
*Integrity ★ Service ★ Excellence*

# Effect of Agent Transparency on Trust in Human-Agent Multi-UxV Management

19 June 2016

**Dr. Laura Humphrey**

**V&V of Complex, Autonomous Systems**

**Autonomous Control Branch (RQQA)**

**Control Science Center of Excellence**

# Preface

- My own work in trust is limited!
- Present work by colleagues at ARL and U. Central Florida

Joseph Mercado, Jessie Chen, Michael Barnes

Michael Rupp, Daniel Barber, Katelyn Procci, Kristin Schaefer, Deborah Billings, James Szalma, Jeff Adams, Tracy Sanders, Peter Hancock

- ***Review*** – "A Meta-Analysis of Factors Influencing the Development of Trust in Automation" [1]
- ***Case Study*** – "Intelligent Agent Transparency in Human–Agent Teaming for Multi-UxV Management" [2]

# REVIEW – FACTORS INFLUENCING TRUST IN AUTOMATION

# Many Definitions of Trust

No consensus on a definition of trust ...

**Fan et al, 1998**

A human's willingness to accept direction from an automated system

**Madsen and Gregor, 2000**

The extent to which a user is confident in, and willing to act on, the basis of recommendations, actions, and decisions of an artificially intelligent agent

**Lee and See, 2004**

The attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability

... but some possible ***antecedents of trust***

AFRL

# Theoretical Model of Human-Automation Trust

Analyzed 42 studies to estimate effects of possible ***antecedents of trust***

| Human Operator | Automation | Environment |
|---|---|---|
| **Human Traits**<br>+* Age<br>+ Gender<br>+ Ethnicity<br>+ Personality<br><br>**Human States**<br>Fatigue<br>Stress<br>Attentional control<br><br>**Cognitive Factors**<br>+* Understanding the automation<br>+* Ability to use automation<br>+* Expectancy of automation<br><br>**Emotive Factors**<br>+ Attitudes towards automation<br>Confidence in automation<br>+ Satisfaction with automation<br>+* Comfort with automation | **Features**<br>* Degrees of automation<br>+* Appearance<br>* Mode of communication<br><br>**Capability**<br>+* Errors in automation<br>+* Automation behavior(s)<br>Quality/Accuracy<br>* Cueing/Feedback/Alarms | **Task Related**<br>+* Proximity<br>+ Risk |

Note: * represents experimental and + represents correlation findings

# Calibrated Trust

- Trust is not always good!
- ***Calibrated trust*** [3] – trust the automaton when it's right, not wrong
- Avoid:
  - ***Misuse*** – over-trust in faulty automation
  - ***Disuse*** – under-trust in correct automation

# CASE STUDY – EFFECTS OF TRANSPARENCY ON TRUST FOR MULTI-UXV MANAGEMENT

# Agent Transparency

*Agent Transparency*

Ability to afford an operator's comprehension about its intent, performance, future plans, and reasoning process

- Antecedent of trust
- Helps operator calibrate trust, build mental model of automation, maintain situational awareness (SA)
- Should be clear and efficient – too much information can increase workload

# SA-Based Agent Transparency (SAT)

- SAT captures information at three levels:

| Level | Name | Content |
|---|---|---|
| 1 | Perception | Basic Info<br>State, goals, intentions, plan of action |
| 2 | Comprehension | Reasoning Process<br>Rationale, capabilities, limitations, trade-offs between options |
| 3 | Projection | Prediction<br>Possible outcomes, predicted consequences, likelihood of success |

# SAT Mission Scenario

- Single human operator uses multiple UxVs (Ground, Air, or Sea) to help defend a base
- UxVs have different capabilities – e.g. stealth, range, sensor types, viable operating conditions
- Operator receives intelligence reports and information on commander's intent
- Assign tasks to UxVs with intelligent agent assistance

# Experimental Setup

- Intelligent Agent (IA) offers operator plans A & B
  - Plan A should be better (IA correct)
  - Plan B is sometimes better (IA incorrect)
- Independent variable – three transparency levels
- 8 missions per transparency level, IA incorrect 3 times
- Dependent Measures (selected)
  - Performance – correct IA usage/rejection
  - Self-assessed trust – "Trust Between People and Automation" survey
  - Workload measures – NASA-TLX survey

