

Architecture

Interaction

Team perspective

Cooperative behavior

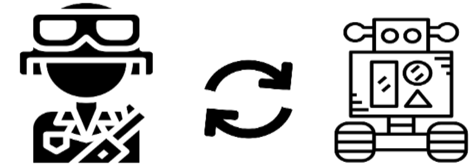
Workload

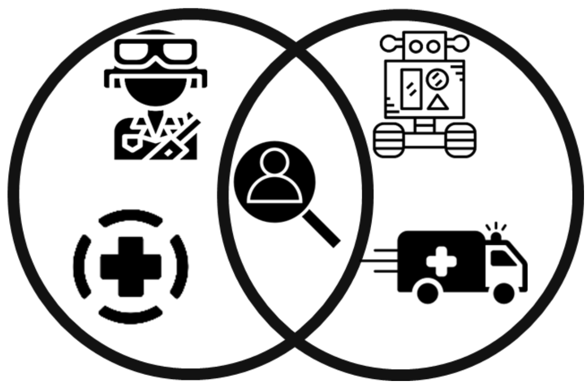
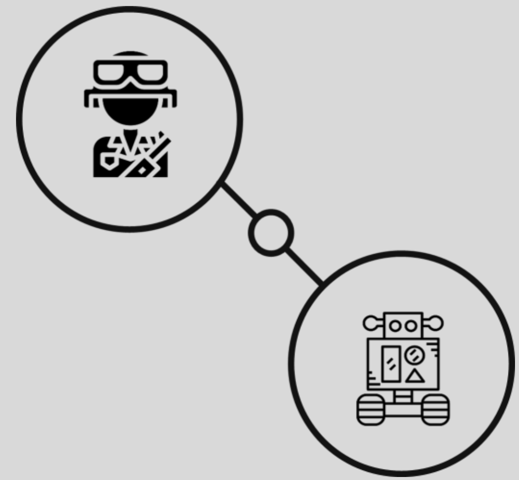
Performance

Collective decision making

Collective task performance

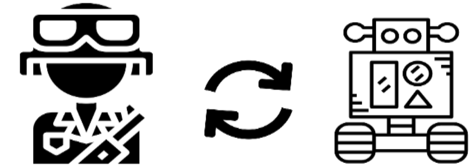
Team capabilities determine what is possible for the team

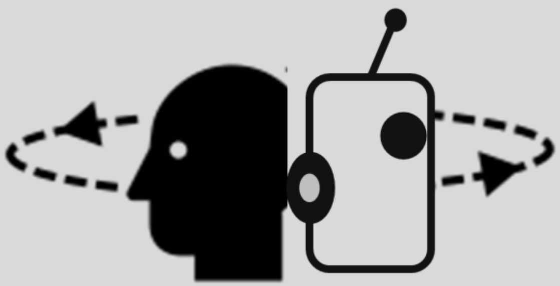
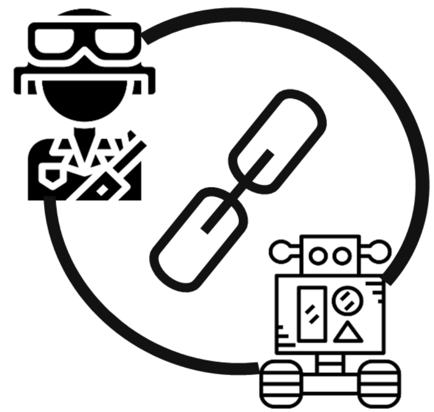
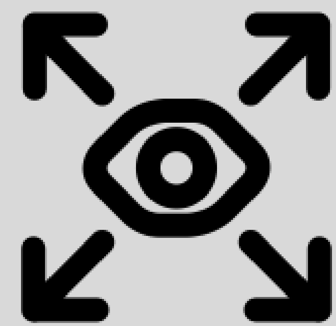


Communication Structure	Structure-Based Interaction	Architecture
		
Joint mission knowledge	Role clarity and adaptability	Hierarchical relationships

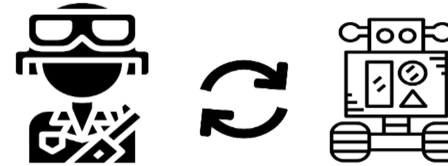
Note: Term definitions and relevant references are in backup slides.

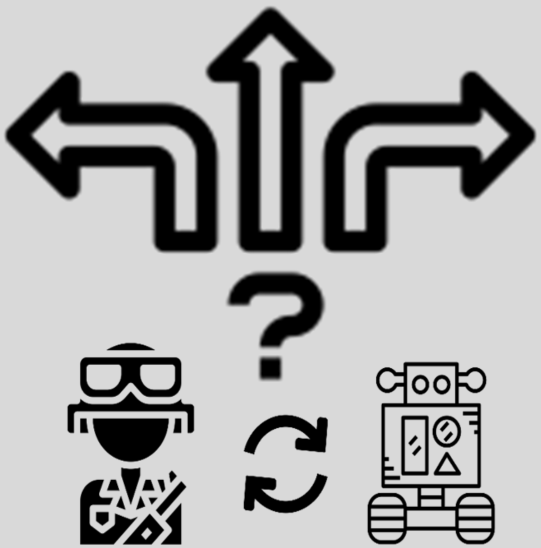
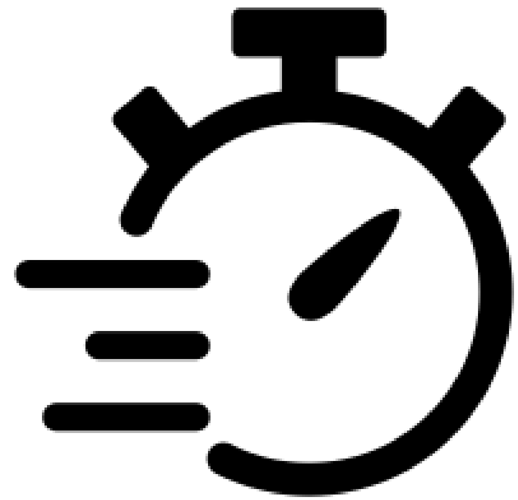
Team interaction captures how the team works and cooperates during the mission



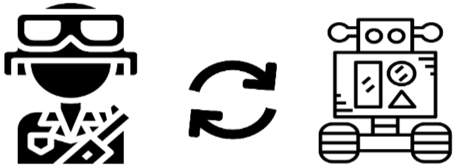
Team Perspective	Cooperative Behavior	Team Resource Allocation
 Team situational awareness	 Team cohesion	 Joint attention allocation

Team performance measures capture mission outcomes



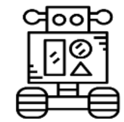
Collective Decision Making	Collective Task Performance
 Decision-making optimality	 Timeliness

Team measures can be assessed only with two or more agents' active involvement



Communication Framework	**Structure-Based Interaction**	**Architecture**
Information flow Joint world model Joint mission knowledge	Influence Role clarity Role adaptability	Hierarchical relationships Uni- and bi-directional relationships Learning patterns
Team Perspective	**Cooperative Behavior**	**Team Resource Allocation**
Situational awareness Information accuracy	Cohesion Intervention Team trust Agency shifting	Joint attention allocation Workload transfer Endurance
Collective Decision Making		**Collective Task Performance**
Optimality Risk level Robustness		Task success/failure Timeliness Planning recognition Efficiency

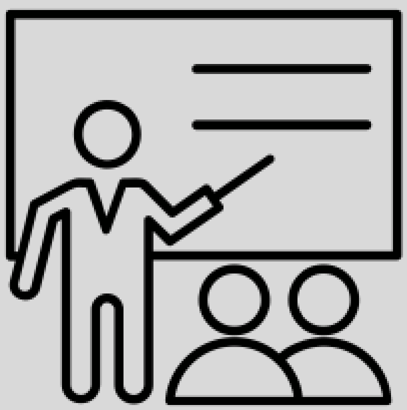
Individual actions cannot account for an entire team's performance, but they might still affect mission outcomes



Human			Machine		
Capability					
Training and experience	Psychological traits	Physical abilities	Cognition structures and algorithms	Cognition software and hardware	Standard platform
Worldview					
Judgment and attitudes	Cognitive allocation	Effort	Perspective	Resource allocation	Resource use
Performance					
Decision making	Task performance		Decision making	Task performance	

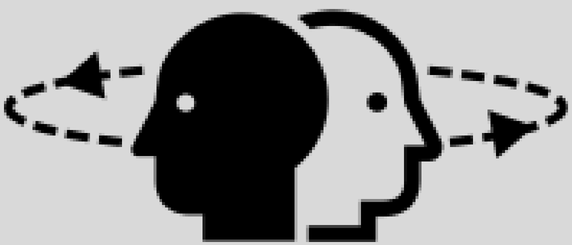
Note: Term definitions and relevant references are in backup slides.

Human capabilities might be the underlying cause of failures in teaming

Training and Experience	Psychological Traits	Physical Abilities
		
Mission knowledge	Intelligence	Physical fitness

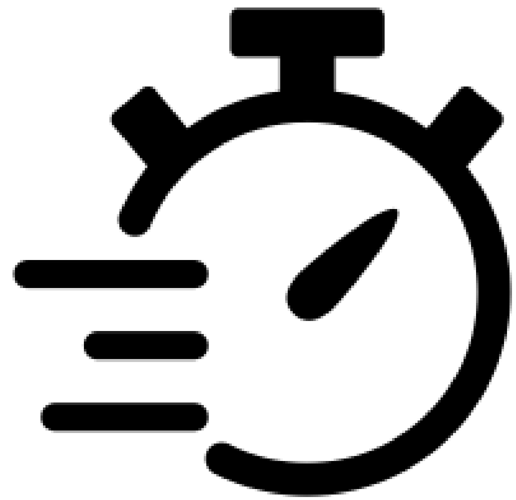
Note: Term definitions and relevant references are in backup slides.

Human judgment, attention, and effort might be the critical areas underlying interaction failures

Judgments and Attitudes	Cognitive Allocation	Effort
		
Situational awareness	Attention allocation	Resource availability (workload, fatigue)

Note: Term definitions and relevant references are in backup slides.

Human performance might be subpar without damaging team outcomes

Decision Making	Task Performance
 **Decision-making optimality**	 **Timeliness**

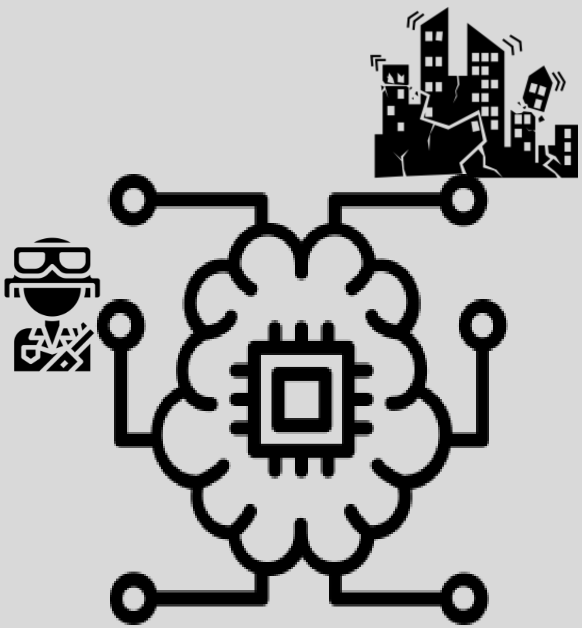
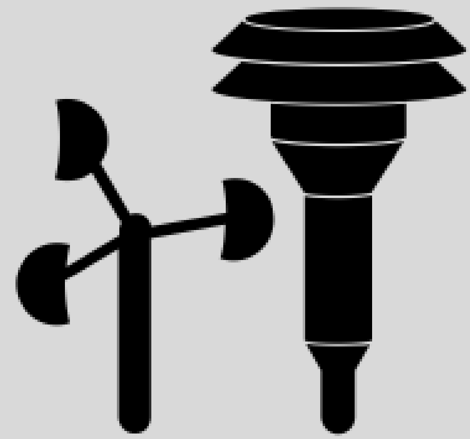
Note: Term definitions and relevant references are in backup slides.

