What is It (like) to Trust a Machine?

Masashi Kasaki

Nagoya University

Dutch-Japanese Workshop in Philosophy of Technology 2018

Rikkyo University, Tokyo

July 7, 2018

Psychology of Robot Use

- Trust: An iRobot PackBot is promoted to staff sergeant (usually, team leader), named “Sgt. Talon,” and given Purple Hearts: “We always wanted him as our main robot. Every time he was working, nothing bad ever happened. He always got the job done. He took a couple of detonations in front of his face and didn't stop working.” (reported by Garreau (2007))

Psychology of Robot Use

- Lack of Trust: WORD (Special Weapons Observation Reconnaissance Detection) was sent to Iraq in 2007. Soldiers never used them in the field because of they did not trust them. (Ogreten et al. 2010)

Psychology of Robot Use

- Lack of Trust: the 2014 survey from Insurance.com: even in the face of dramatic evidence to the contrary, 61% say that they can make better decisions than a computer car.
- Over 75% don't trust a fully autonomous vehicle to drive their child to school.

(for more on the survey, see Vallet (2014))

Psychology of Robot Use

- Overtrust: National Transportation Safety Board report on Grounding of Royal Majesty, 1995: “all the watch-standing officers were overly reliant on the automated position display … and were, for all intents and purposes, sailing the map display instead of using navigation aids or lookout information.” (NTSB 1995: 4)

Trust in Machines

"Trust is important because operators may not use a well-designed, reliable system if they believe it untrustworthy. Conversely, they may continue to rely on automation even when it malfunctions and may not monitor it effectively if they have unwarranted trust in it." (Parasuraman & Miller 2004: 52)

Overtrust⇒Abuse
Mistrust⇒Misuse
Distrust⇒Non-use

Trust-affecting Factors

HUMAN-RELATED	PARTNER-RELATED	ENVIRONMENT-RELATED
Traits Age +* Gender + Ethnicity + Personality + Trust Propensity	**Features** Mode of communication* Appearance/Anthropomorphism+* Level of Automation * Intelligence (Robot/Automation) Personality (Robot/Agent)	**Team Collaboration** Role Interdependence Team Composition Mental Models Cultural/Societal Impact In-group membership
States Attentional Control Fatigue Stress	**Capability** Behavior +* Reliability/Errors +* Feedback/Cueing *	**Task/Context** Risk/Uncertainty + Context/Task Type Physical Environment
Cognitive Factors Understanding +* Ability to use +* Expectancy +*		
Emotive Factors Confidence in the automation Attitudes + Satisfaction + Comfort +*		

Cited from Schaefer et al. (2016: 387)

Definitions of Trust

J. Lee & K. See (2004: 54)	[Trust] is the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability.
K. Jones (2004: 6)	[Trust] is accepted vulnerability to another person's power over something one cares about, where (1) the truster foregoes searching (at the time) for ways to reduce such vulnerability, and (2) the truster maintains normative expectations of the one-trusted that they not use that power to harm what is entrusted.
K. Hawley (2014: 10)	To trust someone to do something is to believe that she has a commitment to doing it, and to rely upon her to meet that commitment. To distrust someone to do something is to believe that she has a commitment to doing it, and yet not rely upon her to meet that commitment.

Trust in Machines

- Lee & See's definition of trust is standardly used in studies on human-machine trust. They regard trust as (at least in part) affective, not fully cognitive. (Jones is representative of this conception of trust.)

- A prominent difference between Lee & See's, on the one hand, and Jones' and Hawley's definitions, on the other hand, is that the expectation involved in trust is not merely predictive but normative.

Normative Expectation

"Normative expectations" may be a wide-ranging notion, and may include:

- The trustee is obligated to act in the expected way.
- The trustee is responsible for the expected action if the trustee does it.
- The trustee is committed to the expected action.
- The trustee understands the value of the expected action for the trustor.
- The trustee takes it as a norm to act in the expected way, and so on.

