AFRL

THE AIR FORCE RESEARCH LABORATORY

LEAD | DISCOVER | DEVELOP | DELIVER

15th ICCRTS
"The Evolution of C2"

Development of Metrics for Trust in Automation

Dr. Janet E. Miller
Senior Electronics Engineer
Sensors Directorate
Air Force Research Laboratory

Motivation

- **Motivation for human-system considerations**
  - **Informed decision making is the essence of command and control**
    - **Can be human or machine**
  - **Increasing trends in *cyber-mediation* and *net-centricity* can make interactions happen more quickly and reliably**
  - **Successful attacks or failures on either human or machine will lead to distrust, disuse, and/or misuse in the joint system**
  - **Issues of *trust,* particularly in human-system teaming, need to be understood as is often mentioned as the silver bullet**

Background

- The US Department of Defense requires the ability to maintain operations in spite of a cyber attack ("fighting through") but there is an equal need for continuously forecasting and attributing malicious cyber conduct.

- Firmly lodged within these concepts is the issue of *trust*—trust in system integrity, trust in information integrity, trust in protections accorded by our own operational opacity, trust in our people.

  - Trust is not a state to be achieved but a multi-faceted and dynamic process.

  - Trust must be incessantly managed and tested, and integrated at the systems-of-systems level which includes hardware, software and humans.

But what is the concept of 'Trust?'

Trust

- ***"Never trust anything that can think for itself if you can't see where it keeps its brain."***
  - **J. K. Rowling, *Harry Potter and The Chamber of Secrets, 1999***
- ***"Trust but verify."***
  - ***Ronald Reagan***
- **"If I always told you the truth, you wouldn't have to trust me."**
  - **Dr. Who**
- **"…trust…"**
  - **US National Security Agency's Information Assurance Framework document – used 352 times!!**

General Issues of Trust

- **Trust is a fundamental social psychology concept**

- **Trust is a critical factor in a number of areas - Lee and See, 2004**
  - **Interpersonal relationships**
  - **Economic exchanges (firms/customers; management/staff)**
  - **Organizational productivity**
  - **Cross-disciplinary and cross-cultural collaboration**
  - **Electronically mediated transactions**

- **Importance of trust grows with…**
  - **Environmental uncertainty, task flexibility, & team structures**

- **…and drops with…**
  - **stable environments & structured hierarchies**
    - **Moorman, Deshpande & Zaltman, 1993**

Definitions
With respect to technology

Use: Voluntary employment of an automation technology

Disuse: Discontinuation or underutilization of technology

Misuse: Overreliance on a specific technology

Abuse: Inappropriate application of technology by designers or managers

-- Parasuraman and Riley (1997)

Trust: "…willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that party."

--Mayer R. C., Davis J. H., and Schoorman F. D. (1995)

BUT:

Vulnerable to what extent? Vulnerable to what outcome? How willing? What are the ramifications of being vulnerable? Does the context matter? Monitored or controlled to what extent?

Examples of Trust Gone Wrong

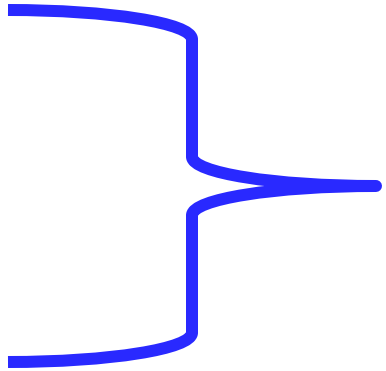
- The DHL B757 and Tu154M mid-air over Germany in 2002
  - The B757 crew, trusting TCAS in a close conflict situation, dove.
  - The Tu154 crew, trusting ATC, dove also. ATC was unaware of the advisories
- December 2009, an elderly couple traveling from Grants Pass, Oregon to Reno, Nevada trusted their GPS
  - Got stuck in snow for three days when their GPS unit sent them down a remote forest road
- February 2008, scores of radiation overdoses at Cedars-Sinai Medical Center
  - A misunderstanding over an 'embedded default setting'
  - Doctors believed it would provide them more useful data to analyze disruptions in the flow of blood to brain tissue.