Human capabilities, judgments, and behaviors affect interaction with the machine and the mission

Training and Experience	Psychological Traits	Physical Abilities
Mental model	Decisiveness and impulsiveness	Physical fitness
Mission knowledge	Flexibility	Sensors (organs/equipment)
Teammate knowledge and experience	Intelligence	
Judgments and Attitudes	**Cognitive Allocation**	**Effort**
Situational awareness	Working memory	Resource availability (workload, fatigue)
Trust	Attention allocation	Usability
	Other dependability monitoring	
Decision Making		**Task Performance**
Optimality; Robustness		Error rates; Timeliness
Risk level; Reliance		Other performance

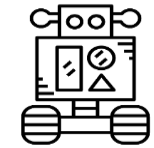
Note: Term definitions and relevant references are in backup slides.

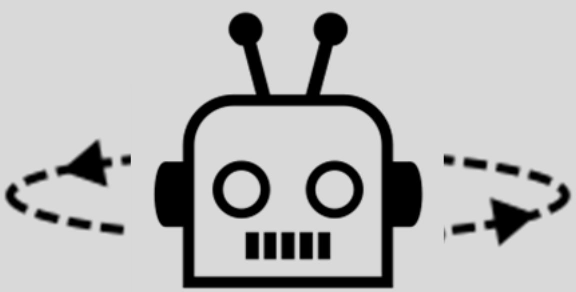
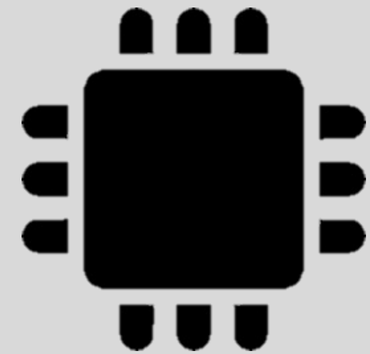
Machine capabilities might be the underlying cause of failures in teaming

Cognition Structures and Algorithms	Cognition Software and Hardware	Standard Platform
		
Mission knowledge	Operating system	Sensors

Note: Term definitions and relevant references are in backup slides.

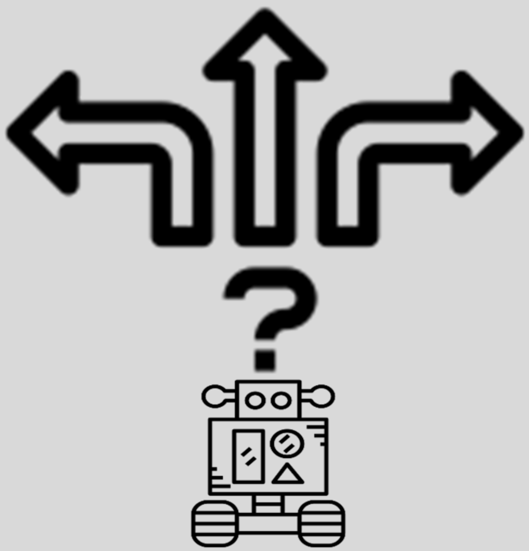
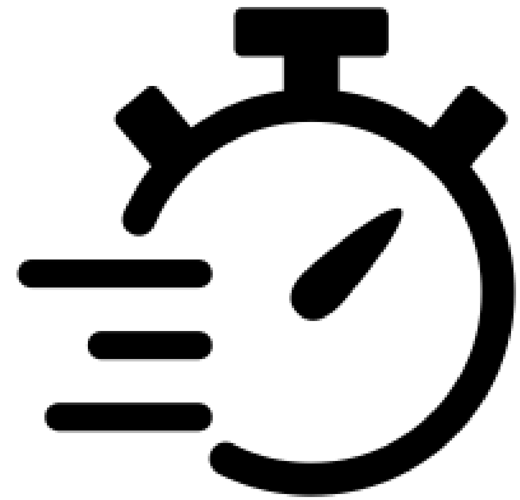
Machine perspective, resource allocation, and resource use critically impact team outcomes



Perspective	Resource Allocation	Resource Use
		
Situational awareness	**Process/threat monitoring**	**Processing**

Note: Term definitions and relevant references are in backup slides.

Machine performance might be subpar even if the team successfully completes the mission

Decision Making/Action	Task Performance
	
Decision-making optimality	**Timeliness**

Note: Term definitions and relevant references are in backup slides.

Machine capabilities, awareness, and actions affect interaction with the human and the mission

Cognition Structures and Algorithms	Cognition Software and Hardware	Standard Platform
World model	Prioritization	Structural and mechanical elements
Mission knowledge	Algorithm flexibility	Sensors
Teammate knowledge and experience	Operating system	Computer and peripherals
		Integration hardware
Perspective	**Resource Allocation**	**Resource Use**
Situational awareness	Activity execution	Platform operating margin
	Process/threat monitoring	Processing
	Other dependability monitoring	
Decision Making/Action		**Task Performance**
Optimality Robustness		Error rates
Risk level Reliance		Timeliness
		Other performance

Note: Term definitions and relevant references are in backup slides.

By focusing on the team, this framework provides a guide for comprehensively evaluating human-machine teams

- Outlined critical approach for assessing effective teaming

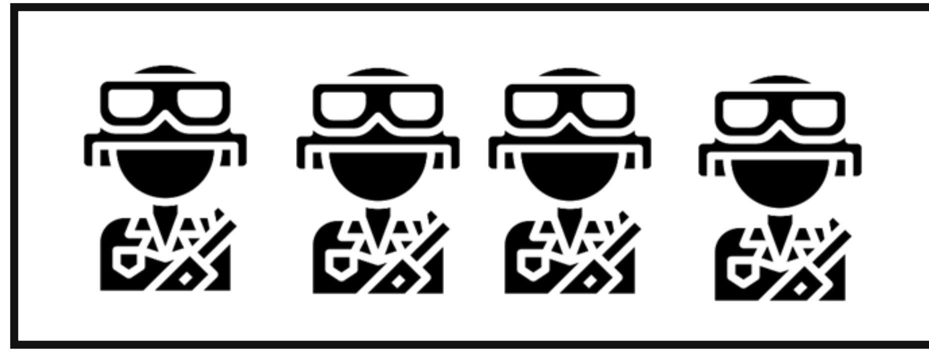
- Emphasized the importance of interaction – focused on the team as the unit of analysis

- Provided a structure for identifying and selecting appropriate measures to evaluate team effectiveness

This framework allows us to understand whether a team is effective ***in general***, not just effective during the observed task.

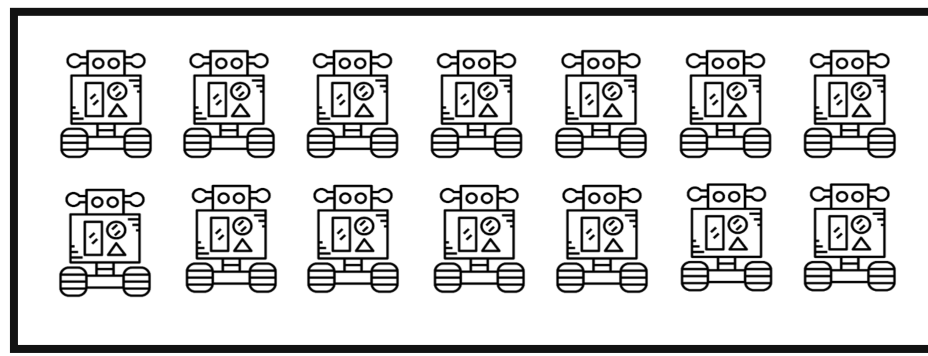
NEXT STEPS

Apply the HMT framework to develop T&E methods for swarms and other unbalanced human-machine teams



+

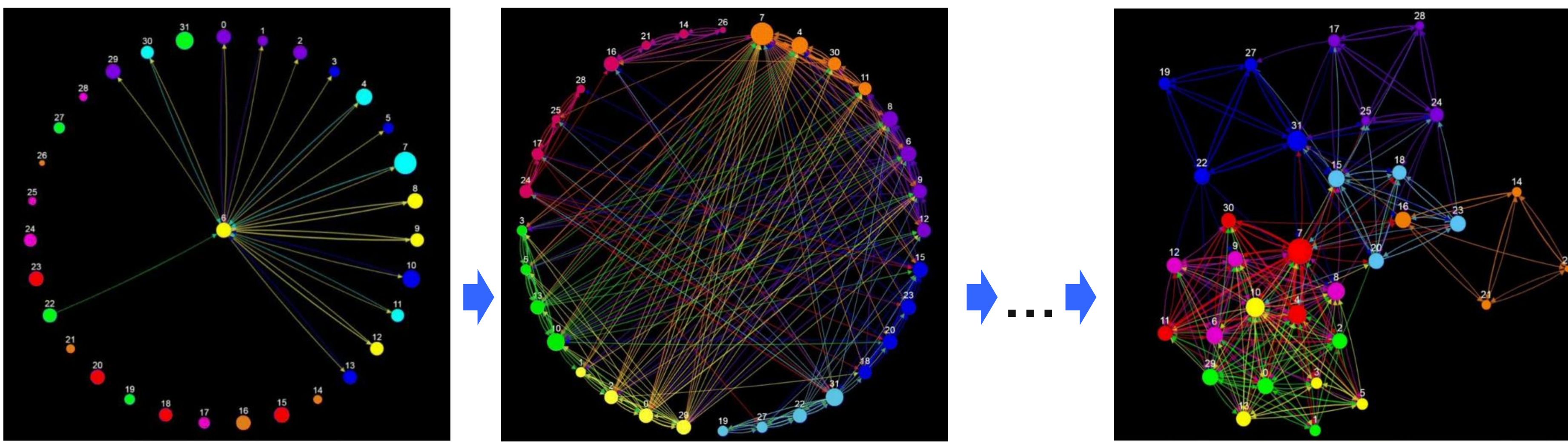
+



Examining human-machine teams (HMTs) with larger, more complex team structures can yield more advanced assessments of operator training, team training, and performance optimization

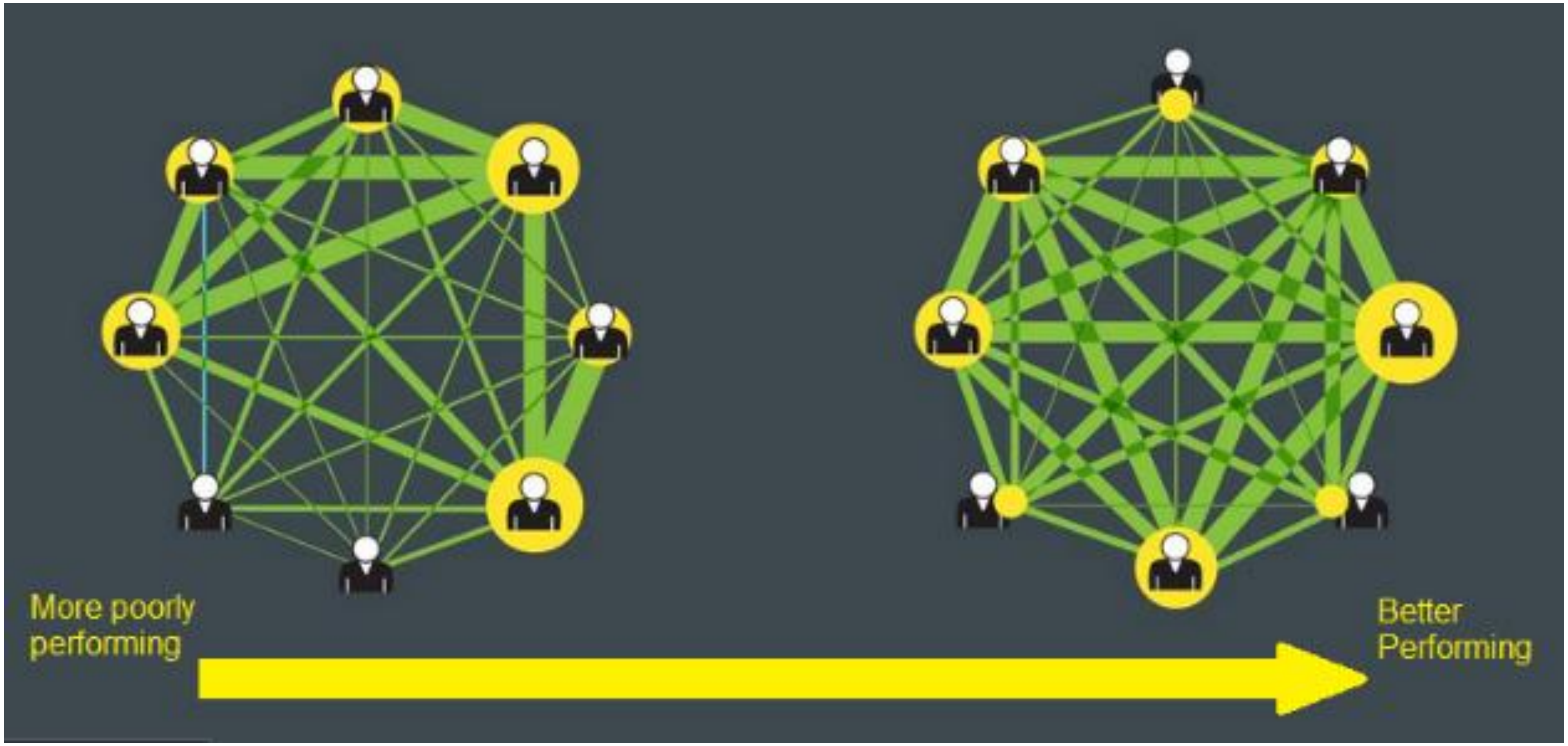
Expand HMT test and evaluation methods to manage additional complexity

