Trusted Autonomy

EDTAS
July 2015

Dr. Brian Mekdeci
Defence and Systems Institute
UNISA

Trusted Autonomy

Automation technology is improving

- Technology is not, and will not be perfect
- Humans are definitely not perfect

Human-machines will need to work together

- Machines should do what they are better at
- Machines should NOT do what they are NOT better at
- Requires trust
  - Humans trust the machines
  - Machines trust the humans*

Calibrated Trust

Trust

firm belief in the reliability, truth, or ability of someone or something

Trustworthiness

How much someone or something should be trusted

Overtrust

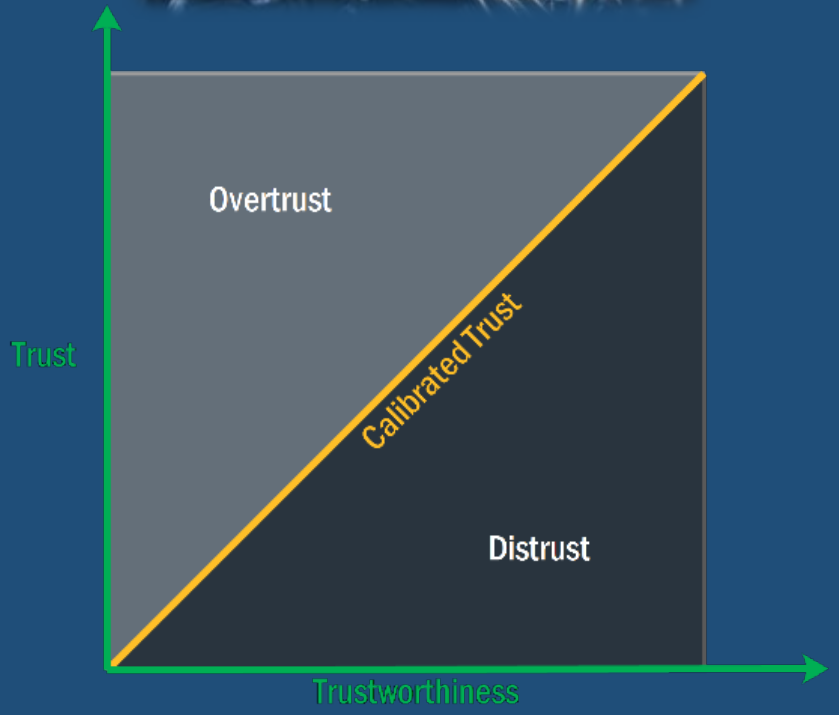
Trust > Trustworthiness

Distrust

Trust < Trustworthiness

Calibrated Trust

Trust = Trustworthiness



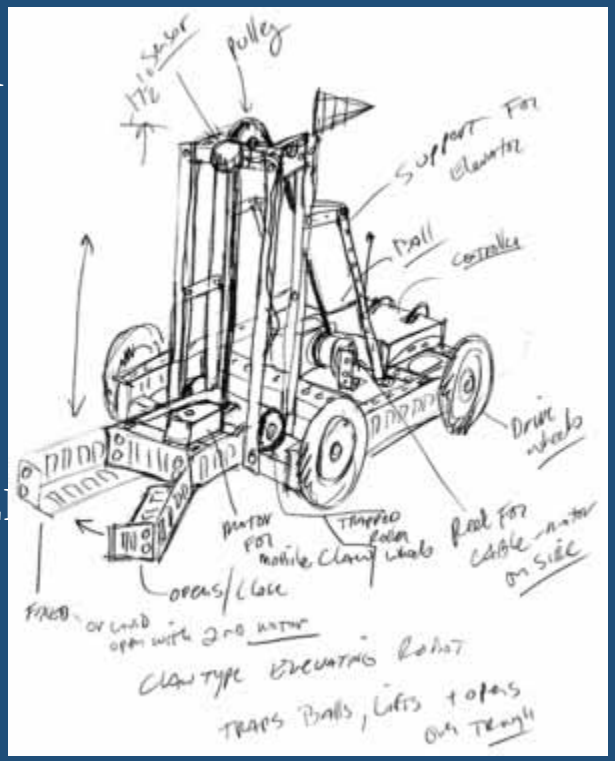
Lee (2004)

