DEVCOM ARMY RESEARCH LABORATORY

U.S. ARMY COMBAT CAPABILITIES DEVELOPMENT COMMAND – ARMY RESEARCH LABORATORY

Transparent Communications for Effective Human-Autonomy Teaming

Jessie Chen, PhD

ST – Soldier Performance in Socio-Technical Systems

US Army Research Laboratory

APPROVED FOR PUBLIC RELEASE

29 July 2021

CHALLENGES

- Human interaction with increasingly sophisticated systems capable of complex reasoning
  - Challenges that need to be resolved in order to achieve assured autonomy and effective human-autonomy teaming (International Organization for Standardization, 2020; Topcu et al., 2020)
  - Top three requirements for high-stake AI systems: **transparency**, traceability, and human control (European Commission 2020)
  - 9 principles: lawful; purposeful and performance-driven; accurate, reliable, and effective; safe, secure, and resilient; understandable; responsible and traceable; regularly monitored; **transparent**; and accountable (US Executive Order on "Promoting the Use of Trustworthy AI in the Federal Government"; Dec 2020)
  - Geoffrey Hinton (2018 Turing Award recipient): "What we need is for neural nets now to begin to be able to **explain** reasoning"

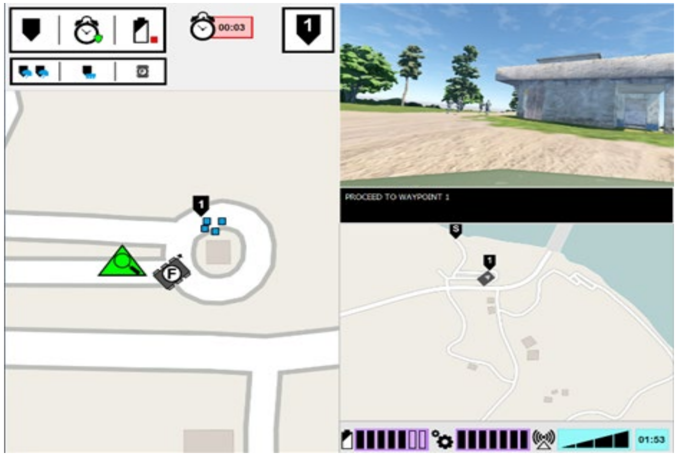
- Military Context
  - Six barriers to human trust in autonomous systems, with 'low **observability**, predictability, directability, and auditability' as well as 'low **mutual understanding** of common goals' being among the key issues (Defense Science Board's *Summer Study on Autonomy,* 2016)

TRANSPARENCY

- Human trust and joint human-system performance
  - Humans interacting with highly automated systems encounter multiple challenges: understanding the current system state, comprehending reasons for its current behavior, and projecting what its next behavior will be (Endsley 1995; Sarter & Woods, 1995).
  - In order to support effective human-autonomy teaming and joint decision making, the human and the machine agent need to understand each other's intent, reasoning, and expected outcomes.
    - Information that the human has but the machine does not have access to (e.g., intelligence reports)
    - Adding or removing constraints
  - Making AI's output more transparent in order to maximize the joint performance of the human-machine team (Matheny et al., 2019; Sanneman & Shah, 2020)
    - DARPA's eXplainable AI (XAI) Program
    - NSF Program on Fairness in Artificial Intelligence (FAI) in Collaboration with Amazon (2020-2023)