Normative Expectation

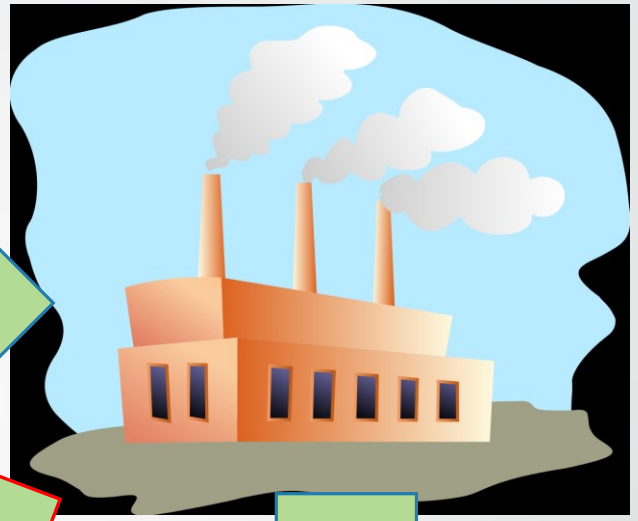
- Predictive expectation is a matter of expected reliability, but normative expectation is not.
- Predictive expectation is all about the trustee's behaviors.
- In the case of human-machine relationships, normative expectation is an attitude towards both a machine and its maker. For the maker is responsible for its behaviors.
- But this may change as technology and moral/legal systems change.

Normative Expectation

- Various factors affect one's forming, maintaining, and revising normative expectations about a machine:
- Makers and designers make sure that they implement the relevant normative codes into machines.
- Social, legal, and institutional systems assure the trustor that it is ok to trust a machine, e.g., its maker will be punished if it does not meet the expectation; and they also assure that the maker designs and maintains machines in such a way that they follow the relevant norms.
- Of course, the trustor learns from interaction with a machine and revise her normative and predictive expectations about it.

Legal, Social, and Institutional Systems

Assuring Norm-following



Assuring Normative Expectation

Normative Expectation

Norm Implementation

Normative Expectation

Predictive Expectation

Conclusions

- Trust is important for the use of machines.
- Normative expectation is the key to trust.
- It is a two-pronged attitude towards both a machine and its maker.
- Various factors affect normative expectations about machines.

References

Garreau, Joel. (2007) “Bots on the Ground”, Washingtonpost.com, http://www.washingtonpost.com/wp-dyn/content/article/2007/05/05/AR2007050501009.html?noredirect=on

Hawley, Katherine. (2010) “Trust, Distrust and Commitment”, *Noûs*, 48(1): 1-20, 2014.

NTSB, Marine Accident Report (1995), *Grounding of the Panamanian Passenger Ship Royal Majesty on Rose and Crown Shoal near Nantucket, Massachusetts, June 10, 1995*. Report Number

NTSB/MAR-97/01

Jones, Karen. (2014) “Trust and Terror,” in *Moral Psychology*, edited by P. DesAutels & M. Urban Walker, Lanham MA: Rowman and Littlefield.

Ogreten, S.; Lackey, S.; Nicholson, D. (2010) “Recommended Roles for Uninhabited Team Members

Within Mixed-Initiative Combat Teams”, The 2010 International Symposium on Collaborative Technology Systems, Chicago, IL, 2010.

Lee, John. D., and See, Katrina A. (2004) “Trust in Automation: Designing for Appropriate Reliance”, *Human Factors*, 46(1): 50-80.

Parasuraman, R., & Miller, C. (2004) “Trust and Etiquette in a Highcriticality Automated Systems”, *Communications of the Association for Computing Machinery*, 47(4): 51-55.

Schaefer, Kristin E., Chen, Jessie Y., Szalma, James L. and Hancock, Peter A. (2016) “A Meta-analysis of Factors Influencing the Development of Trust in Automation: Implications for Understanding Autonomy in Future Systems”, *Human Factors*, 58(3): 377-400.

Vallet, Mark. (2014) “Autonomous Cars: Will You be a Co-pilot or a Passenger”, Insurance.com, https://www.insurance.com/auto-insurance/claims/autonomous-cars-self-driving.html