- More complex teams
- Longer time scales and multiple tasks or missions
- Group and machine learning capabilities
- New collaborative modeling techniques



Images from A. Becheru et al., Analyzing Students' Collaboration Patterns in a Social Learning Environment Using StudentViz Platform, 2018

Use quantitative modeling methods to determine optimal team structures and properties



Teams can develop emergent structures and properties that can influence and potentially optimize decision making, problem solving, task delegation, and overall team functionality

Image from A. Pentland, The New Science of Building Great Teams, 2012

Apply insights from the HMT framework to challenges in machine-machine coordination

Increased swarm capabilities

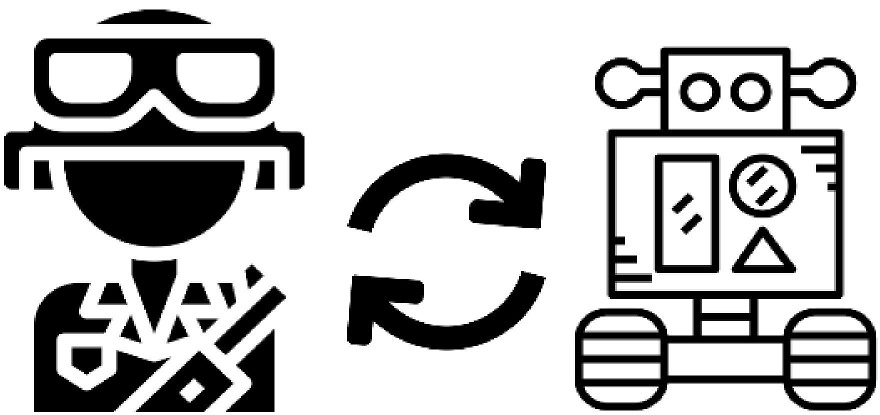
- Independent, but can coordinate with team (and operators)
- Deployable in more complex arrangements
- Deployable across larger range of missions

Future issues in swarms

- Systems integration issues can compound with larger or more complex swarms
- Particular challenges include:
  - Coordinated mobility
  - Communications limitations (bandwidth, physical separation)
  - Increased potential for emergent behavior

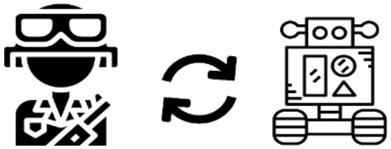
BACKUP

DEFINITIONS



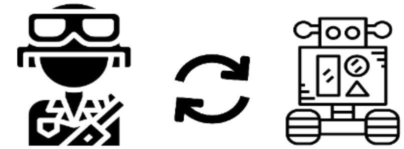
TEAM MEASURE DEFINITIONS

Capability



Team abilities and capacities that are relevant to a particular application environment and the specific applications (i.e., missions) the operator must perform

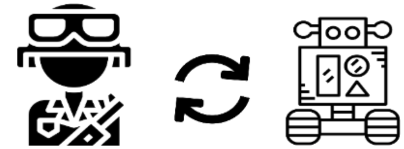
Capability: Communication Framework



Communication framework – The timeliness, capacity, dynamics, precision, and accuracy of information transfer within the team.

- **Information flow** includes the mechanics and capacity for data retention and transfer, including channels, directions, rates, control, interference, and system input-output (Burnham & Anderson, 2002; Arndt, 2004).
- **Joint world model** is the extent to which agents in the team have compatible knowledge structures (compatibility in what they can and cannot represent). Structures need not be alike (e.g., they may be complementary) but they must allow for collaboration with the other agent (Madni & Madni, 2018; Damacharla et al., 2018).
- **Joint mission knowledge** is the compatibility or complementariness of agents' knowledge of the mission, including a strategy for success and a basis for prioritizing competing goals or actions. The team members' knowledge need not be identical (even completely consistent), but must enable collaboration for mutual support in pursuing the mission. Example areas likely to be key include red and blue team tactics, techniques, and procedures (TTPs), standard operating procedures (SOPs), and target weak points (Canonico, 2019).

Capability: Structure-Based Interaction

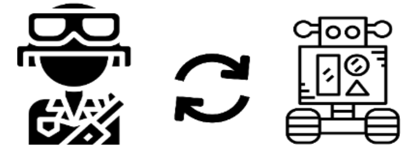


Structure-based interaction – Team agent interactions that are determined and influenced by team organization and architecture.

- **Influence** is the ability of one team agent to elicit a response (cognitive, verbal, emotional, or behavioral) from other teammates (Crichton & Flin, 2004).

- **Role clarity and adaptability** includes awareness, maintenance, and—when necessary—adaptation of action boundaries (within the team and external to it), assigned roles and duties, action, and communication interfaces. Flexibility within teams to recognize the need for adjustment and the ability to implement necessary adjustment (Madni & Madni, 2018; Damacharla et al., 2018). Role clarity under all circumstances and adaptability (if present as a feature) will be of particular importance for the interactions within the team.

Capability: Architecture



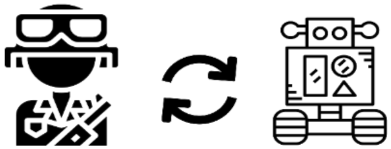
Architecture – Group structure/architecture of team: imposed, implicit, or dynamic.

- **Hierarchical relationships** are the directional influence and command architecture(s) within the team, implied or imposed (Madni & Madni, 2018; Damacharla et al., 2018; Klein et al., 2006; Greening et al., 2015; Greening, 2014; Moore et al., 2012).

- **Uni- and bi-directional relationships** are the characterization of directional command, influence, communication, or action relationships between or among two or more team agents (Damacharla et al., 2018; Greening et al., 2015; Greening, 2014; Moore et al., 2012).

- **Learning patterns** represent the method(s) by which the team assimilates, applies information, adapts, and evolves over time in response to task and team knowledge architecture (Greening et al., 2015; Greening, 2014; Jiang, 2007; Moore et al., 2012).

Interaction



How team members engage in coordination, cooperation, and efficient goal pursuit during execution

Interaction: Perspective

Perspective – What the team knows and/or believes about the world.

- **Situation awareness** is the extent to which a group perceives and distributes information regarding knowledge of collective group actions, tasks, or decisions based on the group's (1) architecture, (2) interaction dynamics, (3) comprehensive environment, and (4) effects, consequences, and influences of environmental factors (Endsley, 1995b). Includes contextual awareness; that is, the team's assessment of objective factors and their potential effects on a task and its completion potential – including physical, social, cultural, economic, and political environments (i.e., the social terrain), friendly vs. adversarial scenarios, etc. (Madni & Madni, 2018; Damacharla et al., 2018).

- **Information accuracy** is the correctness and precision of information passed between team members (Burnham & Anderson, 2002; Arndt, 2004).

Interaction: Cooperative Behavior

Cooperative behavior – The extent to which the team works together to attain a mutual goal or complementary goals (see cooperation definition in *APA Dictionary of Psychology, 2007*).

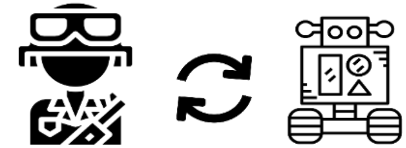
- **Cohesion** is the dynamic process that is reflected in the tendency of a group to remain united in the pursuit of its instrumental objectives and the satisfaction of member needs (Damacharla et al., 2018).

- **Intervention** is the extent to which teammates interact with and affect with each others' actions (Damacharla et al., 2018). Interruption of or direct interaction with a teammate's actions may have a negative impact on overall team performance but it may also be necessary to resolve errors (Damacharla et al., 2018). Previous research indicates that the relationship between intervention and performance is non-monotonic; it is an inverted u-shaped curve where some interactions improve performance but too many hurt it (Damacharla et al., 2018).

- **Team trust** is the attitude that combined group effort "will help achieve [a team's] goals in a situation characterized by uncertainty and vulnerability" (Lee & See, 2004, p. 54). Lee and See's (2004) three general bases of trust in the context of automation (performance, process, and purpose) are likely to play a role in human-machine teams. Based on previous trust in automation research, levels of trust may be ill-calibrated (i.e., overly trusting or insufficiently trusting) due to erroneous judgments of a teammate's performance, process, or purpose (Madhavan & Wiegmann, 2004; Wickens et al., 2000).

- **Agency shifting** involves how and when roles and control in the team change, as well as how the team adapts to filling different roles, having different levels of control, and changes in roles and control (DeRue & Morgeson, 2007).

Interaction: Team Resource Allocation



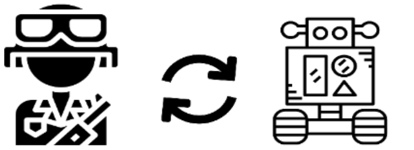
Team resource allocation – Distribution and changing levels of cognitive, physical, and incentive factors among team members.

- **Joint attention allocation** measures the distribution of limited attention paid collectively to the team's tasks and the ability of the team to dynamically attend to and prioritize varied responsibilities such as strategic planning and current assignments (Madni & Madni, 2018; Damacharla et al., 2018).

- **Workload transfer** involves any general process by which one or more tasks are transferred from one subset of team members to another; similar to but more general than human-machine/machine-human intervention (Madni & Madni, 2018; Damacharla et al., 2018).

- **Endurance** is a measure of a team's collective capability to proceed with a task when under stress (Madni & Madni, 2018; Damacharla et al., 2018).

Team Performance



Assessment of decisions, results, and subsequent effects generated by or attributable to team action

Performance definition from APA Dictionary of Psychology, 2007

Team Performance: Collective Decision Making

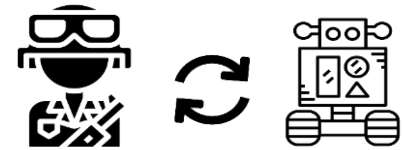
Collective decision making – A team's assessment of objective factors and their potential effects on the task and its completion potential.

- **Optimality** is the extent to which a decision maximizes utility and efficiency. That is, the extent to which the course of action taken maximizes the "value" of a decision. This is usually calculated by a combination of the values of the range of outcomes and the probability of that outcome occurring, given a decision (adapted from Edwards, 1961; Becker & McClintock, 1967). Notably, forming decisions according to statistical and mathematical models of optimal decision making is often complex and time-consuming, and it requires more information than is usually available in real-life decision making (Becker & McClintock, 1967, Kurz-Milcke & Gigerenzer, 2007).