# Simulator Setup

| Icon | Type | Name | Sensor | Strengths | Weaknesses |
|---|---|---|---|---|---|
| A1 | UAV | A1 | HD Zoom (passive)<br>Metal Detector (active) | Visual coverage; searching in clear conditions & for metal objects; heavier | Smoke (unless searching for metal); grounded in rain; louder |
| A2 | UAV | A2 | Day/Night Infrared (passive) | Chasing; finding (even in smoke); quieter & lighter | Long distances |
| G1 | UGV | G1 | Day/Night Infrared (passive) | IED bot; pedestrian avoidance (good on base roads) | High wind; grounded in rain |
| G2 | UGV | G2 | HD Zoom (passive) | Weaponized; large/intimidating | smoke; pedestrian avoidance; unexpected obstacles |
| S1 | USV | S1 | Day/Night Infrared (passive) | Quick; shallow-water friendly | chop (bad weather) |
| S2 | USV | S2 | HD Zoom (passive)<br>Metal Detector (active) | Weaponized; large/intimidating | shallow water |

### Mission Objective

The Commander received a report from a supervisor on base that a man has been spotted walking erratically along the south Base Road. This man's location is currently unknown. The Commander wants that man found immediately so we can determine whether this is a medical emergency or a terrorist. Send on UAV and one UGV to find this man.

### Commander's Intent

08:01 - The Commander received a report from a supervisor on base that a man has been spotted walking erratically along the south Base Road. This man's location is currently unknown. The Commander wants that man found immediately so we can determine whether this is a medical emergency or a terrorist. Send on UAV and one UGV to find this man.

### Alert History

08:00 – Weather Report: The weather is all clear.

08:01 – Patrol Report: A shipping boat has requested access to the Sea Lanes

08:02 – Patrol Report: An unknown person acting erratically has been spotted near the south Base Road.

The Agent suggests that **Plan A** is the best plan.

*What would you like to do?*

Plan A — *accept suggestion*

Plan B — *reject suggestion*

# Information Panel

Asset Capabilities

| Icon | Type | Name | Sensor | Strengths | Weaknesses |
|---|---|---|---|---|---|
| A1 | UAV | A1 | HD Zoom (passive)<br>Metal Detector (active) | Visual coverage; searching in clear conditions & for metal objects; heavier | Smoke (unless searching for metal); grounded in rain; louder |
| A2 | UAV | A2 | Day/Night Infrared (passive) | Chasing; finding (even in smoke); quieter & lighter | Long distances |
| G1 | UGV | G1 | Day/Night Infrared (passive) | IED bot; pedestrian avoidance (good on base roads) | High wind; grounded in rain |
| G2 | UGV | G2 | HD Zoom (passive) | Weaponized; large/intimidating | smoke; pedestrian avoidance; unexpected obstacles |
| S1 | USV | S1 | Day/Night Infrared (passive) | Quick; shallow-water friendly | chop (bad weather) |
| S2 | USV | S2 | HD Zoom (passive)<br>Metal Detector (active) | Weaponized; large/intimidating | shallow water |

Mission Objective

**Mission Objective**

The Commander received a report from a supervisor on base that a man has been spotted walking erratically along the south Base Road. This man's location is currently unknown. The Commander wants that man found immediately so we can determine whether this is a medical emergency or a terrorist. Send on UAV and one UGV to find this man.

Intel

**Commander's Intent**

08:01 - The Commander received a report from a supervisor on base that a man has been spotted walking erratically along the south Base Road. This man's location is currently unknown. The Commander wants that man found immediately so we can determine whether this is a medical emergency or a terrorist. Send one UAV and one UGV to find this man.