Trust in Automation

- Considerable research ongoing in this area, in many domains
  - Cockpit automation and driver decision aiding
  - Teleoperated robots, manufacturing, and process control
  - Distributed computer-supported collaborative work
- Wide range of interactions observed across critical factors
  - System-associated factors
    - Level of aiding/automation/autonomy (full manual to full “auto”; in-the-loop, on-the-loop, out-of-the-loop)
    - Reliability of the aid/automation (failure rates, consistency,..)
    - Timeliness…
  - Operator/user-associated factors
    - Operator/user skill level, experience, personality
    - Reliance on the aid/automation by the operator/user
    - Operator familiarity/trust in the aid/automation
    - Workload,…
  - Other factors
    - Organizational, social, cultural,…

Trust Specification

- To investigate 'trust,' must specifically identify:
  - Object of Trust
    - Such as: Automation? Person? Inanimate object?
  - Context of Interest
    - Such as: Hazardous? Unknown? Stable?
  - Lower Level Attributes (aka Components)
    - Such as:
      - Competence
      - Predictability
      - Dependability
      - Consistency
      - Confidence



Critical for measurement!

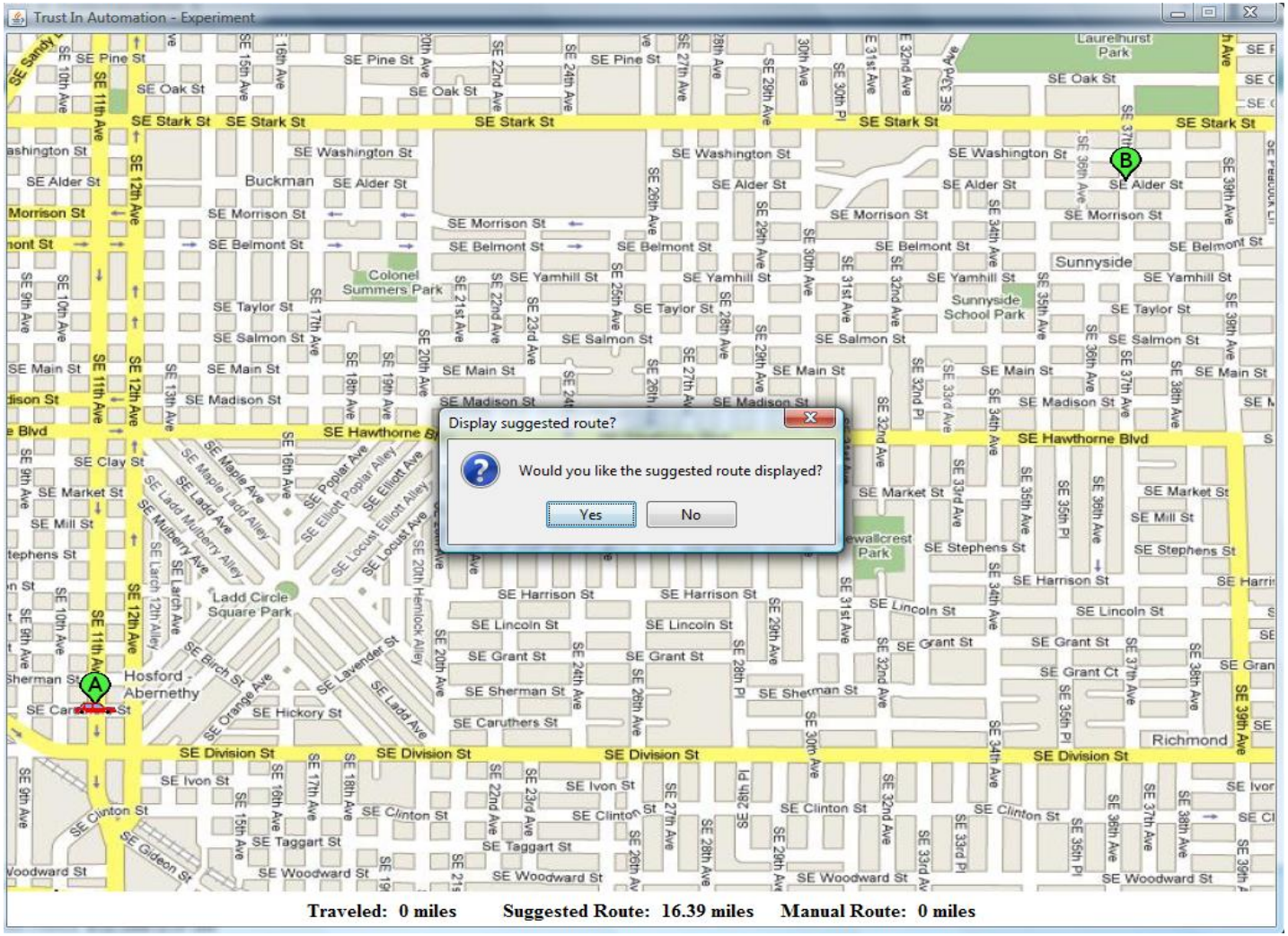
Trust Experiment on Attributes

- **Goal: Investigate whether the five attributes are reasonable for defining the qualifier of 'lower level components' in a trust in automation situation**
- **Participants**