- **Risk level** is extent to which the team's collective decisions or actions take into account the perceived threat of a given situation and the vulnerability expected from taking a certain action (Krokhmal et al., 2003).

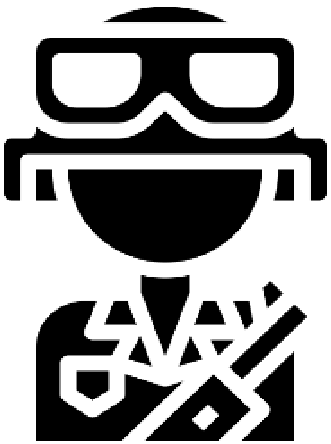
- **Robustness** is a measure of how effective the team's decision-making framework is at choosing actions for different (types of) situations (Krokhmal et al., 2003). The extent to which the decision-making process "takes into account the stochastic nature of risk-inducing factors, and generates decisions that are not only effective on average (in other words, have good 'expected' performance), but also safe enough under a wide range of possible scenarios" (Krokhmal et al., 2003 p. 2; see also Gigerenzer & Gaissmaier, 2015).

Team Performance: Collective Task Performance



Collective task performance – The quality of team productivity, efficiency, and other important outcome measures for a given task.

- **Task success/failure** is the team's accuracy on a given task, regardless of resource use.
- **Timeliness** is the time to complete the task or a ratio of "time focused on task" to "time assigned to task" (Chen & Chen, 2011).
- **Planning recognition** is the ability of the team to identify (one or more possible sets of) steps and subtasks necessary for the completion of the task as a whole (Bolia et al., 2006).
- **Efficiency** is a measure of how effectively the team performs while using the least amount of resources within its resource constraints (Ho, 1972).



HUMAN MEASURE DEFINITIONS

Capability

A human agent's physical and psychological abilities and capacities that are relevant to a particular environment and the missions the agent must perform

Wang et al., 2010

Capability: Training and Experience

Training and experience – Relates to an operator's psychological and physical readiness for a mission, including knowledge of the mission, knowledge of mission systems and agents, natural talent, and prior interaction with other agents and systems that impacts the operator's ability to effectively perform the mission.

- **Mental model** is an internal representation of knowledge (the world, teammates, objects, processes), which can then be manipulated for reasoning processes (Johnson-Laird, 1989). While this concept encapsulates all types of knowledge, for test and evaluation purposes this can only be imperfectly estimated, with different communities using different terms (discussed in Staggers & Norcio, 1993).

- **Mission knowledge** is the range of one's information about and understanding of the mission (knowledge). Consists of representations within one's mental model allowing for flexible mission accomplishment (e.g., red and blue team TTPs and SOPs), target information (e.g., weak points, identifiers), strategies for mission success, and a basis for prioritizing competing goals or actions, among other components (Dept. of the Army, 2019).

- **Teammate knowledge and experience** is the range of one's understanding of or information on the machine teammate (knowledge). This might include operational guidelines and capabilities, system specifications, and system troubleshooting routines (c.f. Ericsson et al., 2018). Also includes the operator's ability or proficiency with the system (human interacting with machine), acquired through previous training or practice with the system (skill; c.f. Ericsson et al., 2018). This will be related primarily to their time and use of the relevant system(s) in the past, including the range of domains, functions, contexts, etc.

Capability: Psychological Traits

Psychological traits – "Relatively *enduring* characteristics" that "appreciably *differentiate* the individual from others" (Anastasi, 1948, emphasis in original). Traits are expected to manifest in the observable behavior of individuals; however, behavior also is impacted by the situation and by the individual's state ("systematic changes in the condition of the organism over time periods of moderate duration") (Anastasi, 1983, p. 348).

- **Decisiveness** is the tendency to engage in the decision-making process, particularly a tendency to simplify information which simplifies making decisions regardless of the quality of the decision (Kruglanski & Webster, 1996; Neuberg & Newsom 1993; Roets & Hiel, 2007; Wichary et al., 2008; Weissman, 1976). Decisiveness has also been related to need for closure (NFC) (Webster & Kruglanski, 1994) as a tendency to make decisions quickly in order to mitigate uncertainty. However, at least one study has disputed decisiveness as a facet of NFC (Kossowska, Van Hiel, Chun, & Kruglanski, 2002), and Roets and Hiel's (2007) findings indicate that decisiveness can be operationalized as both an ability (more closely represented by Weissman's 1976 conceptualization) and a need (consistent with a facet of the NFC). Decisiveness can be generalized or be context- or domain-specific (Potworoski, 2010), so selection of the appropriate scale needs to be based on researcher's intent.

- **Impulsiveness** or impulsivity "is the tendency to be impulsive, spontaneous, and careless as opposed to controlled, reflective, and cautious" (Tellegen, 1982). Dickman (1990) distinguishes two kinds of impulsivity: dysfunctional impulsivity and functional impulsivity. "Dysfunctional impulsivity is the tendency to act with less forethought than most people" in a way that results in poor decision making and increased difficulty (Dickman, 1990, 95). Functional impulsivity "is the tendency to act with relatively little forethought when such a style is optimal" (Dickman, 1990, 95). More recent work characterizing impulsiveness is in Whiteside and Lynam (2001) and MacKillop et al. (2016).

Capability: Psychological Traits

Flexibility includes the ability to choose among approaches to tasks in the execution trade-space, including the use of multi-tasking (e.g., taking a different approach when the current one is not succeeding).

- **Multi-tasking** is the extent to which the operator is capable of engaging in concurrent or switching activities and their likelihood of engaging in such activities. Multi-tasking capability and frequency of multi-tasking engagement have been found to be negatively correlated (see Sanbonmatsu et al., 2013; Watson & Strayer, 2010). For a thorough discussion of multi-tasking, see Salvucci and Taatgen, 2010 (Janssen provides a summary in her 2012 review).

- **Adaptability** is "the capacity to make appropriate responses to changed or changing situations; the ability to modify or adjust one's behavior in meeting different circumstances of different people" (VandenBos, 2007, p. 18). See Martin, 2017, for further discussion of the conceptual bounds of this construct and Martin et al., 2012, for validation of an adaptability scale.

- **Intelligence** is the ability to derive information, learn from experience, adapt to changing environments, understand, and correctly utilize thought and reason (definition of intelligence adapted from VandenBos, 2007, *APA Dictionary of Psychology*). See also Sternberg, 2005 and 2019, and McGrew, 2009, for psychometric approaches to intelligence. There is no generally agreed-upon definition of intelligence; here, it refers to the many executive and related functions (e.g., fluid intelligence, Raven et al., 1988; and cognitive flexibility, Monsell, 1996) that may impact T&E, not just Gf-Gc (fluid intelligence-crystalized intelligence) theory (Horn & Noll, 1997).

Capability: Physical Abilities

Physical abilities – The capability to perform some physical activity. Physical ability is defined by the requirements of a given task, occupation, or mission—one may possess physical ability to excel at performing in one occupation but lack the physical ability to perform in another (Hogan, 1991).

- **Physical fitness** includes strength, speed, and endurance.
  - **Strength** is the degree to which muscular force can be applied to perform a task. Physical strength generally is increased through repeated resistance training, and maximal strength can be measured by one's highest possible load (Friedl et al., 2015; Hogan, 1991). Improved strength is thought to reduce the risk of physical injury (Friedl et al., 2015).
  - **Speed** is the velocity at which an agent is able to progress from one physical location to another. Speed has been found to be closely related to agility and jumping ability among competitive-level, young sports athletes, indicating that these capabilities stem from the same physical attributes (Negra et al., 2017).
  - **Endurance** is the degree to which physical effort can be sustained or prolonged (Hogan, 1991). Physical muscle endurance can be increased through high repetitions of high loads (Friedl et al., 2015). Endurance generally is measured by the total time for which a particular activity can be maintained (Hogan, 1991; Vago et al., 1987).

- **Sensors (organs and equipment)** includes all input devices (e.g., sense organs) and capabilities of equipment, such as radar, chemical detectors, pressure sensors, video cameras, and microphones, among others. This includes sensors supporting the cognitive teaming function.

Worldview

The agent's resource-constrained processes to represent the current and changing state(s) of the environment(s), and the representation that results

Simulation Interoperability Standards Organization, 2007

Worldview: Judgment and Attitudes

Judgment and attitudes – Judgment refers to the process by which "multiple, fallible, incomplete, and sometimes conflicting cues" are integrated "to infer what is happening in the external world" as well as the inferences made by that process (Hastie, 2001, p. 657). Attitudes are "affective evaluation[s] of beliefs that guides people to adopt a particular intention" (Lee & See, 2004, p. 53). Attitudes are based on the informational foundation of perceptions and beliefs, but they are distinct from that foundation; similarly, attitudes guide human intentions, which translate into behaviors, but attitudes are not equivalent to the intentions or behaviors they influence (Lee & See, 2004).

- **Situational awareness** as defined by Endsley (1995, p. 65) "is the perception of environmental elements and events with respect to time or space, the comprehension of their meaning, and the projection of their future status."

- **Trust** is "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability" (Lee & See, 2004, p. 54). Lee and See's (2004) three general bases of trust in the context of automation (performance, process, and purpose) are likely to play a role in human-machine teams. Based on previous trust in automation research, levels of trust may be ill-calibrated (i.e., overly trusting or insufficiently trusting) due to erroneous judgments of a machine teammate's performance, process, or purpose (Madhavan & Wiegmann, 2004; Wickens et al., 2000).

Worldview: Cognitive Allocation

Cognitive allocation – "A state in which cognitive resources are focused on certain aspects of the environment rather than on others, and the central nervous system is in a state of readiness to respond to stimuli. Because it has been presumed that human beings do not have an infinite capacity to attend to everything—focusing on certain items at the expense of others—much of the research in this field has been devoted to discerning which factors influence attention and to understanding the mechanisms that are involved in the selective processing of information. For example, past experience affects perceptual experience (we notice things that have meaning for us), and some activities (e.g., reading) require conscious participation (i.e., voluntary attention). However, attention can also be captured (i.e., directed involuntarily) by qualities of stimuli in the environment, such as intensity, movement, repetition, contrast, and novelty" (Attention, as defined in *APA Dictionary of Psychology*; see also McCallum, 2015).

- **Working memory** is the cognitive system or group of cognitive systems that are used to keep necessary information in mind to enable complex tasks such as reasoning, comprehension, and learning (Baddeley, 2010; Engle, 2002). Many measures of working memory have been developed (see Conway et al., 2005). For a comprehensive treatment of working memory, see Conway, 2016.

- **Attention allocation** refers to the focused direction or distribution of limited cognitive resources (Archibald, Levee, & Olino, 2015). Humans do not possess an infinite capacity to attend to stimuli; attention allocation thus captures the direction or distribution of stimuli that cognitive resources are devoted to for a given period of time.

- **Other dependability monitoring** refers to [typically implicit] processes used for health monitoring systems and indicated mitigations; it includes monitoring of teammates. It also includes safety and reliability, legal/moral/ethical considerations, and assurances that warfighters not only will do what you want but also will not do what you do not want.

Worldview: Effort

Effort – The expending of physical or mental exertion.