**Alert History**

| |
|---|
| 08:00 – Weather Report: The weather is all clear. |
| 08:01 – Patrol Report: A shipping boat has requested access to the Sea Lanes |
| 08:02 – Patrol Report: An unknown person acting erratically has been spotted near the south Base Road. |
| |

Decision

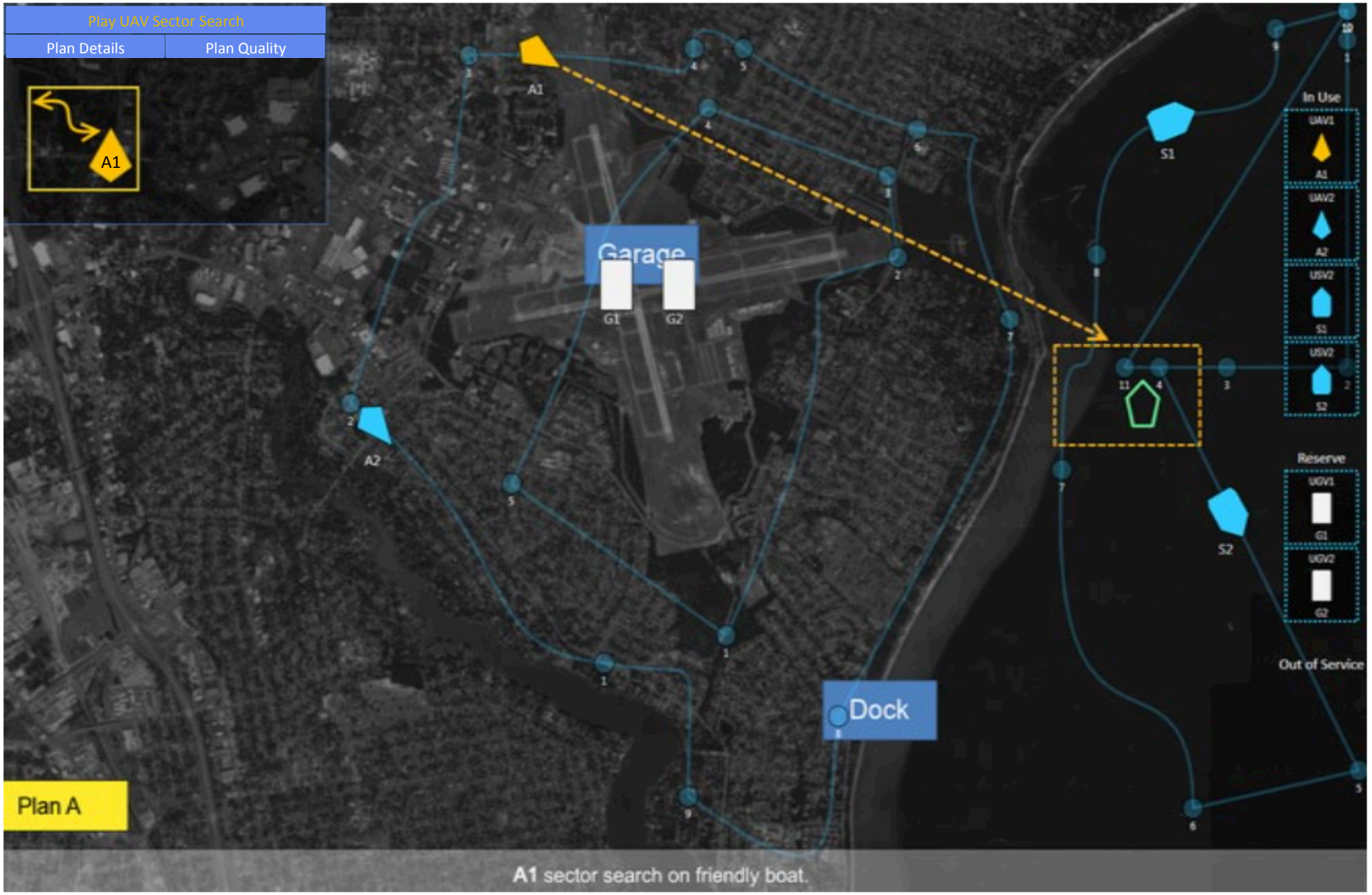
The Agent suggests that **Plan A** is the best plan.