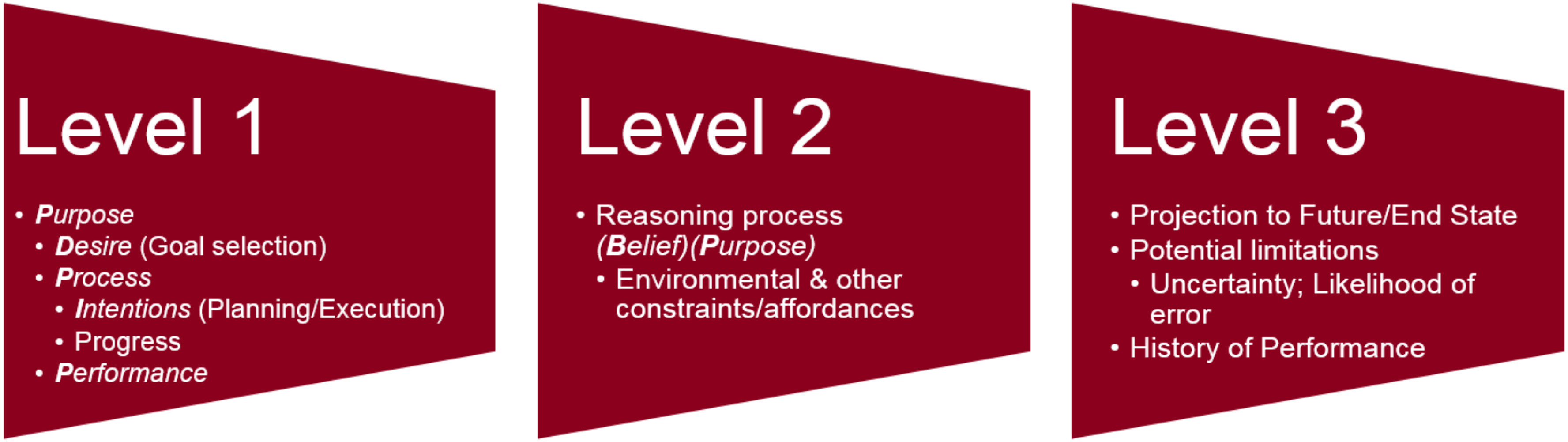
TRANSPARENCY RESEARCH

- ***Transparency Frameworks***
  - Situation awareness-based Agent Transparency (SAT) (Chen et al. 2014, 2018)
  - Human-Robot Transparency (Lyons et al. 2014)
  - Coactive System: Observability, Predictability, Directability (Johnson et al. 2014)
- ***Human-Robot Interaction***
  - Small ground robots (Olatunji et al. 2020; Pynadath et al. 2018; Selkowitz et al. 2017; Wright et al. 2020)
  - Multiagent management via an intelligent planning agent (Bhaskara et al. 2021; Mercado et al. 2016; Stowers et al. 2020; Vered et al. 2020)
  - Robotic swarms (Hepworth et al. 2020; Roundtree et al. 2019; Roundtree et al, 2020)
- ***Automated/Autonomous Driving*** (Krause et al., 2020; Kunze et al. 2019)
- ***Aviation***
  - Emergency landing planning agent (Lyons et al. 2016)
  - Workload management agent in a helicopter cockpit environment (Roth et al. 2020)
- ***Explainable AI*** (Chien et al., in press; Holder & Wang, 2021; Miller, 2019; Sanneman and Shah, 2020)
- ***Individual and Cultural Differences*** (Matthews et al. 2020; Chien et al. 2020)



SA-based **A**gent **T**ransparency (**SAT**)

- Definition of Agent Transparency: *"A quality of an interface pertaining to its abilities to afford an operator's comprehension of an intelligent agent's intent, performance, future plans, and reasoning process"* (Chen et al., 2014)
- Focus on operator task performance and trust calibration



Level 1

- ***Purpose***
  - ***Desire*** (Goal selection)
  - ***Process***
    - ***Intentions*** (Planning/Execution)
    - Progress
  - ***Performance***

Level 2

- Reasoning process (***Belief***)(***Purpose***)
  - Environmental & other constraints/affordances

Level 3

- Projection to Future/End State
- Potential limitations
  - Uncertainty; Likelihood of error
- History of Performance

What's going on and what is the agent trying to achieve?

Why is the agent doing it?

What should the operator expect to happen?

