Integrity ★ Service ★ Excellence

Trust in Automation: Measurement Considerations

Date: 4 APR 17

Joseph Lyons, PhD
Technical Advisor
Human Trust & Interaction Branch
Air Force Research Laboratory

Trust Background

- Trust = willingness to be vulnerable w/out the ability or capacity to monitoring (Mayer et al., 1995)
- Necessary Conditions
  - Human intention
  - Relational – must have a referent (otherwise its dispositional)
  - Risk – without it trust is obsolete
- What to Measure?
  - Intention to be vulnerable
  - Trustworthiness
    - Ability, benevolence, integrity (Mayer et al., 1995)
    - E.g., Machine characteristics (reliability, consistency etc.)
  - Reliance behavior
    - Must be volitional, should be in risky context – more than mere cooperation
  - Dispositional trust
  - Physiology?

The Problem – Inaccurate Trust

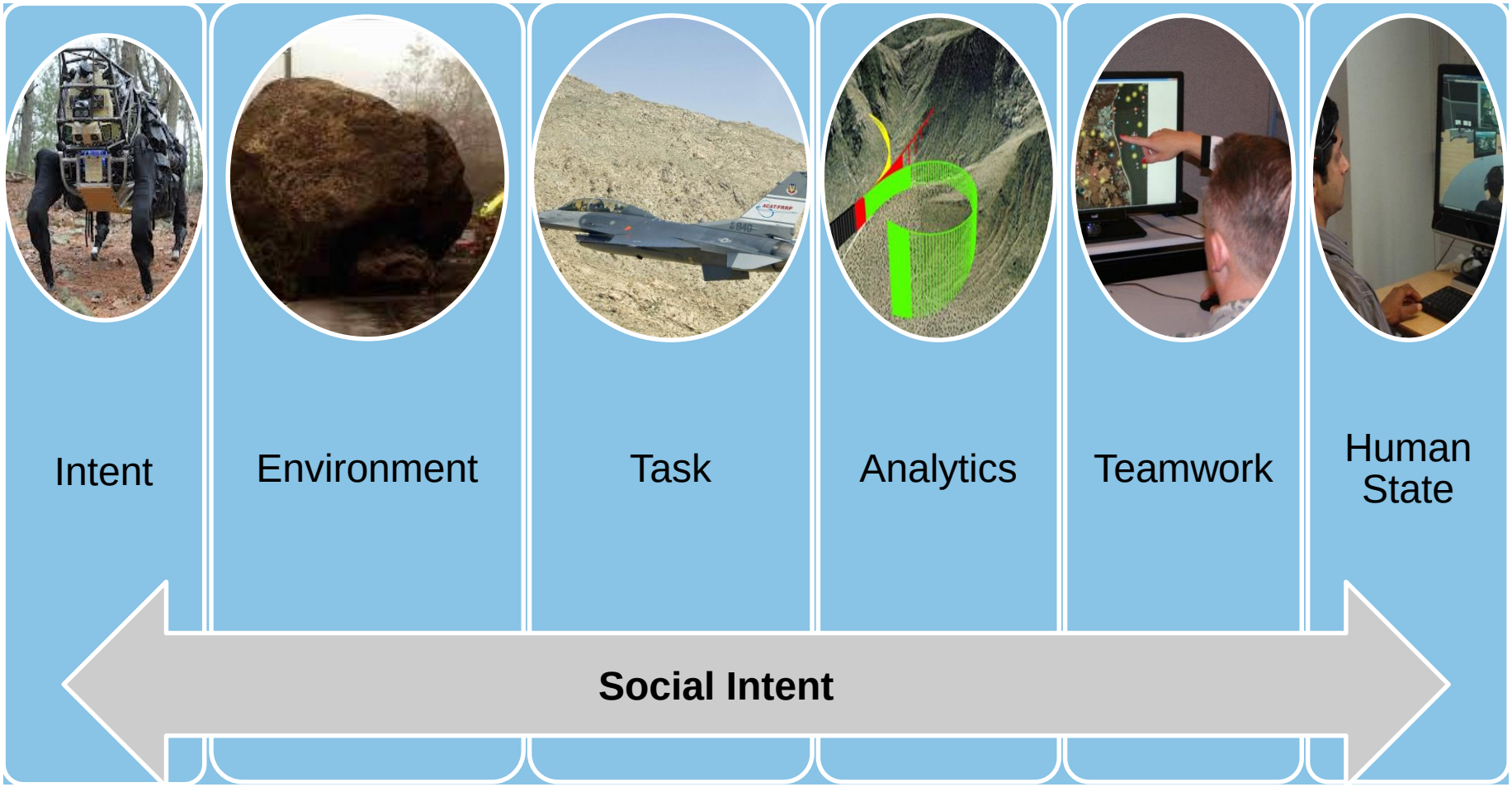
- Why is this so complicated???
  - Errors for highly automated systems can be catastrophic (Onnasch et al., 2014)
  - Trust is dynamic & driven by multiple factors
    - Social, Affective, Cognitive, Dispositional…
    - Absolute values will change with changing context
      - Importance of predictors may also change
  - Perceptions of trustworthiness may be inaccurate
  - Unintended consequences (Parasuraman & Riley, 1997)
  - Few studies using real technology & real-world consequences & real operators (R3)