*What would you like to do?*

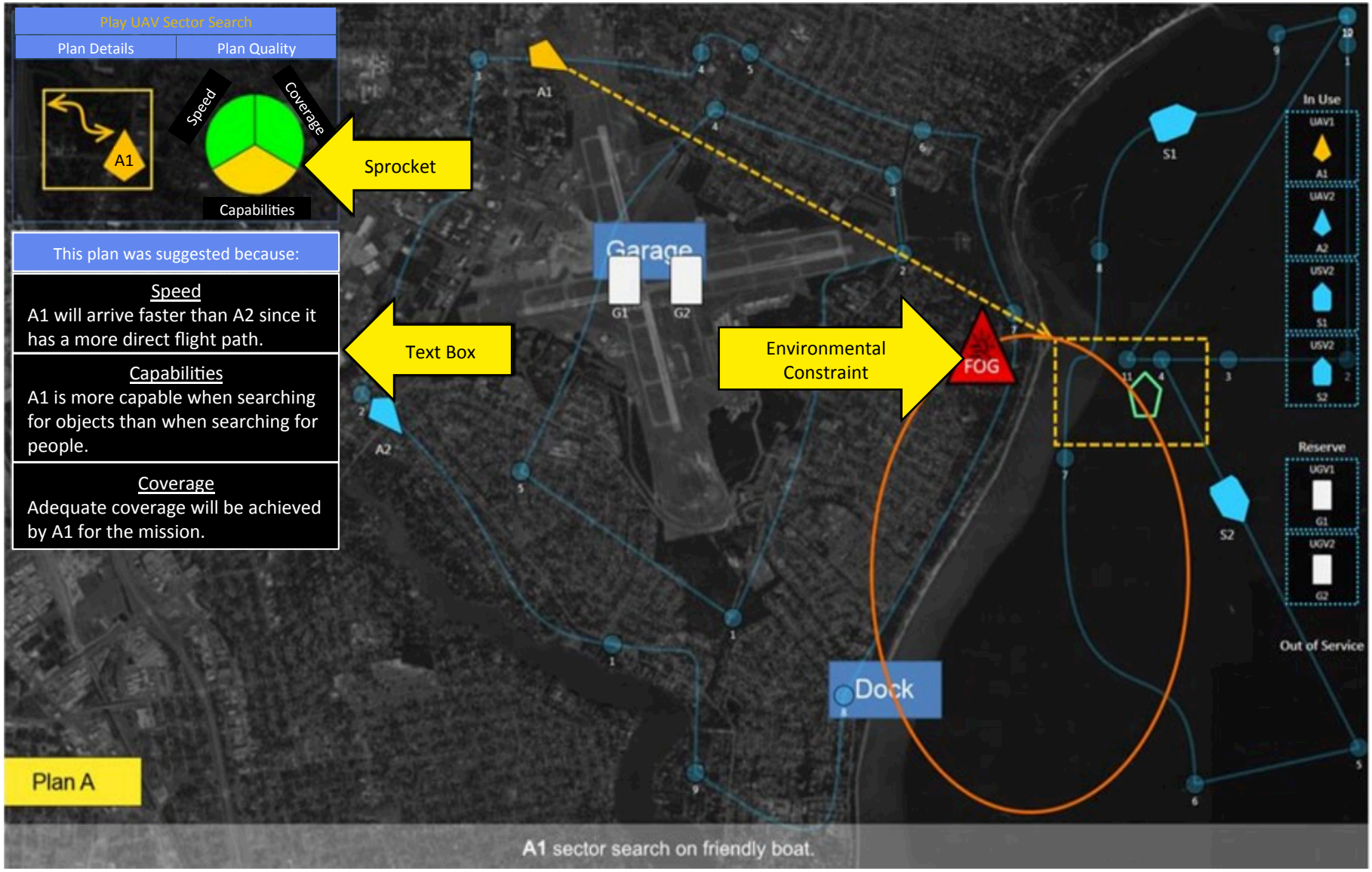
Plan A — *accept suggestion*

Plan B — *reject suggestion*

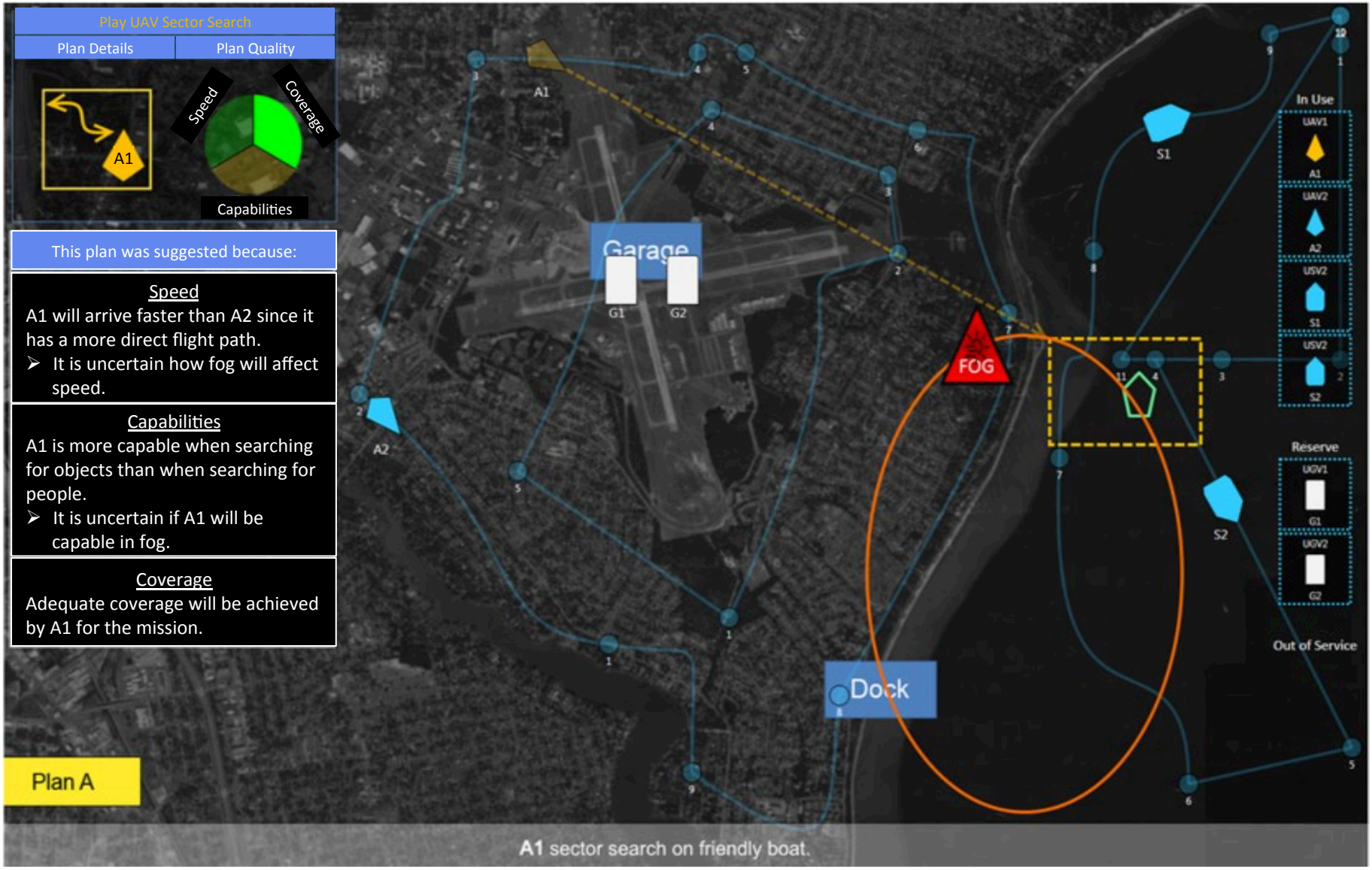
# Plan Panel – Level 1



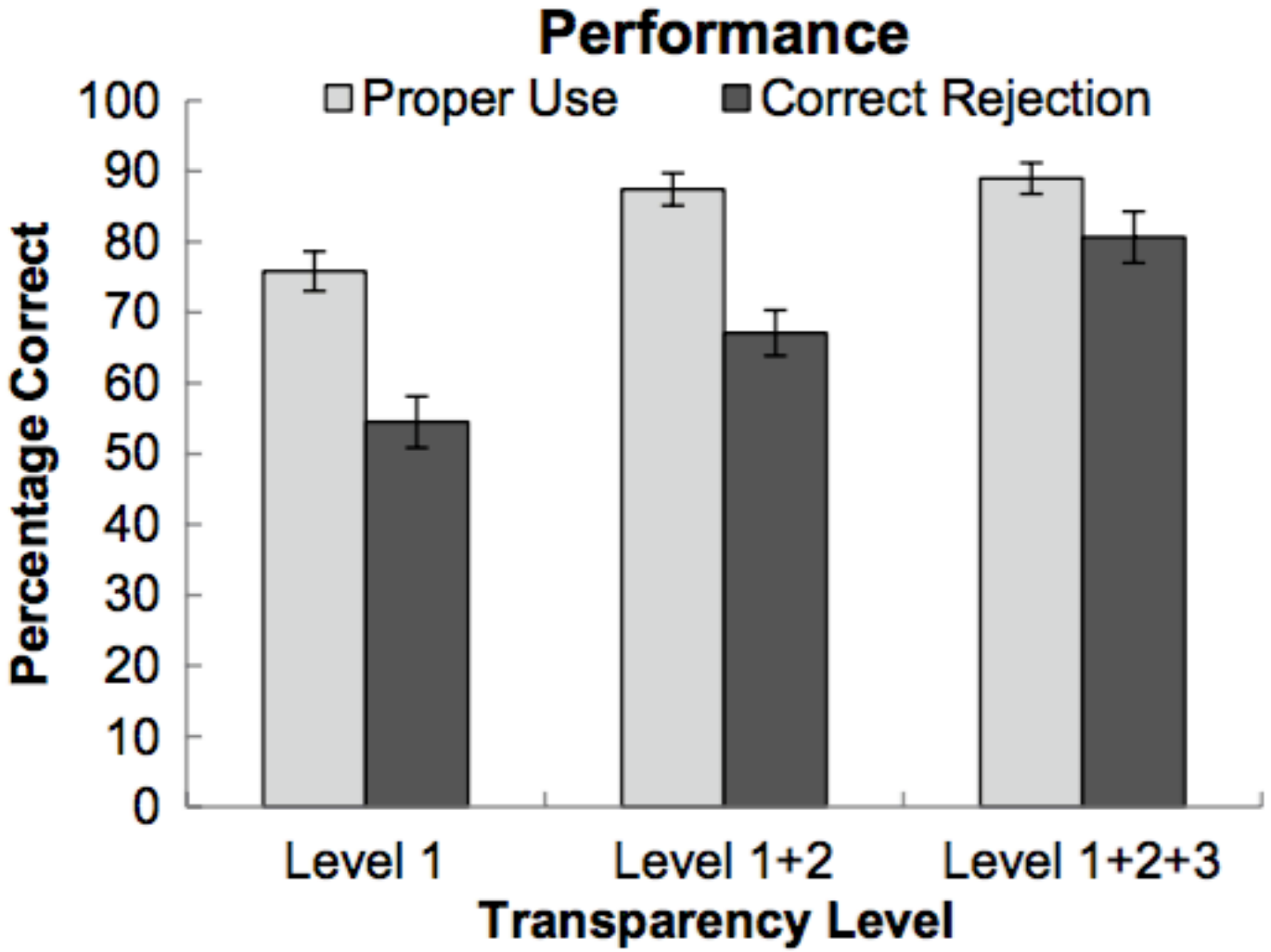
# Plan Panel – Level 2



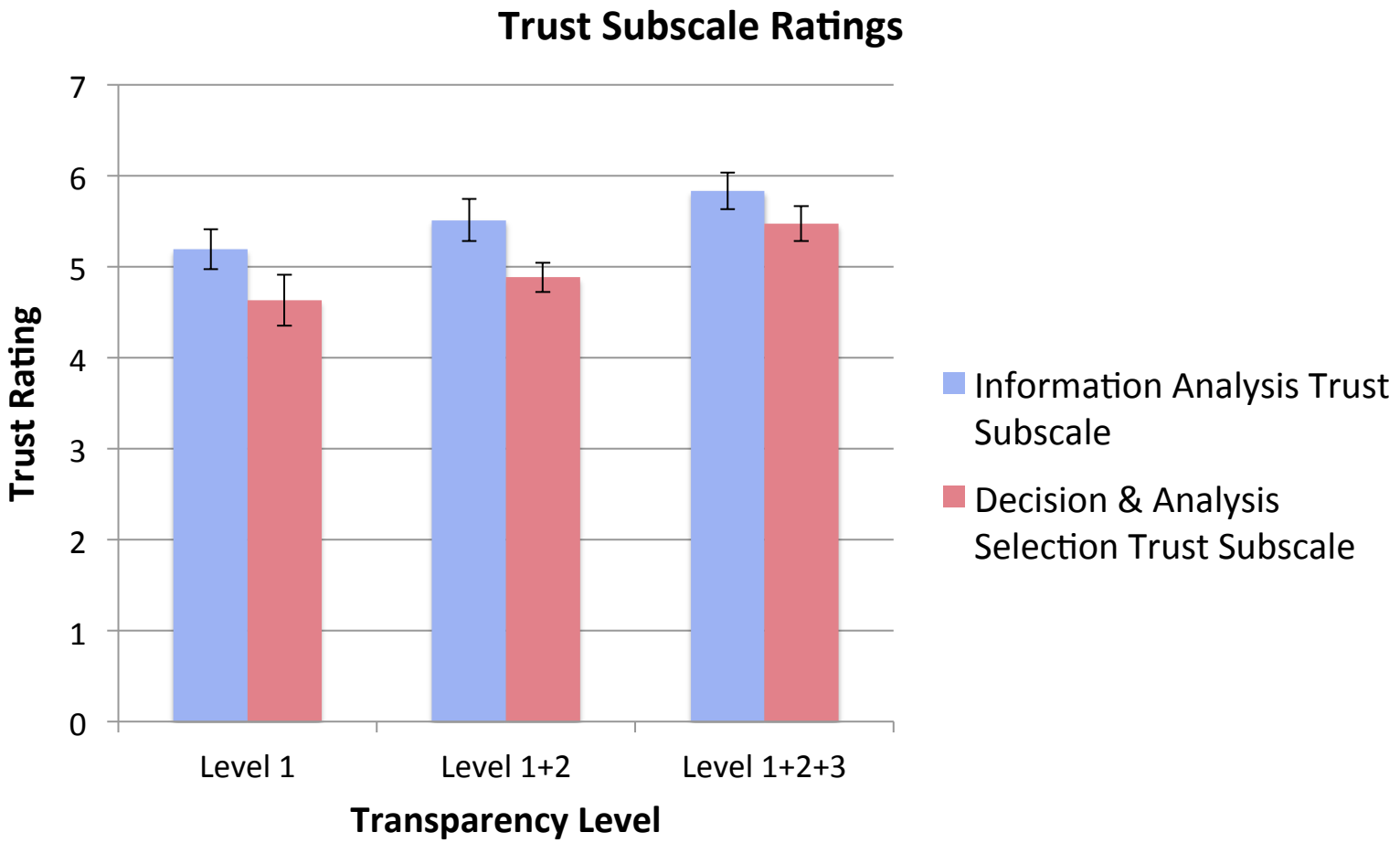
# Plan Panel – Level 3



# Outcome – IA Usage/Rejection



# Outcome – Trust

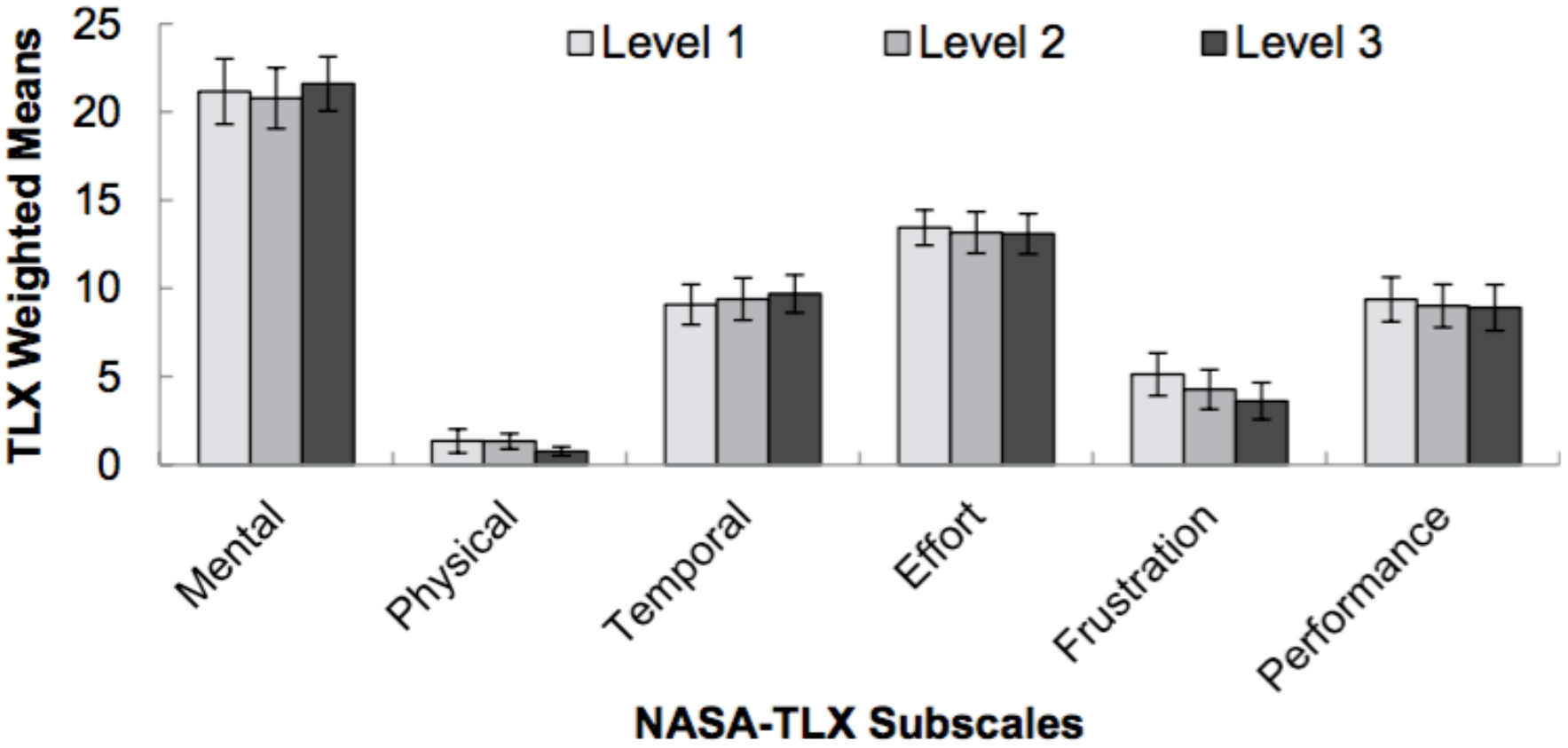