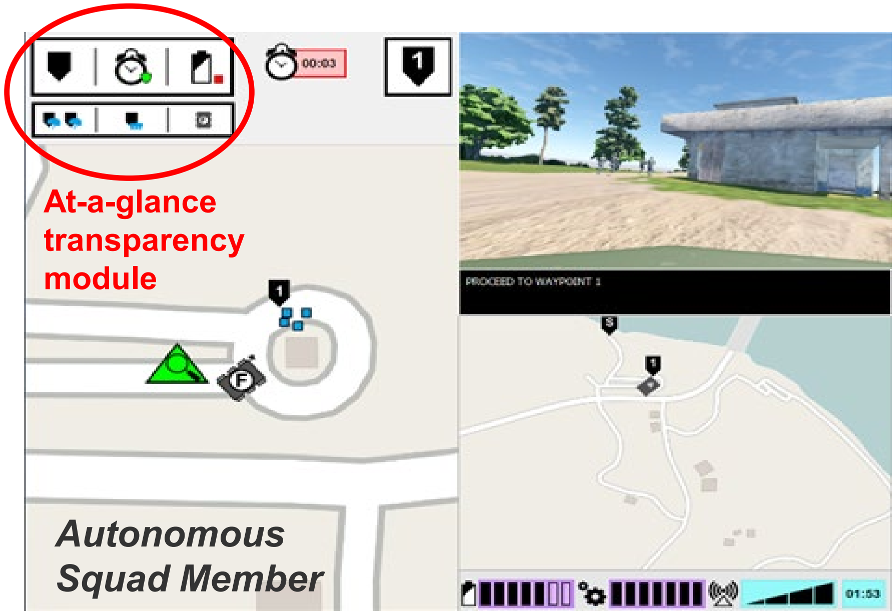
Situation Awareness (SA) (Endsley, 1995)
BDI Agent Framework (Rao & Georgeff, 1995)
Trust calibration (Lee & See, 2004)

Chen, J.Y.C. et al. (2014). *Situation Awareness-based Agent Transparency* (ARL-TR-6905).

Chen, J.Y.C. et al. (2018). Situation awareness-based agent transparency and human-autonomy teaming effectiveness. *Theoretical Issues in Ergonomics Science, 19*(3), 259-282.

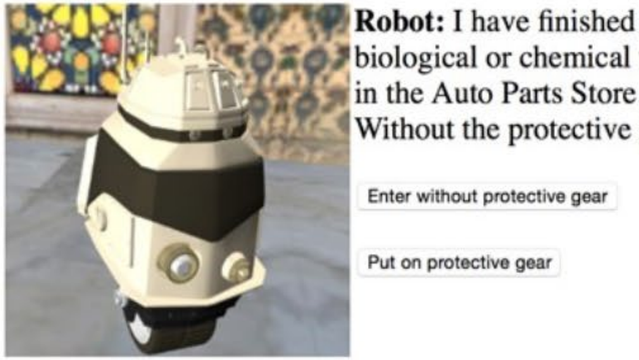
SAT-BASED RESEARCH

- ***Human-Robot Interaction***
  - Human interaction with a small ground robot
    - Robotic support of an infantry squad (Autonomous Squad Member) (Selkowitz et al. 2017; Wright et al. 2020)
    - Robotic support for threat detection (Pynadath et al. 2018)
    - Older adults’ interaction with an assistive robot (Olatunji et al. 2020)