Trust

- Intention to be vulnerable
  - Questionnaires
    - Focus should be on person, not the technology
    - Positive expectations
    - Intent to use
    - Willingness to engage in risk
  - Behavior
    - Reliance
    - Requires a scenario with an affordance for vulnerability
      - Remember “choice” is important…

Transparency

- **Human-Robot Transparency (Lyons, 2013)**



Trustworthiness

- What are the factors that shape trustworthiness?
- Context: Automated Data Fusion System
  - Source Characteristics
  - Data Retrieval Methods
  - Data Fusion Methods
  - User Interaction
  - Social Characteristics

Trustworthiness

- Source Characteristics
  - Pedigree of Information
    - Institutional trust
  - Information Recency
    - How old is this data?
  - Credibility of Source
    - Sensor types
    - Individual credentials/expertise

Trustworthiness

- Data retrieval methods
  - Traceability
    - Can you track back to the origin of the data?
  - Search Vectors
    - Where, how, how often, when
    - Excluded areas
  - Artifact Selection
    - Selection criteria, assumptions
  - Ingestion Methods

Trustworthiness

- Data Fusion Methods
  - Functional Transparency
    - What is done to the data?
    - How is it done?
  - Predictability
    - Logical link between inputs and outputs
    - Can focus on process versus product
  - Consistency
    - Relationship between inputs and outputs logically stable?
    - Learning system?

Trustworthiness

- User Interaction
  - User Interface
    - Intuitive
    - Readable
    - Useable
  - Familiarity
    - Common terms, assumptions, structure
  - Convergent Validity
    - Use of multiple factors to support the same outcome

Trustworthiness

- Social Characteristics
  - Psychological Business case
    - Why is it needed?
    - What's wrong with the status quo?
    - Purpose/Intent
  - User Bias
  - Organizational Constraints
    - Culture – values/motivation
    - Training – knowledge
    - Process – capabilities
    - Structure – capacity

Trust Background… "a small sliver"

- Management, e-commerce, Human Factors, Robotics, Social Psych… trust research is pervasive
- Mayer et al. (1995)
- Lee & See (2004)
- Parasuraman & Manzey (2010)
- Hancock et al (2011)
- Chen & Barnes (2014)
- Technology Acceptance Model
- Merritt et al. (2015)
- Hoff & Bashir (2015)

Thanks!

Joseph Lyons, PhD

Technical Advisor, Human Trust and Interaction Branch

711 HPW/RHXS

Air Force Research Laboratory

Joseph.Lyons.6@US.AF.MIL

(937) 255-8734