# Measuring Trust

- "Trust Between People and Automation" questionnaire [4]
- Respond to statements on scale of 1 (not at all) - 7 (extremely):

| | |
|---|---|
| The system is deceptive | The system provides security |
| The system behaves in an underhanded manner | The system has integrity |
| I am suspicious of the system's intent, action, or outputs | The system is dependable |
| I am wary of the system | The system is reliable |
| The system's actions will have a harmful or injurious outcome | I can trust the system |
| I am confident in the system | I am familiar with the system |

# Outcome – Workload



# References


1) K. E. Schaefer, D. R. Billings, J. L. Szalma, J. K. Adams, T. L. Sanders, J. Y. Chen, and P. A. Hancock. *A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Human-Robot Interaction*, Tech. Rep. ARL-TR-6984. ARL, 2014.

2) J. E. Mercado, M. A. Rupp, J. YC Chen, M. J. Barnes, D. Barber, and K. Procci. "Intelligent agent transparency in human–agent teaming for Multi-UxV management." *Human Factors: The Journal of the Human Factors and Ergonomics Society,* 58(3):401-415, 2016.

3) J. Y. Chen, K. Procci, M. Boyce, J. Wright, A. Garcia, and M. Barnes. *Situation Awareness-Based Agent Transparency*, Tech. Rep. ARL-TR-6905, ARL, 2014.

4) J. Jian, A. M. Bisantz, and C. G. Drury. “Foundations for an empirically determined scale of trust in automated systems.” *International Journal of Cognitive Ergonomics,* 4:53-71, 2000.