  - **Ninety-five undergraduate students (M = 20, SD = 3.96) from a medium Midwestern university participated in the GPS simulation experiment. A within subjects experimental design was adapted where all participants completed all the GPS conditions.**
- **Platform**
  - **The automated tool used for the experiment was a Route Planner that resembles a GPS in that it assists in determining directions to a destination (figures 1-3). However, the Route Planner only displayed the entire map for an area of interest on the screen while a standard GPS could displayed either the current intersection or the entire area map. In addition, the Route Planner had the following simulated wireless updating capabilities for use in different experiments in this research: traffic jams, car accidents, burning buildings, unsafe neighborhoods, riot outbreaks, and drive by shootings.**

Trust Experiment on Components

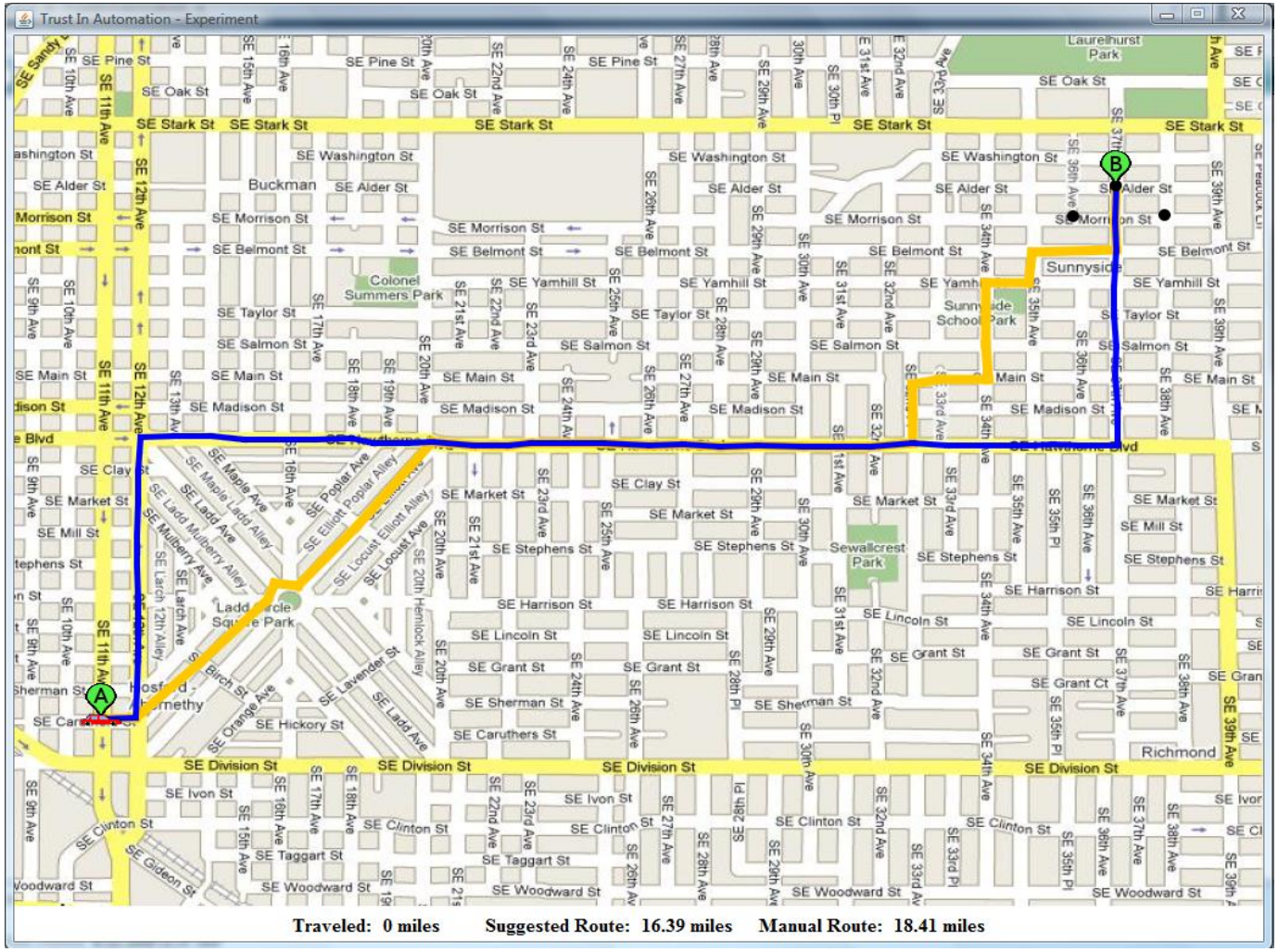
- Experiment One: Control Scenario
    - Travel from point A to B using the shortest distance.
- Experiment Two: Low Risk: Time Pressure Scenario
    - Time pressure was added.
- Experiment Three: Medium Risk: Common Hazards Scenario
    - A risk context was added
- Experiment Four: High Risk: Uncommon Hazards Scenario
    - Combination of a risk context and a time constraint added to the initial navigational goal. Participants were asked to avoid all hazards and to get to destination “B” in twenty minutes or less
- Randomization of Maps and Scenarios
    - Four maps and four scenarios