Engineering Trusted Autonomy

Improve Trustworthiness

Improve capability (hard engineer

- Better sensors
- Better actuators
- Better artificial intelligence



Improve Trust Calibration

Improve interaction (soft engineer

- Training
- User interface
  - Feedback

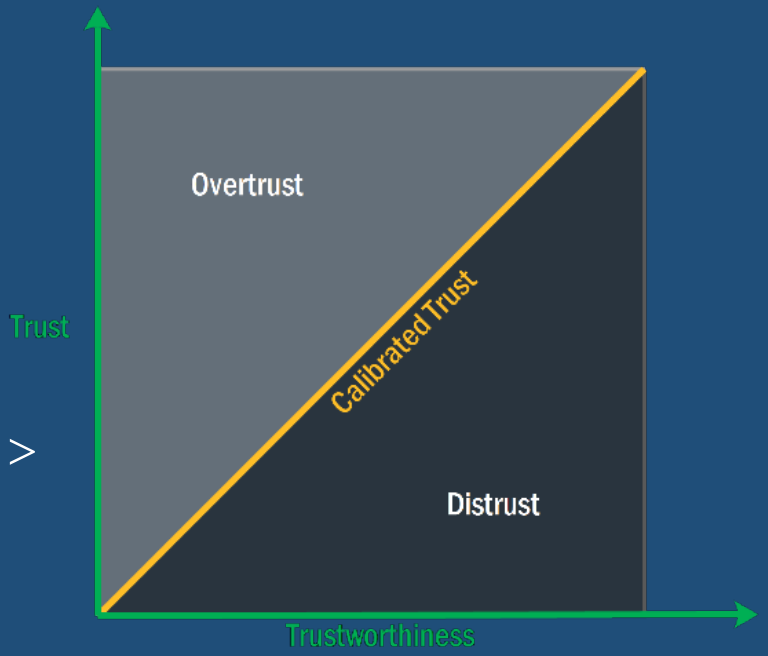
Is it really that easy?

Trustworthiness

Problems

What are the metrics for trustworthiness?

- Consider Google car
  - Number of collisions per 100,000 km
  - Detection rate of objects > 1 m^3?
  - Number of flops / s?



Lee (2004)

How are multiple dimensions of trustworthiness handled?

How is this data gathered?

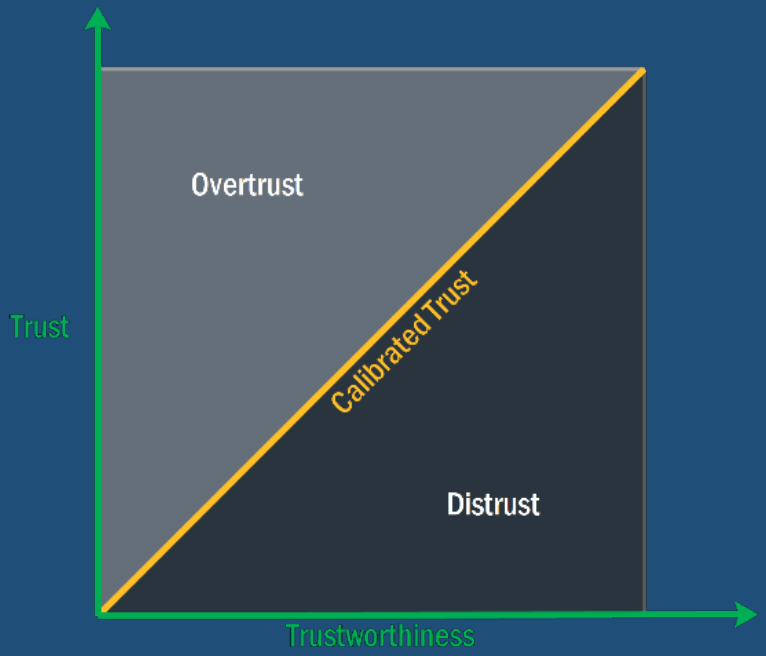
- Historical data / Trial and error
- Modelling / Simulation

Trust

Problems

What are the metrics for trust?

- Trust is a belief not an action
  - How is trust actually measured?
- Realizing trust
  - Is reliance what we are really interested in?



Lee (2004)

How are multiple dimensions of trust handled?

How is this data gathered?

- Historical data / Trial and error

Example: Contraband Detection

Goal

Detect contraband in luggage using automatic machine

Problem 1 – Determining Trustworthiness

How is the machine's trustworthiness determined?

- Historical data / trial and error
- Simulation and modelling

Assume 97% accuracy

- 2% False positives
- 1% Miss

Problem 2

How should an operator trust this imperfect machine?

- Rely on it 97% of the time
- Very subjective

Example: Vehicle Routing

Goal

Assign vehicles to targets so that overall response time is minimized

Problem 1 – Determining Trustworthiness

How is the machine's trustworthiness determined?

- Belongs to a system of systems
  - Reliance on other entities
- Conditions change
- Never know where next target is going to be
- No perfect solution

Problem 2

How should an operator trust this imperfect machine?

Trust Vs. Reliance

Reliance is often used as a measure of trust

Trust but not rely

An operator may trust the machine, but not rely on it

Rely but not Trust

An operator may not trust the machine, but rely on it anyway

Do we really care about trust or do we care about reliance?