- **Resource availability** includes workload and fatigue.
  - **Workload** is the portion of operator physical or mental resources that are required to meet task demands (Lysaght et al., 1989); the costs to the operator to accomplish the mission. Workload might include the amount of work such as the number of tasks to complete and the difficulty of those tasks, the time that one has to complete the task, and the subjective experience of the human operator (Cain, 2007; Lysaght et al., 1989). Thus, workload might include objective and subjective measures but often workload is measured using a subjective scale of workload that measures the operator's experience of workload. The most well-known such scale is the NASA-TLX, which includes six subscales measuring different facets of workload (Hart & Steveland, 1988; Hart 2006), but validated single-item measures also exist (i.e., the AFFTC Revised Workload Estimate Scale or ARWES, Ames & George, 1993).
  - **Fatigue** is a "specific form of human inadequacy in which the individual experiences an aversion to exertion" and a sensation of an inability to continue activity (fatigue, physiology). Fatigue may be physical, experienced as a "an unpleasant bodily state, including headaches, tension, and vague pains in muscles and joints," but it may also be cognitive, manifested by an "unfocused mental state (distraction, frustration, discomfort)," or affective, manifested by low mood, tiredness, or lethargy (Hockey, 2013, p. 1). Fatigue can be a consequence of everyday shifts in mood and quality of life; in more extreme cases it can be "felt as physical exhaustion, a total incapacity for any exertion, a profound lack of motivation, or depression" (Hockey, 2013, p. 1).

Worldview: Effort cont.

- **<u>Usability</u>** is generally defined as the ease with which an operator can use a system, but it is a multi-dimensional concept that is associated with five underlying attributes: learnability (or the ease with which a new operator can learn to employ the system), efficiency (the productivity an operator can attain once they learn to use the system), memorability (the ease with which an operator can return to use of the system after a period away, without having to relearn the system), errors (the error rate of the system should be low and it should recover quickly from errors that are made), and satisfaction (operators should be subjectively satisfied by the system; it should be pleasant to use) (Nielsen, 1994). The System Usability Scale (SUS) is generalizable to a wide range of systems and is often used as a subjective measure of usability (Brooke, 1996).

Performance

Assessment of decisions, results, and subsequent effects generated by or attributable to agent action

Performance definition from APA Dictionary of Psychology, 2007

Performance: Decision Making

Decision making – "The entire process of choosing a course of action," including the consideration of alternative courses of action, uncertain conditioning events, and consequences associated with potential outcomes (Hastie, 2001, p. 657).

- **Optimality** is the extent to which a decision maximizes utility. That is, the extent to which the course of action taken maximizes the "value" of a decision. This is usually calculated by a combination of the values of the range of [expected] outcomes and the probability of that outcome occurring, given a decision (adapted from Edwards, 1961; Becker & McClintock, 1967). Notably, forming decisions according to statistical and mathematical models of optimal decision making is often complex and time- consuming, and it requires more information than is usually available in real-life decision making (Becker & McClintock, 1967; Kurz-Milcke & Gigerenzer, 2007).

- **Robustness** is the extent to which the decision-making process "takes into account the stochastic nature of risk-inducing factors, and generates decisions that are not only effective on average (in other words, have good 'expected' performance), but also safe enough under a wide range of possible scenarios" (Krokhmal et al., 2003, p. 2; see also Gigerenzer & Gaissmaier, 2015).

Performance: Decision Making cont.

- **Risk level** is the extent to which a decision involves a gamble on its desired outcome or expected value as opposed to ensuring that desired or expected outcome (Kahneman & Tversky, 1980). Risk-averse decision making is characterized by decisions that show a preference for “the certain prospect (x) over any risky prospect with the expected value x” (Kahneman & Tversky, 1980, p. 264). In contrast, risk-seeking decision making is characterized by a preference for the gamble over its expected value (e.g., choosing to participate in a gamble to receive $100 or nothing with a 50% chance of winning over winning a sure $50). As outlined by prospect theory, humans tend to be risk averse when it comes to positive outcomes or gains—preferring higher certainty even if the gained value is lower (e.g., rather win $45 than have a 50% chance to win $100 and 50% chance to win nothing), but to be risk seeking when it comes to negative outcomes or losses—preferring to maximize the probability of avoiding loss even if the expected value of those decisions are lower (e.g., rather have a 50% chance of losing $100 and 50% chance of losing nothing than a certain loss of $45) (Kahneman & Tversky, 1980).

- **Reliance** is the behavioral act of dependence on another agent for the accomplishment of some goal. Reliance is influenced by trust but not determined by it (Lee & See, 2004; Madhaven & Wiegmann, 2007).

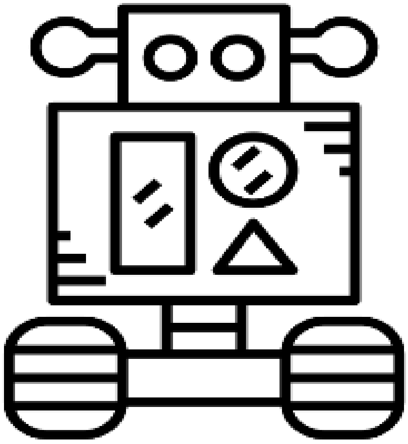
Performance: Task Performance

Task performance –The effectiveness with which agents engage in and complete activities that contribute to the overall mission either directly, such as through implementing an action or decision, or indirectly, such as by providing necessary information or services (Borman & Motowidlo, 1997). Task performance has been distinguished from contextual performance, which influences “effectiveness in ways that shape the organizational, social, and psychological context that serves as the catalyst for task activities and processes” (Borman & Motowidlo, 1997, p. 100). Contextual performance includes such factors as trust (found under worldview), whereas task performance is best measured using objective factors such as those described below.

- **Error rate** is the frequency or duration of deviations from correct task procedure throughout mission accomplishment relative to total tasks or task time performed for the mission. Examples of error rate measures include the number of errors in a given time period and the probability of a miss given the number of shots taken.

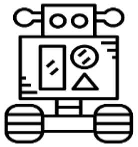
- **Timeliness** is the extent to which to which the agent completes the required task(s) within an allotted amount of time (Cothier & Levis, 1986). Examples of timeliness measures include time to complete [a task] (TTC), time to target (TTT), and time to detect (D_T or T_D).

- **Persistence** (the action of persevering) is the sustainment of effort on a task or towards a goal despite obstacles (persistence); that is, “directed effort extended over time” (Locke et al., 1981). Persistence has a positive connotation and persisting behavior has been shown to enhance performance (Maddi et al., 2012). Despite this, in some instances this behavior can impede completion of a higher goal (Lucas et al., 2015).



MACHINE MEASURE DEFINITIONS

Capability



A system's capability includes the key attributes of the system as a function of its application environment and specific applications (e.g., missions)

Wang et al., 2010

Capability: Cognition Structures and Algorithms

Cognition structures and algorithms – The knowledge structures and processes that the machine uses to represent information—stored and dynamic from sensors and communications—about the world, mission, and teammate(s). These capabilities inform task selection, task accomplishment, teaming, and communication, among related performance metrics.[1] While there are many ways to computationally represent and store the information to be used by the machine, here we build upon Laird's (2012) Soar model as an example, with different types of knowledge stored in networked structures, similar to expert systems (Buchanan et al., 2018). While these may or may not be explicit or "explainable" (Samek et al., 2019), we note that during T&E critical capabilities representing the world, the mission, and teammate(s) must be testable.

- **World model** provides the structure for the machine knowledge (what it can and cannot represent, Laird, 2019). The key point is that the structure is not required to be complete or high fidelity—the structure must only be sufficient to support collaboration with any teammate(s) operating with a similar and/or different worldview.

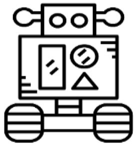
- **Mission knowledge** is the range of the system's information about and understanding of the mission (knowledge) within an information representation (c.f. Laird, 2019). Consists of representations allowing for mission support broadly defined, in and beyond breadth of Joint Forces Command, joint command, and operation doctrines (e.g., U.S. Army, 2017, including Red and Blue team TTPs and SOPs), target information (e.g., weak points, identifiers), strategies for mission success, a basis for prioritizing competing goals or actions, etc.

[1] As such, this is analogous to the training and experience, and to some extent the psychological traits, in the human. Unlike the human, a machine needs to be provided a structure or ontology (usually explicit) that allows for interaction with humans.

Capability: Cognition Structures and Algorithms cont.

- **Teammate knowledge and experience** is made up of representations allowing the machine to coordinate activities with any teammate(s). These include activities up to and beyond the relevant agent's capabilities and procedures: training and skills, mission orientation, best and most appropriate communication method(s), etc. Potentially, these activities could be updated as they are executed. Also includes the machine's ability or proficiency with the human acquired through previous training or practice with the human. This proficiency is related primarily to the machine's time and "experience" (possibly as a form of learning data), including the range of domains, functions, contexts, etc.

Capability: Cognition Software and Hardware



Cognition software – Software that executes or supports higher-level capabilities (cognitive algorithms). Cognition software includes processes that are analogous to lower-level human cognitive functions and that allow processing to occur (e.g., operating system, firmware, transport bus procedures, and other basic processes for executing computational processing) (Sandini et al., 2007). Controller and interface functions may be separated into software and hardware components during developmental T&E, but typically not during operational T&E.

- **Prioritization** is the process of deconflicting or deferring competing subsystem claims for cases of CPU, memory, bandwidth overload, etc. Given limits on system resources, rules or algorithms are required to decide where resources will be allocated, and these rules may depend on the system and mission (Wang et al., 2010).

- **Algorithm flexibility** is the ability to choose among algorithms in the execution trade-space (e.g., use a lower performing but lower latency algorithm when latency becomes an issue); the visible portions acting on the world are built into the action selection mechanisms within the agent architecture (Bryson, 2000). We are scoring this as a “controller” function rather than a “cognitive” function, thus its inclusion in cognition software.

- **Operating system/memory** includes the software, hardware, and firmware enabling computations, as well as any “controller” software specific to autonomous operation or teaming. This can include a range of processes for accessing, interfacing with, and engaging with hardware components.

Capability: Standard Platform

Standard platform – Platform hardware features make it possible for the machine to receive, process, and transmit information and to interact with the world, including aspects required for information integration (c.f. Sage & Lynch, 1999). These features allow for computation, sensing, communications, and other interactions, including movement, engines, tires, wings, radar, audiovisual sensing, etc.

- **Structural and mechanical elements** include structural components such as frames, bearings, and axles; control mechanisms such as gear trains, brakes, and engines; control components such as actuators and controllers for other systems; and sensors that allow for integration (Lelikov, 2009; automotive applications discussed in Fleming, 2001).

- **Sensors** include all input devices not part of integration hardware or computer and peripherals (e.g., radars, chemical detectors, pressure sensors, video cameras, and microphones). These include sensors supporting the cognitive teaming function.

- **Computer and peripherals** involve additional mechanics not related to computation software. These can include hard drives, I/O hardware, etc. for platform operation. In some architectures, the teaming-specific function is handled by processors that also handle platform operation (Whitten & Bentley, 2005).

- **Integration hardware** overlaps with I/O hardware, wireless capabilities, and other items needed for successful teaming.

Worldview

The agent’s resource-constrained processes to represent the current and changing state(s) of the environment(s), and the representation that results

Simulation Interoperability Standards Organization, 2007

Worldview: Perspective

Perspective – What the machine knows and/or believes about the world. This is consistent with the technical definitions from art or drawing—with explicit acknowledgment that a perspective can be misleading. This largely maps to the *sense* portion of robotics' "sense-plan-act" design philosophy (Siciliano & Khatib, 2008), which allows for planning and acting (decision making in this framework).

- **Situational awareness** is "the perception of environmental elements and events with respect to time or space, the comprehension of their meaning, and the projection of their future status" (Endsley, 1995, p. 65).

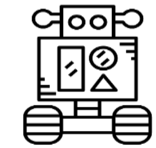
- **Physical world** refers to knowledge and processing about the machine's environment, primarily used for physical navigation (i.e., locomotion and wayfinding; Montello, 2005), to include agents other than teammates. It includes but is not limited to terrain, weather, where red units and structures are, and possible threats (Anderson et al., 2016).

- **Internal** refers to the machine's awareness of its own design performance and deviations therefrom enabled by sensors, integration, and other capabilities (Anderson et al., 2016).

- **Cyber situational awareness** is the collection and processing of data to provide an understanding of cyber health or compromise (Jajodia et al., 2010; reviewed in Franek & Brynielsson, 2014), including the Identify, Protect, Detect, Respond, Recover framework (NIST, 2018).