At-a-glance transparency module

Autonomous Squad Member



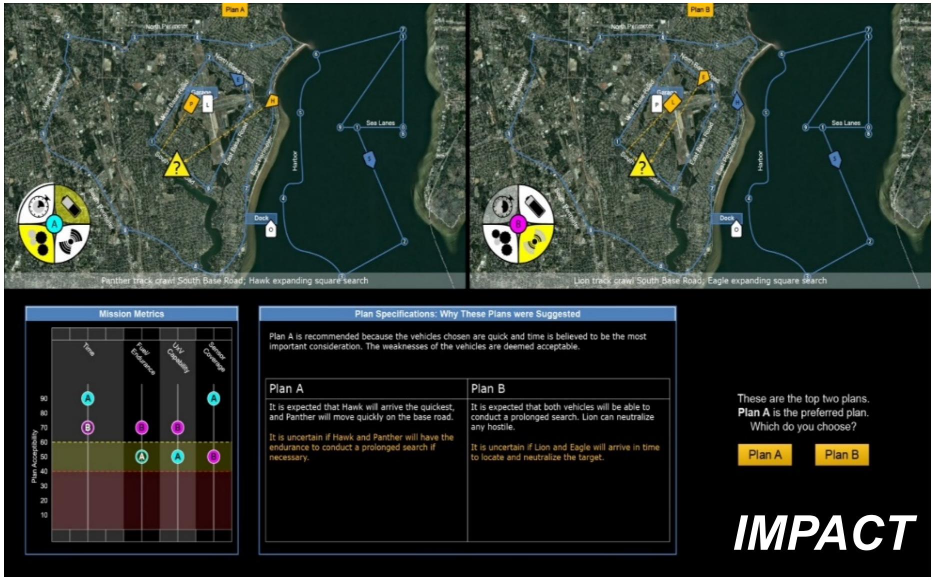
Robot: I have finished surveying the Auto Parts Store. I think the place is safe. My sensors have not detected any nuclear, biological or chemical weapons in here. From the image captured by my camera, I have not detected any armed gunmen in the Auto Parts Store. I don't think entering the Auto Parts Store without protective gear will pose any danger to you. Without the protective gear, you will be able to search the building a little faster.

Pynadath, D. et al. (2018). Transparency communication for machine learning in human-automation interaction. In: *Human and Machine Learning. Human–Computer Interaction Series* (ed. J. Zhou and F. Chen) Springer, Cham.

SAT-BASED RESEARCH

Multiagent Management

- Mission planning involving autonomous aerial and ground vehicles
  - IMPACT (Mercado et al. 2016; Stowers et al. 2020)
  - Defense S&T Group of Australia (Bhaskara et al. 2021)
- Workload-adaptive cognitive agent in helicopter cockpit environments (Roth et al. 2020)



Plan Specifications: Why These Plans were Suggested

Plan A is recommended because the vehicles chosen are quick and time is believed to be the most important consideration. The weaknesses of the vehicles are deemed acceptable.

Plan A	Plan B
It is expected that Hawk will arrive the quickest, and Panther will move quickly on the base road. It is uncertain if Hawk and Panther will have the endurance to conduct a prolonged search if necessary.	It is expected that both vehicles will be able to conduct a prolonged search. Lion can neutralize any hostile. It is uncertain if Lion and Eagle will arrive in time to locate and neutralize the target.

These are the top two plans. **Plan A** is the preferred plan. Which do you choose?

