A Heuristic-Based Framework for Assessing Operator Trust in Autonomous Systems

Kim Jackson, Emily Vincent, Eric Jones, Zahar Prasov

NDIA Human Systems Conference

February 10-11, 2015
Develop actionable guidelines to determine how to design and evaluate autonomous systems that will support appropriate levels of operator trust.
What do we mean by “trust”?

- Trust (working definition): “adoption of and reliance on the system”
  - Adoption: Is the operator willing to use the system?
  - Reliance: Is the operator willing to rely on the system for mission-critical tasks?

- Trust needs to be appropriate for the given task.
Mismatched expectations lead to a lack of trust

How can we build trusted (or trustable) systems?

**Needs:**
- Identify unstated & unmet operator needs for trusted autonomous systems
- Understand what we know (academically) about how trust is built, measured, and understood

**Tools:**
- Literature review
- Human-Centered Engineering operator study
Data Collection

Tigershark

Guided Airdrop

Firescout

X-47B (UCAS-D)
Synthesis
Heuristics for Trusted Autonomy

- Visibility of Current System Behavior
- Visibility of Probable System Behavior
- Awareness of Latencies and Delays
- Visibility of System Capabilities & Limitations
- Transparency of Failure
- Fit with Users and Operations

**Accessibility of System Rationale**

- Display information about the system’s decision making process in a language familiar to the user
- Include, where appropriate
  - information about the decision making algorithm
  - Alternative behaviors
  - The system’s situational awareness of the environment
  - Levels of confidence
  - Algorithm performance metrics
Example: Accessibility of System Rationale

Planning constraints hidden

Terrain and Line-of-sight constraints shown
Heuristics as tools for design or evaluation

Heuristic Evaluation

Expert evaluation method to determine whether a system adheres to each heuristic. Produces actionable information for any deficiencies along with associated severity ratings.

Frequency
Impact
Persistence

Severity

*Nielsen, J. (1994). Usability engineering
# Standardized Scoring Examples

## Frequency

<table>
<thead>
<tr>
<th></th>
<th>Occurs rarely - less than once per mission</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Occurs once or twice per mission</td>
</tr>
<tr>
<td>2</td>
<td>Occurs once or twice per hour</td>
</tr>
<tr>
<td>3</td>
<td>Occurs many times an hour</td>
</tr>
</tbody>
</table>

## Impact

...  

## Persistence

...  

## Severity

### Persistence = 1

<table>
<thead>
<tr>
<th>Impact</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1 2 3 3</td>
</tr>
<tr>
<td>3</td>
<td>1 2 2 3</td>
</tr>
<tr>
<td>2</td>
<td>1 1 2 2</td>
</tr>
<tr>
<td>1</td>
<td>1 1 1 1</td>
</tr>
</tbody>
</table>

### Persistence = 2

<table>
<thead>
<tr>
<th>Impact</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>2 3 4 4</td>
</tr>
<tr>
<td>3</td>
<td>2 3 3 4</td>
</tr>
<tr>
<td>2</td>
<td>1 2 3 3</td>
</tr>
<tr>
<td>1</td>
<td>1 1 2 2</td>
</tr>
</tbody>
</table>

### Frequency

<table>
<thead>
<tr>
<th>Impact</th>
<th>Persistence</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>3</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>2</td>
<td>1 2 3 4</td>
</tr>
<tr>
<td>1</td>
<td>1 2 3 4</td>
</tr>
</tbody>
</table>

### Persistence

...  

...
Next Steps

- Validate heuristics
- Refine and validate evaluation method
- Investigate extensibility to other domains (beyond autonomous vehicle operations)
- Investigate applicability to other “user groups”
A Heuristic-Based Framework for Assessing Operator Trust in Autonomous Systems

Kim Jackson
Draper Laboratory
617-258-1543
kjackson@draper.com