Experiment Display



Representative Display Screen: Asking user if wanted suggested route displayed

Experiment Display



Questions

	Not At All	A little	Sometimes	Frequently	All the Time
each item and then circle the er of the response that best ribes the extent to which you would he Route Planner's performance. ate to what extent you generally his way.					
o what extent is the Route Planner ompetent in mapping out the outes?	1	2	3	4	5
o what extent can the Route lanner's routes be predicted?	1	2	3	4	5
o what extent can you rely on the oute Planner to plan the routes?	1	2	3	4	5
o what extent is the Route Planner onsistent in planning the routes?	1	2	3	4	5
o what extent are you confident in he Route Planner's performance?	1	2	3	4	5

Attribute Definitions

- **Competence is the ability to do the task at hand**
- **Predictability is the matching of performance with expectations**
- **Dependability is always being there to perform**
- **Consistency is being free from variation or contradiction**
- **Confidence is the user's certainty that the automation will perform appropriately**

Results

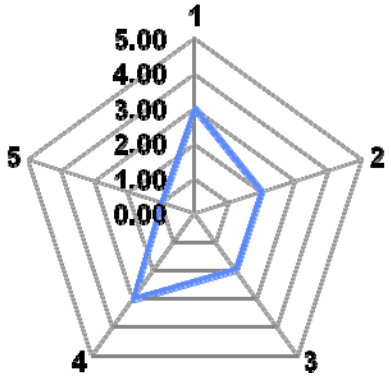
	Mean	Range	SD
1 - Competence	**3.91**	**2-5**	**.65**
2 - Predictability	**3.44**	**2-5**	**.86**
3 - Dependability	**3.94**	**2-5**	**.77**
4 - Consistency	**3.99**	**2-5**	**.78**
5 - Confidence	**3.81**	**2-5**	**.87**
Mean of Five Factors	**3.81**	**2.2-4.8**	**.71**

Highly moderate correlation between the participant assigned Likert scale value of the factors of competence, predictability, dependability, consistency and confidence and the participant assigned value for overall trust in the GPS system

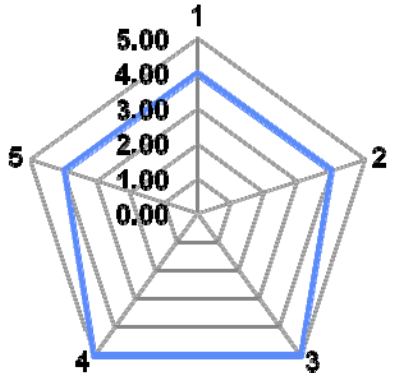
Metrics for Trust in Automation

Metrics: Attributes after identifying Object and Context

Measurement: Using Likert scale



Low Degree of Trust (Participant was 2.2 of 5)



High Level of Trust (Participant was 4.4 of 5)

Discussion

Implementation issues for active trust management

- **Who to monitor responses**
  - **Network control center**
  - **Security operations**
  - **Functional area operations**
    - **--- It depends!**
- **How to monitor responses**
  - **Intrusiveness**
  - **Burden to user**
  - **Tying to specific apps**

Conclusion

Trust loosely identified as silver bullet

- For system integrity
- For information integrity
- For protections accorded by our own operational opacity
- For people

BUT:

To be useful

- More specification required
  - Object of interest
  - Context
  - Attributes
- Metrics need to be actively applied