IMPACT

Human-Swarm Interaction

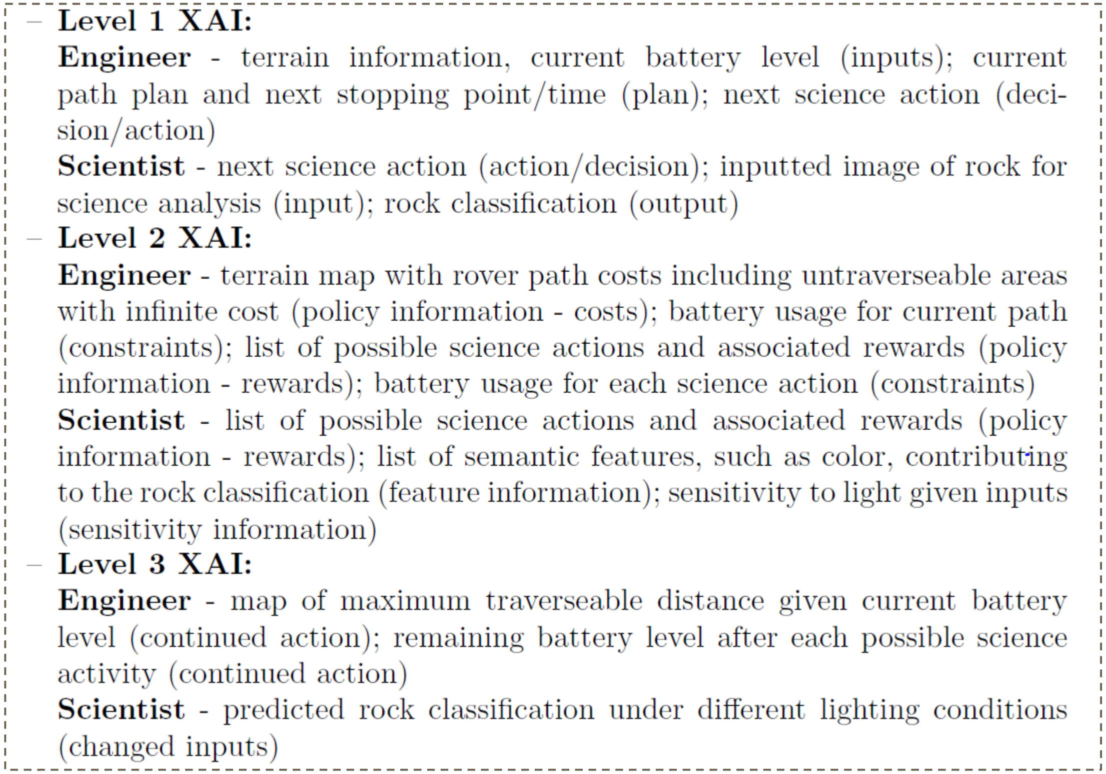
- Roundtree et al. (2019) identify key challenges associated with applying transparency design principles to achieve the three levels of SAT.
  - Design guidelines on transparent human-swarm interface visualizations (Roundtree et al. 2020).

SAT-BASED RESEARCH

Explainable AI

- ***Planetary Rover***
  - XAI design and evaluation framework (similar to SAT) based on human users' informational needs related to their situation awareness in the human-agent tasking environments (Sanneman & Shah, 2020)
- ***Fake News Detector***
  - SAT-based explainer for a fake news detection system (Chien et al. in press)
- ***"Junior Cyber Analyst"***
  - SAT-based HMI for a "junior cyber analyst" AI agent that works with the "senior" human analysts to identify cyber threats (Holder & Wang, 2021)

Planetary rover example



- **Level 1 XAI:**
  Engineer - terrain information, current battery level (inputs); current path plan and next stopping point/time (plan); next science action (decision/action)
  Scientist - next science action (action/decision); inputted image of rock for science analysis (input); rock classification (output)
- **Level 2 XAI:**
  Engineer - terrain map with rover path costs including untraverseable areas with infinite cost (policy information - costs); battery usage for current path (constraints); list of possible science actions and associated rewards (policy information - rewards); battery usage for each science action (constraints)
  Scientist - list of possible science actions and associated rewards (policy information - rewards); list of semantic features, such as color, contributing to the rock classification (feature information); sensitivity to light given inputs (sensitivity information)
- **Level 3 XAI:**
  Engineer - map of maximum traverseable distance given current battery level (continued action); remaining battery level after each possible science activity (continued action)
  Scientist - predicted rock classification under different lighting conditions (changed inputs)

Sanneman, L., & Shah, J. A. (2020). A Situation Awareness-Based Framework for Design and Evaluation of Explainable AI. In *Proc. EXTRAAMAS* 2020.