- **Teammate** refers to the machine's knowledge of the human: location, activity, health, processing capability, etc. (Chakraborti et al., 2017)

Worldview: Resource Allocation



Resource allocation – Processes that enable resource claims on machine operation (e.g., CPU usage, sensor operation), especially those processes that enable teaming (e.g., explore-exploit tradeoffs in task completion and communication) (Wilson, 1996).

- **Activity execution** refers to the subsystems the machine employs to organize, schedule, and integrate tasking of various OSs, devices, algorithms, etc. (i.e., processor management to larger control-systems, parallel to human "working memory"). Can be among a range of techniques, including dynamic programming, reinforcement learning, combinatorial optimization (Busoniu et al., 2010), and lower-level task scheduling (Cerotti et al., 2015).

- **Process/threat monitoring** denotes the I/O streams that feed normal operations and may be associated with cyber intrusions, such as computer, user, access, SIEM, externals, certificates, credentials, and other details for security assurance (Splunk, Snort, Wireshark) (passive and active sources in Curry et al., 2013). *Note:* Machine Cyber SA above is the awareness that results.

- **Other dependability monitoring** refers to health monitoring systems and indicated mitigations; includes machine monitoring of humans, if data are available to machine. Also includes safety and reliability, legal/moral/ethical considerations, and assurances that the machine not only will do what you want but also will not do what you do not want.

Note 1. The output of the resources allocated to monitoring feeds the situational awareness under "perspective."
Note 2. "Allocation," as opposed to "use" which follows, is biased toward the cognitive rather than the physical resources, allowing this to be more parallel with the human.

Worldview: Resource Use

Resource use – Resource utilization in machine operation (electrical power, CPU usage, sensors, integration of inputs). Measurements required for the machine to function include the following (most of these are lower level processes that permit the dependable functioning of complex mechanical systems; Siciliano & Khatib, 2008):

- **Platform operating margin** refers to SWaP-C limits; analogous to limits on non-autonomous systems.
- **Processing** refers to CPU and memory margins (thermal and power constraints on processing could go here or line above).

Note 1. The monitoring in “resource allocation” is a “demand” signal on resource use for the sensors and other subsystems. Resource use is the provided resources, whatever the demand.

Performance

Assessment of decisions, results, and subsequent effects generated by or attributable to agent action

Performance definition from APA Dictionary of Psychology, 2007

Performance: Decision Making

Decision making/action – "The entire process of choosing a course of action," including the consideration of alternative courses of action, uncertain conditioning events, and consequences associated with potential outcomes (Hastie, 2001, p. 657; various autonomous forms discussed in Veres et al., 2011). Also includes the enacted choice itself.

- **Optimality** is the extent to which a decision maximizes utility. That is, the extent to which the course of action taken maximizes the [expected] value of a decision. This is usually calculated by a combination of the values of the range of outcomes and the probability of that outcome occurring, given a decision and available information (adapted from Edwards, 1961; Becker & McClintock, 1967).

- **Robustness** is the extent to which the decision-making process "takes into account the stochastic nature of risk-inducing factors, and generates decisions that are not only effective on average (in other words, have good 'expected' performance), but also safe enough under a wide range of possible scenarios" (Krokhmal et al., 2003 p. 2; see also Gigerenzer & Gaissmaier, 2015).

Performance: Decision Making/Action cont.

- **Risk level** is the extent to which a decision involves a gamble on its desired outcome or expected value as opposed to ensuring that desired or expected outcome (Kahneman & Tversky, 1980). Risk-averse decision making is characterized by decisions that show a preference for “the certain prospect (x) over any risky prospect with the expected value x” (Kahneman & Tversky, 1980, p. 264). In contrast, risk-seeking decision making is characterized by a preference for the gamble over its expected value (e.g., choosing to participate in a gamble to receive $100 or nothing with a 50% chance of winning over winning a sure $50). Design of autonomous machines can attempt to bias the machines’ decision making in either risk-tolerant or risk-averse fashions, leading calibration to be important for T&E. Incomplete understanding of contributors to the decision making by designers, commanders, or teammates may result in a risk tolerance incompatible with the commander’s intent. This is a subject for T&E determination.

- **Reliance** is the behavioral act of dependence on another agent to accomplish some goal.

Performance: Task Performance

Task performance – The effectiveness with which agents engage in and complete activities that contribute to the overall mission either directly (such as through implementing an action or decision) or indirectly (such as by providing necessary information or services) (Borman & Motowidlo, 1997). Task performance has been distinguished from contextual performance, which influences "effectiveness in ways that shape the organizational, social, and psychological context that serves as the catalyst for task activities and processes" (Borman & Motowidlo, 1997, p. 100). Contextual performance includes such factors as situational awareness of teammates (found under worldview), whereas task performance is best measured using objective factors such as those described below.

- **Error rate** is the frequency or duration of deviations from correct task procedures throughout mission accomplishment relative to total tasks or task time for the mission.

- **Timeliness** is the extent to which to which the agent completes the required task(s) within an allotted amount of time (Cothier & Levis, 1986).

- **Persistence** (the action of persevering) is the sustainment of effort on a task or towards a goal despite obstacles (persistence). Persistence has a positive connotation and persisting behavior has been shown to enhance performance (e.g., demonstrated in humans in Maddi et al., 2012). Despite this, in some instances this behavior can impede completion of a higher goal (Lucas et al., 2015).

REFERENCES

Ackerman, P. L., Beier, M. E., & Bowen, K. R. (2002). What we really know about our abilities and our knowledge. *Personality and Individual Differences*, *33*(4), 587–605. https://doi.org/10.1016/S0191-8869(01)00174-X

Adams, B. D., & Webb, R. D. (2002). Trust in small military teams. In *7th International Command and Control Technology Symposium* (pp. 1-20).

Aircraft Control—An overview | ScienceDirect Topics. (n.d.). Retrieved June 9, 2020.

Allport, A. (1989). Visual attention. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 631–682). Cambridge, MA: MIT Press.

Ames, L. L., & George, E. J. (1993). *Revision and Verification of a Seven-Point Workload Estimate Scale.* Fort Belvoir, VA: Defense Technical Information Center.

Anastasi, A. (1983). Traits, states, and situations: A comprehensive view. In H., Wainer & S. Messick (Eds.), *Principles of modern psychological measurement* (pp. 345–356). New York, NY: Routledge.

Arndt, C. (2001). *Information measures: Information and its description in science and engineering*. New York, NY: Springer Science & Business Media.

Attention | psychology. (n.d.). Encyclopedia Britannica. Retrieved June 3, 2020, from https://www.britannica.com/science/attention

Archibald, L. M., Levee, T., & Olino, T. (2015). Attention allocation: Relationships to general working memory or specific language processing. *Journal of Experimental Child Psychology*, *139*, 83–98.

Baddeley, A. (2010). Working memory. *Current Biology*, *20*(4), R136–R140. https://doi.org/10.1016/j.cub.2009.12.014

Baumeister, R. F. (2016). Toward a general theory of motivation: Problems, challenges, opportunities, and the big picture. *Motivation and Emotion*, *40*(1), 1–10.

Becker, G. M., & McClintock, C. G. (1967). Value: Behavioral decision theory. *Annual Review of Psychology*, *18*(1), 239–286.

Bolia, R. S., Nelson, W. T., Summers, S. H., Arnold, R. D., Atkinson, J. L., Taylor, R. M., Cottrell, R., & Crooks, C. L. (2006). Collaborative decision making in network-centric military operations. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(3), 284–288.

Book review | Elsevier Enhanced Reader. (n.d.). https://doi.org/10.1016/j.cogsys.2011.03.001

Borman, W. C., & Motowidlo, S. J. (1997). Task performance and contextual performance: The meaning for personnel selection research. *Human Performance*, *10*(2), 99–109.

Brooke, J. (1996). SUS: A “quick and dirty” usability scale. *Usability Evaluation in Industry*, 189.

Bruemmer, D. J., Marble, J. L., & Dudenhoeffer, D. D. (2002). Mutual initiative in human-machine teams. *Proceedings of the IEEE 7th Conference on Human Factors and Power Plants*, 7–7.

Burnham, K. P., & Anderson, D. R. (Eds.). (2002). Model selection and multimodel inference: A practical information-theoretic approach (2nd ed.). New York, NY: Springer.

Cain, B. (2007). *A review of the mental workload literature*. Toronto, Ontario, Canada: Defence Research and Development.

Canonico, L. (2019). Human-Machine Teamwork: An Exploration of Multi-Agent Systems, Team Cognition, and Collective Intelligence. *All Dissertations*. https://tigerprints.clemson.edu/all_dissertations/2490

Cantimur, Y., Rink, F., & van der Vegt, G. S. (2016). When and why hierarchy steepness is related to team performance. *European Journal of Work and Organizational Psychology*, *25*(5), 658–673.

Cardwell, M. (2014). *Dictionary of psychology*. New York, NY: Routledge.

Chen, C., & Chen, J. (2011). Timeliness evaluation of task-oriented networked space-based information system. *Journal of Systems Engineering and Electronics*, *22*(4), 621–627.

Conway, A. R. A., Kane, M. J., Bunting, M. F., Hambrick, D. Z., Wilhelm, O., & Engle, R. W. (2005). Working memory span tasks: A methodological review and user's guide. *Psychonomic Bulletin & Review*, *12*(5), 769–786.

Cothier, P. H., & Levis, A. H. (1986a). Timeliness and measures of effectiveness in command and control. *IEEE Transactions on Systems, Man, and Cybernetics*, *16*(6), 844–853. https://doi.org/10.1109/TSMC.1986.4309003

Cowan, N. (2016). *Working memory capacity: Classic edition*. New York, NY: Psychology Press.

Crichton, M. T., & Flin, R. (2004). Identifying and training non-technical skills of nuclear emergency response teams. *Annals of Nuclear Energy*, *31*(12), 1317–1330. https://doi.org/10.1016/j.anucene.2004.03.011

Damacharla, P., Javaid, A. Y., Gallimore, J. J., & Devabhaktuni, V. K. (2018). Common metrics to benchmark human-machine teams (HMT): A review. *IEEE Access*, *6*, 38637–38655.

Datu, J. A. D., Valdez, J. P. M., & King, R. B. (2016). Perseverance counts but consistency does not! Validating the Short Grit Scale in a collectivist setting. *Current Psychology*, *35*(1), 121–130.

DeRue, D. S., & Morgeson, F. P. (2007). Stability and change in person-team and person-role fit over time: The effects of growth satisfaction, performance, and general self-efficacy. *Journal of Applied Psychology*, *92*(5), 1242–1253. https://doi.org/10.1037/0021-9010.92.5.1242

Dickman, S. J. (1990). Functional and dysfunctional impulsivity: Personality and cognitive correlates. *Journal of Personality and Social Psychology*, *58*(1), 95–102. https://doi.org/10.1037/0022-3514.58.1.95

Dunning, D. (2011). The Dunning–Kruger effect. In *Advances in Experimental Social Psychology* (Vol. 44, pp. 247–296). New York, NY: Elsevier.

Edwards, W. (1961). Behavioral decision theory. *Annual Review of Psychology*, *12*(1), 473–498.

Endsley, M. R. (1995). Measurement of situation awareness in dynamic systems. *Human Factors*, *37*(1), 65–84.

Engle, R. W. (2016). Working memory capacity as executive attention. *Current Directions in Psychological Science, 11*(1), 19–23.

Fallahi, M., Motamedzade, M., Heidarimoghadam, R., Soltanian, A. R., & Miyake, S. (2016). Assessment of operators' mental workload using physiological and subjective measures in cement, city traffic, and power plant control centers. *Health Promotion Perspectives*, *6*(2), 96–103. https://doi.org/10.15171/hpp.2016.17

Fatigue | physiology. (n.d.). Encyclopedia Britannica. Retrieved June 10, 2020, from https://www.britannica.com/science/fatigue-physiology.

Federal Aviation Administration. (1991). *Criteria for Autopilot Engagement*. Washington, DC: FAA Aviation Rulemaking Advisory Committee.