Current Gaps

- Determining trustworthiness
  - Suitable testing and certification
- Setting appropriate trust level
  - Given a particular trustworthiness and context, what is the appropriate level of trust?
- Determining user's level of trust
  - How well does the user trust the machine?
- Guiding humans towards appropriate trust
  - Knowing what the appropriate trust is and the human's trust level, guiding the human to trust more appropriately

Lab

Hardware:

- 3 Computer Workstations
- 3 monitors each
- Standard keyboard and mouse
- Internet connection

Software:

- Real-time, multi UAV simulation in a spatio-temporal scenario
- Multi-operator capable
- Platform Agnostic, Internet browser-based
- Desktop, mobile
- Windows, iOS, Android, Li
- 2D/3D graphics

Operator Tasks

Routing:

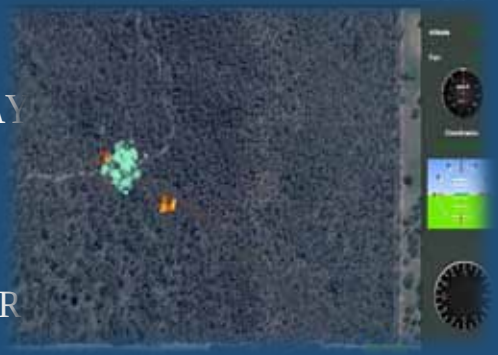
- Assign best vehicle to targets
  - Best strategy is somewhat subjective

Detection:

- Determine if target is located within a specific image
  - Right or wrong

Firefighting:

- Put out fire using water spray payload
  - Hand/eye coordination (2D)
  - Wind is the difficulty factor
  - Time is the key metric (skill based)



Rescue:

- Land on object, retrieve it
  - Hand/eye coordination (3D)
  - Momentum is the difficulty

Experimental Variables

Independent Variables

- Frequency / Severity of Errors
  - Low, High
- Ground Control Station
- Desktop, Portable
- Teammates
  - Human, Robot
- Error Reporting
  - Honest, None
- Work Load
  - High, Low
- Task Complexity
  - High, Low

Dependent Variables

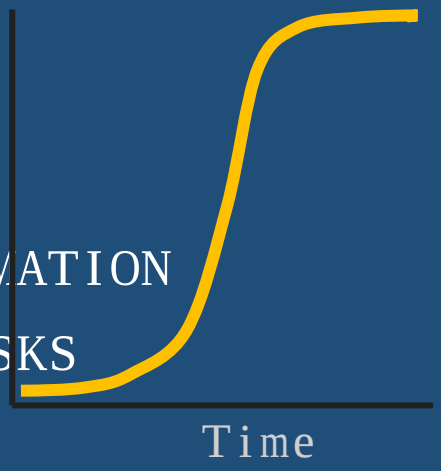
- Mission performance
- Trust
  - Likert scale
  - Number of interventions (reliance)
- Autonomy metrics
  - How often robots need assistance (neglect time)
- Communication and coordination
  - Number of messages, quality of the messages
- Workload
  - Ratio of service time to neglect time

Summary: Trends

- Drivers
  - Technology Improvement
    - Better trustworthiness
    - Private companies, like Google, are heavily involved
  - Cost savings
    - Replace expensive human resources with cheaper automation
- Barriers
  - Lack of proper testing, certification, etc.
  - Fear
    - Distrust of government, big business, Technology*

Summary: Forecast

- < 15 Years
  - Technology improvement
    - Better trustworthiness
  - More public acceptance for automation
  - Limited autonomy for complex tasks
- > 15 years
  - Attempts at full autonomy for complex tasks
  - Technology advancements start to slow down
  - Accidents and fear will prevent full public acceptance / trust
  - Economy problems (jobs) will be an issue



End of Presentation

Questions?

Hypotheses

1. Performance decreases as trust between human-machine, or human-human decreases
2. Trust in automation varies with task load
   - When humans are under high workloads, they will trust machines more
3. Trust in automation increases with task complexity
   - The less they have confidence in their own abilities, the more humans will trust the automation
4. Trust of machines decreases as failure of other technologies decrease
   - Robot may be distrusted if other parts of the overall system of systems fail
5. User interface of portable devices is not-ideal, reliance will increase because of convenience
   - This will be more pronounced during high-task load situations
6. Error reporting increases trust (reliance) on autonomy
   - Automation is not perfect

Research Questions

1. How can trust / trustworthiness be characterized and measured for different types of human-robot interaction?
   - How is trust and reliance related?
   - If we know how well an imperfect robot performs certain tasks, how do we calibrate trust?
2. Do humans trust other humans more than robots more than humans?
3. How do disruptions change trust?
   - Does it matter if the disruption as caused by
     - Human failure
     - Robot failure
     - System failure