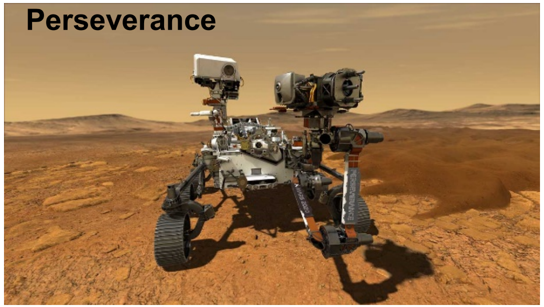


Image: NASA/JPL-Caltech

Approved for Public Release/ Distribution Unlimited

FINDINGS OF SAT-BASED RESEARCH

Operator Performance

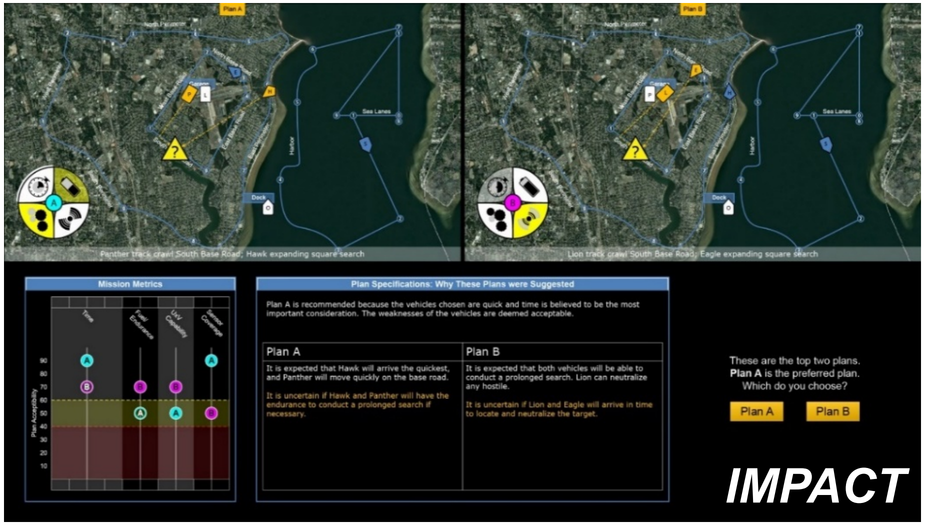
- Significant improvements as agent transparency (AT) increases
  - More effective trust calibration

Operator Workload

- No sig. increases as AT increases

Operator's Perception of Agent

- Operator's perceived *trust* tends to increase as AT increases
  - Factors such as agent reliability may impact trust more
  - Effects of uncertainty information on trust not always consistent
    - Level 3 (projection) information may lead to over-trust in an unreliable agent (Bhaskara et al., 2021)
- Operator's perceived *humanness* (anthropomorphism and intelligence) increases as AT increases



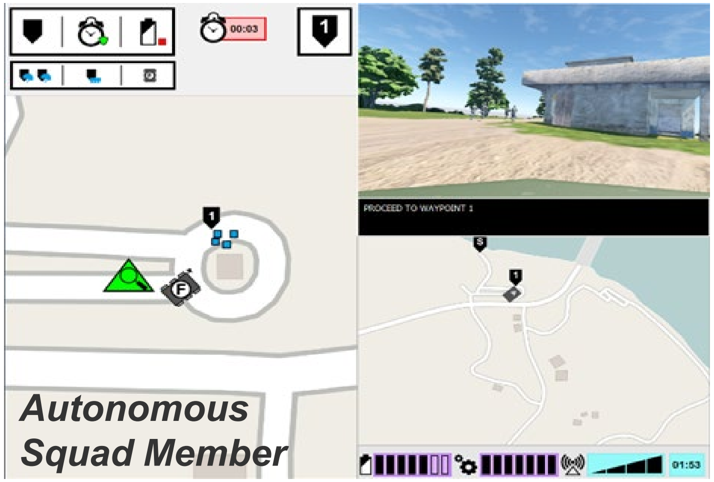
Plan Specifications: Why These Plans were Suggested

Plan A is recommended because the vehicles chosen are quick and time is believed to be the most important consideration. The weaknesses of the vehicles are deemed acceptable.