Federal Aviation Administration. (2014). *Advisory Circular—Approval of Flight Guidance Systems*. Washington, DC: FAA.

Friedl, K. E., Knapik, J. J., Häkkinen, K., Baumgartner, N., Groeller, H., Taylor, N. A., ... & Nindl, B. C. (2015). Perspectives on aerobic and strength influences on military physical readiness: Report of an international military physiology roundtable. *The Journal of Strength & Conditioning Research*, *29*, S10–S23.

Gigerenzer, G., & Gaissmaier, W. (2015). Decision making: Nonrational theories. In *International Encyclopedia of the Social & Behavioral Sciences* (pp. 911–916). New York, NY: Elsevier. https://doi.org/10.1016/B978-0-08-097086-8.26017-0

Greening, B. R., Pinter-Wollman, N., & Fefferman, N. H. (2015). Higher-order interactions: Understanding the knowledge capacity of social groups using simplicial sets. *Current Zoology*, *61*(1), 114–127.

Greening Jr, B. (2014). *Higher-order analysis of knowledge capacity and learning potential in social animal groups* [Rutgers University - Graduate School - New Brunswick]. https://doi.org/10.7282/T3RB7325

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. *Advances in Psychology* (Vol. 52, pp. 139–183). New York, NY: Elsevier.

Hastie, R. (2001). Problems for judgment and decision making. *Annual Review of Psychology*, *52*(1), 653–683.

Heckhausen, J., & Heckhausen, H. (2018). Motivation and action: Introduction and overview. In J. Heckhausen & H. Heckhausen (Eds.), *Motivation and Action* (pp. 1–14). New York, NY: Springer International Publishing.

Hoc, J.-M. (2000). From human – machine interaction to human – machine cooperation. *Ergonomics*, *43*(7), 833–843.

Hockey, B., & Hockey, R. (2013). *The psychology of fatigue: Work, effort and control*. Cambridge, UK: Cambridge University Press.

Hogan, J. (1991). Structure of physical performance in occupational tasks. *Journal of Applied Psychology*, *76*(4), 495–507.

Jiang, B., & Omer, I. (2007). Spatial topology and its structural analysis based on the concept of simplicial complex. *Transactions in GIS*, *11*(6), 943–960.

Jiang, X. (2010). How to motivate people working in teams. *International Journal of Business and Management*, *5*(10), 7.

Johnson, M., Bradshaw, J. M., Feltovich, P. J., Jonker, C. M., Van Riemsdijk, M. B., & Sierhuis, M. (2014). Coactive design: Designing support for interdependence in joint activity. *Journal of Human-Robot Interaction*, *3*(1), 43. https://doi.org/10.5898/JHRI.3.1.Johnson

Johnson, M., Bradshaw, J. M., Hoffman, R. R., Feltovich, P. J., & Woods, D. D. (2014). Seven cardinal virtues of human-machine teamwork: Examples from the DARPA Robotic Challenge. *IEEE Intelligent Systems*, *29*(6), 74–80. https://doi.org/10.1109/MIS.2014.100

Jones, D. N., & Paulhus, D. L. (2011). The role of impulsivity in the dark triad of personality. *Personality and Individual Differences*, *51*(5), 679–682. https://doi.org/10.1016/j.paid.2011.04.011

Justusson, B. (2017, December 22). *Autopilot: Computerized Type plus Glass Cockpits—Test Flight Birrfeld to Bern*. http://www.justus2.se/flyg/simulator/sim_ap_rt1/sim_ap_glass_cockpit_rt1.htm

Kahneman, D. (1973). *Attention and effort*. Upper Saddle River, NJ: Prentice-Hall.

Kahneman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 263–291.

Klein, K. J., Ziegert, J. C., Knight, A. P., & Xiao, Y. (2006). Dynamic delegation: Shared, hierarchical, and deindividualized leadership in extreme action teams. *Administrative Science Quarterly*, *51*(4), 590–621.

Kossowska, M., Van Hiel, A., Chun, W. Y., & Kruglanski, A. W. (2002). The Need for Cognitive Closure Scale: Structure, cross-cultural invariance, and comparison of mean ratings between European-American and East Asian samples. *Psychologica Belgica*, *42*(4), 267–286.

Krokhmal, P., Murphey, R., Pardalos, P., Uryasev, S., & Zrazhevski, G. (2003). Robust decision making: Addressing uncertainties in distributions. In S. Butenko, R. Murphey, & P. M. Pardalos (Eds.), *Cooperative control: Models, applications and algorithms* (pp. 165–185). New York, NY: Springer.

Kruger, J., & Dunning, D. (1999). Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. *Journal of Personality and Social Psychology*, *77*(6), 1121.

Kruger, J., Wirtz, D., Van Boven, L., & Altermatt, T. W. (2004). The effort heuristic. *Journal of Experimental Social Psychology*, *40*(1), 91–98.

Kurz-Milcke, E., & Gigerenzer, G. (2007). Heuristic decision making. *Marketing: Journal of Research and Management*, *3*(1), 48–56.

Laird, J., Ranganath, C., & Gershman, S. (n.d.). *Future Directions in Human Machine Teaming Workshop*. Arlington, VA: Virginia Tech Applied Research Corporation.

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors; Santa Monica*, *46*(1), 50–80.

Lichtenstein, S., & Fischhoff, B. (1977). Do those who know more also know more about how much they know? *Organizational Behavior and Human Performance*, *20*(2), 159–183.

Locke, E. A., Shaw, K. N., Saari, L. M., & Latham, G. P. (1981). Goal setting and task performance: 1969–1980. *Psychological Bulletin*, *90*(1), 125.

Lucas, G. (2015). *Routledge handbook of military ethics*. New York, NY: Routledge.

Lucas, G. M., Gratch, J., Cheng, L., & Marsella, S. (2015). When the going gets tough: Grit predicts costly perseverance. *Journal of Research in Personality*, *59*, 15–22. https://doi.org/10.1016/j.jrp.2015.08.004

Lysaght, R. J., Hill, S. G., Dick, A. O., Plamondon, B. D., & Linton, P. M. (1989). *Operator workload: Comprehensive review and evaluation of operator workload methodologies.* Fort Belvoir, VA: Defense Technical Information Center.

MacKillop, J., Weafer, J., Gray, J., Oshri, A., Palmer, A., & de Wit, H. (2016). The latent structure of impulsivity: Impulsive choice, impulsive action, and impulsive personality traits. *Psychopharmacology*, *233*(18), 3361–3370.

Maddi, S. R., Matthews, M. D., Kelly, D. R., Villarreal, B., & White, M. (2012). The role of hardiness and grit in predicting performance and retention of USMA Cadets. *Military Psychology*, *24*(1), 19–28.

Madhavan, P., & Wiegmann, D. A. (2004, September). A new look at the dynamics of human-automation trust: Is trust in humans comparable to trust in machines? In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 48, No. 3, pp. 581–585). Los Angeles, CA: SAGE Publications.

Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: An integrative review. *Theoretical Issues in Ergonomics Science*, *8*(4), 277–301.

Madni, A. M., & Madni, C. C. (2018). Architectural framework for exploring adaptive human-machine teaming options in simulated dynamic environments. *Systems*, *6*(4), 44. https://doi.org/10.3390/systems6040044

Martin, A. J., Nejad, H., Colmar, S., & Liem, G. A. D. (2012). Adaptability: Conceptual and empirical perspectives on responses to change, novelty and uncertainty. *Australian Journal of Guidance and Counselling*, *22*(1), 58–81. https://doi.org/10.1017/jgc.2012.8

Mathieu, J. E., Tannenbaum, S. I., Donsbach, J. S., & Alliger, G. M. (2014). A review and integration of team composition models: Moving toward a dynamic and temporal framework. *Journal of Management*, *40*(1), 130–160. https://doi.org/10.1177/0149206313503014

Martin, A. J. (2017). Adaptability—what it is and what it is not: Comment on Chandra and Leong (2016). *American Psychologist, 72*(7), 696–698.

McDermott, P., Dominguez, C., Kasdaglis, N., Ryan, M., Trahan, I., & Nelson, A. (2018). *Human-machine teaming systems engineering guide*. Bedford, MA: MITRE Corporation.

Naquin, C. E., & Kurtzberg, T. R. (2009). Team negotiation and perceptions of trustworthiness: The whole versus the sum of the parts. *Group Dynamics: Theory, Research, and Practice*, *13*(2), 133–150.

McGrew, K. S. (2009). CHC theory and the human cognitive abilities project: Standing on the shoulders of the giants of psychometric intelligence research. *Intelligence*, *37*(1), 1–10.

Mingyue Ma, L., Fong, T., Micire, M. J., Kim, Y. K., & Feigh, K. (2018). Human-robot teaming: Concepts and components for design. In M. Hutter & R. Siegwart (Eds.), *Field and service robotics* (pp. 649–663). Springer International Publishing. https://doi.org/10.1007/978-3-319-67361-5_42

Mohammed, S., & Dumville, B. C. (2001). Team mental models in a team knowledge framework: Expanding theory and measurement across disciplinary boundaries. *Journal of Organizational Behavior*, *22*(2), 89–106. https://doi.org/10.1002/job.86

Moore, T. J., Drost, R. J., Basu, P., Ramanathan, R., & Swami, A. (2012). Analyzing collaboration networks using simplicial complexes: A case study. *2012 Proceedings IEEE INFOCOM Workshops*, 238–243.

Moray, N. (1982). Subjective mental workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *24*(1), 25–40.

Mraz, S. (2014, June 12). What's the difference between pitch, roll, and yaw? *Machine Design*.

Naatanen, R., & Näätänen, R. (1992). *Attention and brain function*. New York, NY: Psychology Press.

Negra, Y., Chaabene, H., Hammami, M., Amara, S., Sammoud, S., Mkaouer, B., & Hachana, Y. (2017). Agility in young athletes: Is it a different ability from speed and power? *Journal of Strength and Conditioning Research*, *31*(3), 727–735. https://doi.org/10.1519/JSC.0000000000001543

Neuberg, S. L., & Newsom, J. T. (1993). Personal need for structure: Individual differences in the desire for simpler structure. *Journal of Personality and Social Psychology*, *65*(1), 113–131.

Nielson, J. (1994). Usability inspection methods. *In Conference Companion on Human Factors in Computing Systems* (pp. 413–414). New York, NY: Association for Computing Machinery. https://doi.org/10.1145/259963.260531

Oakley, B., Felder, R. M., Brent, R., & Elhajj, I. (2004). Turning student groups into effective teams. *Journal of Student Centered Learning*, *2*(1), 9–34.

Ocasio, W. (2011). Attention to attention. *Organization Science*, *22*(5), 1286–1296. https://doi.org/10.1287/orsc.1100.0602

Pashler, H. E. (1999). *The psychology of attention*. Cambridge, MA: MIT Press.

Pavel, M., Wang, G., & Li, K. (2003). Augmented cognition: Allocation of attention. *36th Annual Hawaii International Conference on System Sciences,* Big Island, HI.

Pellerin, C. (2015). Work: Human-machine teaming represents defense technology future. *DoD News*, *8*.

Potworowski, G. A. (2010). *Varieties of indecisive experience: Explaining the tendency to not make timely and stable decisions*. Doctoral dissertation. The University of Michigan.

Roets, A., & Van Hiel, A. (2007). Separating ability from need: Clarifying the dimensional structure of the Need for Closure Scale. *Personality and Social Psychology Bulletin*, *33*(2), 266–280.

Rolfhus, E. L., & Ackerman, P. L. (1996). Self-report knowledge: At the crossroads of ability, interest, and personality. *Journal of Educational Psychology*, *88*(1), 174.

Romme, A. G. L. (1996). A note on the hierarchy–team debate. *Strategic Management Journal*, *17*(5), 411–417.

Salvucci, D. D., & Taatgen, N. A. (2010). *The multitasking mind*. New York, NY: Oxford University Press.