Plan A	Plan B
It is expected that Hawk will arrive the quickest, and Panther will move quickly on the base road. It is uncertain if Hawk and Panther will have the endurance to conduct a prolonged search if necessary.	It is expected that both vehicles will be able to conduct a prolonged search. Lion can neutralize any hostile. It is uncertain if Lion and Eagle will arrive in time to locate and neutralize the target.

These are the top two plans.
Plan A is the preferred plan.
Which do you choose?

IMPACT



Autonomous Squad Member

Increasing AT => better operator performance without increases in workload

INDIVIDUAL & CULTURAL DIFFERENCES

Individual Differences

- Individual differences in attitudes toward robots (e.g. unreasonable expectations of robot capability or negative attitudes toward humanlike robots) may impact humans' mental models of robots' task performance, which in turn, may affect their trust calibration and SA (Matthews et al. 2020)
  - Transparent interface design suggestions based on the SAT framework
  - Transparency content should be compatible with the operator's mental model by highlighting appropriate aspects of robots' capabilities

Cultural Differences

- Effects of cultural differences on human-agent interaction in the context of multiagent management (Chien et al. 2020)
  - Three distinct cultural backgrounds (based on the Cultural Syndromes Theory) were assessed in the experiment: United States (Dignity), Taiwan (Face), and Turkey (Honor).
  - Transparency had an impact on operator's interaction with the planning agent (i.e. compliance with agent's recommendations), but the effects of agent transparency were significantly influenced by participants' culture. For example, Face culture participants had a higher tendency to accept recommendations from an opaque agent.
  - When transitioning autonomy technologies from one culture to another, user interface modifications and training interventions may be required

Approved for Public Release/ Distribution Unlimited

CHALLENGES & FUTURE RESEARCH

- Architecture of transparent interfaces and info requirements
  - System users vs. evaluators
  - Operator vs. scientists

- Transparent interfaces (all 3 levels of SAT) for systems that continue to learn and evolve
  - How to convey newly-acquired capabilities?

- Real-time generation of transparent/explainable content
  - Modalities of transparent/explainable interfaces
  - Effects on operator workload (Kunze et al. 2019; Skraaning & Jamieson, 2021)

Approved for Public Release/ Distribution Unlimited

REFERENCES

References

- ❑ Chen, J.Y.C., Lakhmani, S., Stowers, K., Selkowitz, A., Wright, J., & Barnes, M. (2018). Situation awareness-based agent transparency and human-autonomy teaming effectiveness. *Theoretical Issues in Ergonomics Science, 19*(3), 259-282.
- ❑ Chen, J.Y.C. et al. (2014). *Situation Awareness-based Agent Transparency* (ARL-TR-6905).
- ❑ Chen, J.Y.C., & Barnes, M.J. (2014). Human-agent teaming for multi-robot control: A review of human factors issues. *IEEE Transactions on Human-Machine Systems, 44*(1), 13-29.

Additional Resources

- ❑ *International Journal of Human-Computer Interaction*: Special Issue on “Transparent Human-Agent Communications” (2022)
- ❑ *IEEE Transactions on Human-Machine Systems*: Special Issue on “Agent and System Transparency” (2020)
- ❑ *Theoretical Issues in Ergonomics Science*: Thematic Issue on “Human-Autonomy Teaming” (2018)

ISSN: 1044-7318
International Journal of
Human–Computer Interaction
Editors: CONSTANTINE STEPHANIDIS & GAVRIEL SALVENDY
IEA Published in cooperation with the International Ergonomics Association and the Human Interface Society
Volume 37, 2021
Included in this print edition:
Number 4
Special Issue: Human-Computer Interaction in Autonomous Vehicle and Mobility
Guest Editors: Heidi Krömker, Cindy Mayas
Taylor & Francis
Theoretical Issues in Ergonomics Science
IEA
TIES
Taylor & Francis

IEEE TRANSACTIONS ON

HUMAN-MACHINE SYSTEMS

A PUBLICATION OF THE IEEE SYSTEMS, MAN, AND CYBERNETICS SOCIETY