Sanbonmatsu, D. M., Strayer, D. L., Medeiros-Ward, N., & Watson, J. M. (2013). Who multi-tasks and why? Multi-tasking ability, perceived multi-tasking ability, impulsivity, and sensation seeking. *PLOS ONE*, *8*(1), e54402.

Sandini, G., Metta, G., & Vernon, D. (2007). The iCub cognitive humanoid robot: An open-system research platform for enactive cognition. In M. Lungarella, F. Iida, J. Bongard, & R. Pfeifer (Eds.), *50 Years of artificial intelligence: Essays dedicated to the 50th anniversary of artificial intelligence* (pp. 358–369). New York, NY: Springer.

Sarter, N. B., & Woods, D. D. (1995). Autonomy, authority, and observability: Properties of advanced automation and their impact on human-machine coordination. *IFAC Proceedings Volumes*, *28*(15), 149–152.

Schlösser, T., Dunning, D., Johnson, K. L., & Kruger, J. (2013). How unaware are the unskilled? Empirical tests of the “signal extraction” counterexplanation for the Dunning–Kruger effect in self-evaluation of performance. *Journal of Economic Psychology*, *39*, 85–100.

Shah, J., Kim, B., & Nikolaidis, S. (2012, October 19). Human-Inspired Techniques for Human-Machine Team Planning. *2012 AAAI Fall Symposium Series*. 2012 AAAI Fall Symposium Series.

Shaw, R. (2014, June 12). *Dynamics of Flight*. Washington, DC: NASA.

Singer, S., & Akin, D. (2011, July 17). A Survey of Quantitative Team Performance Metrics For Human-Robot Collaboration. *41st International Conference on Environmental Systems*. 41st International Conference on Environmental Systems, Portland, Oregon. https://doi.org/10.2514/6.2011-5248

Solberg, A. (2018, September 12). *System Capability Model* [Text]. NASA. http://www.nasa.gov/consortium/SystemCapabilityModel

Spruyt, A., De Houwer, J., Everaert, T., & Hermans, D. (2012). Unconscious semantic activation depends on feature-specific attention allocation. *Cognition*, *122*(1), 91–95. https://doi.org/10.1016/j.cognition.2011.08.017

Sternberg, R. J. (2005). The theory of successful intelligence. *Interamerican Journal of Psychology*, *39*(2), 189–202.

Sternberg, R. J. (2019). A theory of adaptive intelligence and its relation to general intelligence. *Journal of Intelligence*, *7*(4), 23.

Sukthankar, G., & Sycara, K. (2006). Robust recognition of physical team behaviors using spatio-temporal models. *Proceedings of the Fifth International Joint Conference on Autonomous Agents and Multiagent Systems*, 638–645.

Simulation Interoperability Standards Organization. (2007). *Reference for guide: DIS plain and simple* (Report No. SISO-REF-020-2007). Orlando, FL: SISO.

Tellegen, A. (1982). *Brief manual for the Multidimensional Personality Questionnaire*. Unpublished Manuscript, University of Minnesota, Minneapolis, 8, 1031–1010.

Vago, P., Mercier, J., Ramonatxo, M., & Prefaut, C. (1987). Is ventilatory anaerobic threshold a good index of endurance capacity? *International Journal of Sports Medicine*, *8*(3), 190–195. https://doi.org/10.1055/s-2008-1025654

van Dijk, L., & Withagen, R. (2014). The horizontal worldview: A Wittgensteinian attitude towards scientific psychology. *Theory & Psychology*, *24*(1), 3–18. https://doi.org/10.1177/0959354313517415

VandenBos, G. R. (2007). *APA dictionary of psychology.* Washington, DC: American Psychological Association.

Vidal, C. (2008). Wat is een wereldbeeld? [What is a worldview?] In H. Van Belle & J van der Veken (Eds.), *De wetenschappen en het creatieve aspect van de werkelijkheid* [*The sciences and the creative aspect of reality*]. Leuven, Belgium: Acco.

Wainer, H., & Messick, S. (Eds.). (2012). *Principals of modern psychological measurement: A festschrift for Frederic M. Lord*. New York, NY: Routledge.

Walliser, J. C., de Visser, E. J., Wiese, E., & Shaw, T. H. (2019). Team structure and team building improve human–machine teaming with autonomous agents. *Journal of Cognitive Engineering and Decision Making*, *13*(4), 258–278. https://doi.org/10.1177/1555343419867563

Wang, J. J., Jing, Y. Y., & Zhang, C. F. (2010). Optimization of capacity and operation for CCHP system by genetic algorithm. *Applied Energy*, *87*(4), 1325–1335.

Warner, S., Bowers, M. T., & Dixon, M. A. (2012). Team dynamics: A social network perspective. *Journal of Sport Management*, *26*(1), 53–66.

Watson, J. M., & Strayer, D. L. (2010). Supertaskers: Profiles in extraordinary multitasking ability. *Psychonomic Bulletin & Review*, *17*(4), 479–485.

Webster, D. M., & Kruglanski, A. W. (1994). Individual differences in need for cognitive closure. *Journal of Personality and Social Psychology*, *67*(6), 1049.

Weissman, M. S. (1976). Decisiveness and psychological adjustment. *Journal of Personality Assessment*, *40*(4), 403–412.

Whiteside, S. P., & Lynam, D. R. (2001). The Five Factor Model and impulsivity: Using a structural model of personality to understand impulsivity. *Personality and Individual Differences*, *30*(4), 669–689.

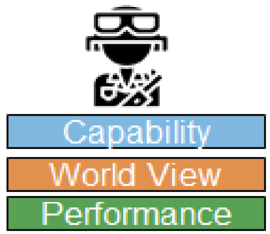
Yeh, M., & Wickens, C. D. (2000). *Attention and trust biases in the design of augmented reality displays*. Fort Belvoir, VA: Defense Technical Information Center.

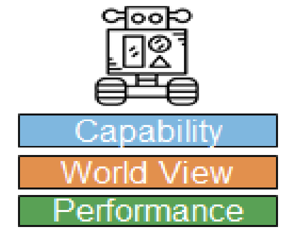
Young, M. S., Brookhuis, K. A., Wickens, C. D., & Hancock, P. A. (2015). State of science: Mental workload in ergonomics. *Ergonomics*, *58*(1), 1–17.

Yu-Chi Ho, & K’ai-Ching Chu. (1972). Team decision theory and information structures in optimal control problems–Part I. *IEEE Transactions on Automatic Control*, *17*(1), 15–22.

BACKUP SLIDES

We are interested in measuring individual human and machine agents on similar concepts





Capability

Physical and psychological abilities and capacities that are relevant to a particular application environment and the specific applications (i.e., missions) the operator must perform

Worldview

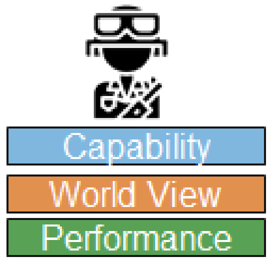
The agent's resource-constrained processes to represent the current and changing state(s) of the environment(s), and the representation that results

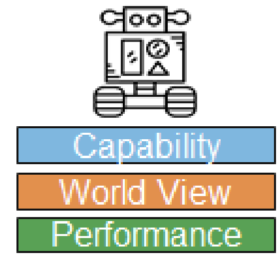
Performance

Assessment of decisions, results, and subsequent effects generated by or attributable to agent action

Performance definition from APA Dictionary of Psychology, 2007; see also Simulation Interoperability Standards Organization, 2007

Joint Force Command doctrine describes HOW you use MEANS to achieve ENDS. Each is important for evaluation.





Capability

MEANS: Abilities of agents of the team to get closer to the goal state. What the agents can do coming into T&E.

Worldview

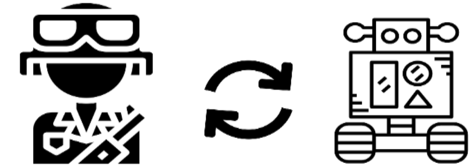
HOW: Agent engagement processes with the environment and the mission. Factors that change during T&E.

Performance

ENDS: Outcomes that characterize mission accomplishment. Large amount of MOEs, MOSs, etc.

Acronyms: MOE – Measure of Effectiveness; MOS – Measure of Success; T&E – Test and Evaluation

Team measures can be assessed only with two or more agents' active involvement



Communication Framework	Structure-Based Interaction	Architecture
Information flow	Influence	Hierarchical relationships
Joint world model	Role clarity	Uni- and bi-directional relationships
Joint mission knowledge	Role adaptability	Learning patterns
Team Perspective	**Cooperative Behavior**	**Team Resource Allocation**
Situational awareness	Cohesion	Joint attention allocation
Information accuracy	Intervention	Workload transfer
	Team trust	Endurance
	Agency shifting	
Collective Decision Making		**Collective Task Performance**
Optimality; Risk level		Task success/failure; Timeliness
Robustness		Planning recognition; Efficiency

Human capabilities, judgments, and behaviors affect interaction with the machine and the mission

Training and Experience	Psychological Traits	Physical Abilities
Mental model	Decisiveness and impulsiveness	Physical fitness
Mission knowledge	Flexibility	Sensors (organs/equipment)
Teammate knowledge and experience	N/A*	N/A*
	Intelligence	N/A*
Judgments and Attitudes	**Cognitive Allocation**	**Effort**
Situational awareness	Working memory	Resource availability (workload, fatigue)
Trust	Attention allocation	Usability
	Other dependability monitoring	
Decision Making		**Task Performance**
Optimality Robustness		Error rates Timeliness
Risk level Reliance		Other performance

*No human equivalent to the machine version of this concept.

Machine capabilities, awareness, and actions affect interaction with the human and the mission

Cognition Structures and Algorithms	Cognition Software and Hardware	Standard Platform
World model	Prioritization	Structural and mechanical elements
Mission knowledge	Algorithm flexibility	Sensors
Teammate knowledge and experience	Operating system	Computer and peripherals
	N/A*	Integration hardware
Perspective	**Resource Allocation**	**Resource Use**
Situational awareness	Activity execution	Platform operating margin
N/A*	Process/threat monitoring	Processing
	Other dependability monitoring	
Decision Making/Action		**Task Performance**
Optimality; Robustness		Error rates
Risk level; Reliance		Timeliness
		Other performance

*No machine equivalent to the human version of this concept.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
09-02-2021

2. REPORT TYPE
Final

3. DATES COVERED *(From - To)*
January 2021 - April 2021

4. TITLE AND SUBTITLE
DATAWorks 2021:
Characterizing Human-Machine Teaming Metrics for Test and Evaluation

5a. CONTRACT NUMBER
HQ0034-19-D-0001

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Heather Wojton; Brian Vickers; Kristina Carter; David Sparrow; Leonard Wilkins; Caitlan Fealing

5d. PROJECT NUMBER
BD-9-2299

5e. TASK NUMBER
2299(90)

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Institute for Defense Analyses
4850 Mark Center Drive
Alexandria, Virginia 22311-1882

8. PERFORMING ORGANIZATION REPORT NUMBER
NS-D-21564

H 2021-000050

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Director, Operational Test and Evaluation
1700 Defense Pentagon
Washington, DC 20301

10. SPONSOR/MONITOR'S ACRONYM(S)
DOT&E

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for Public Release. Distribution Unlimited.

13. SUPPLEMENTARY NOTES
Project Leader: Heather Wojton

14. ABSTRACT
The Department of Defense and partner organizations are developing advanced machine systems that are intended to partner with humans and engage in teamwork to accomplish missions. Given that these human-machine team (HMT) systems have never gone through test and evaluation (T&E), this initial effort looks at the new challenges and provides guidance associated with HMTs. This briefing (1) defines human-machine teaming, (2) describes new challenges in evaluating HMTs, and (3) provides a framework for the categories of metrics that are important for the T&E of HMTs.

15. SUBJECT TERMS
assessment framework; human-machine teaming; test & evaluation

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
Unclassified	Unclassified	Unclassified	Unlimited	114	Heather Wojton
					19b. TELEPHONE NUMBER *(Include area code)* 703-845-6811

Standard Form 298 (Rev. 8/98)
Prescribed by ANSI Std. Z39.18