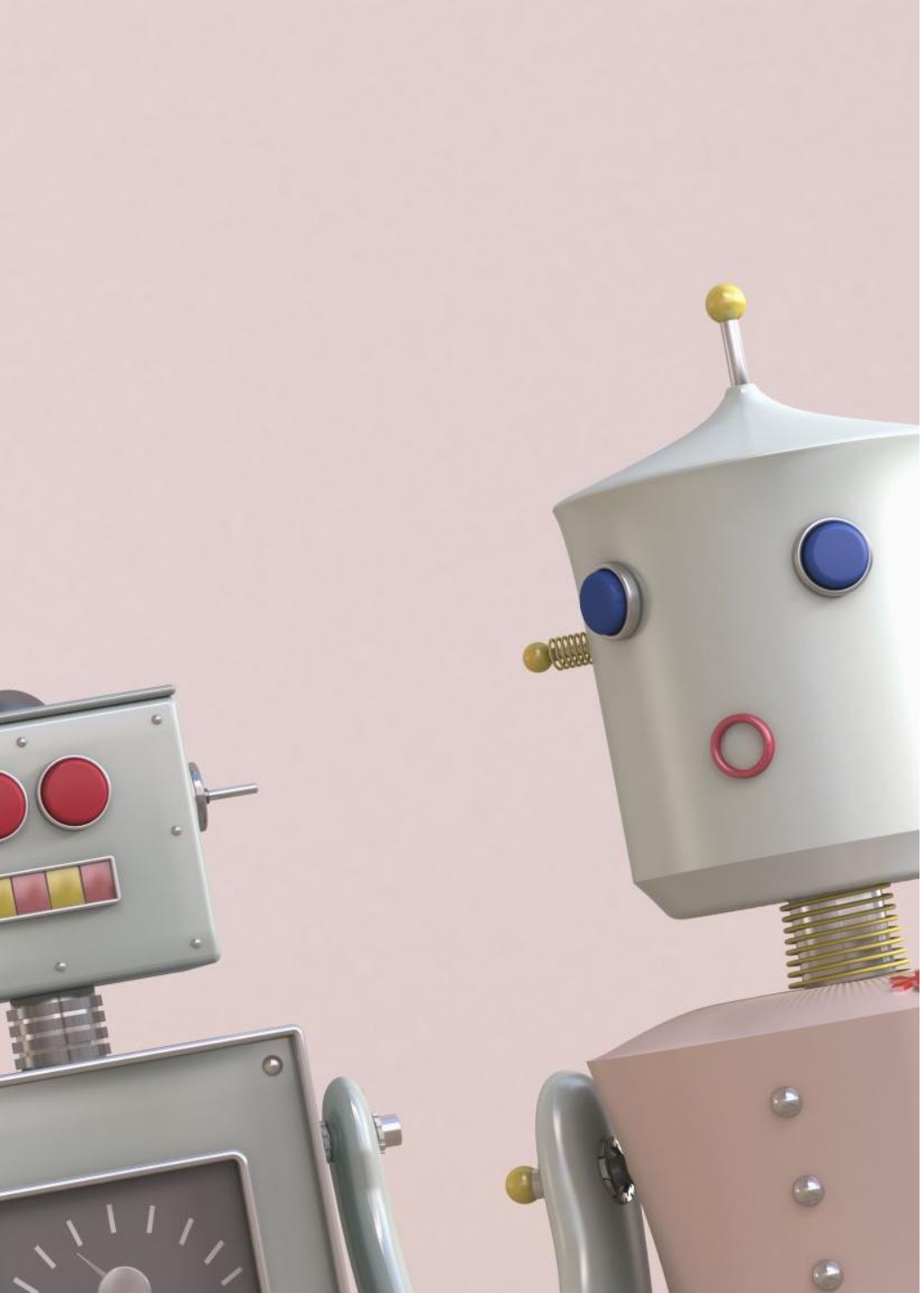


SOCIAL ROBOTS AND HUMAN-ROBOT INTERACTION

Week 4, Trust in HRI

Ana Paiva

P2
2025/2026

STRUCTURE OF LECTURE

- What is Trust and why study trust?
- Trust in AI Systems
- Trust in HRI: what factors affect trust in a robot?
- How to measure trust in HRI?
- Do humans over-trust robots (and technology in general)?
- Discussion

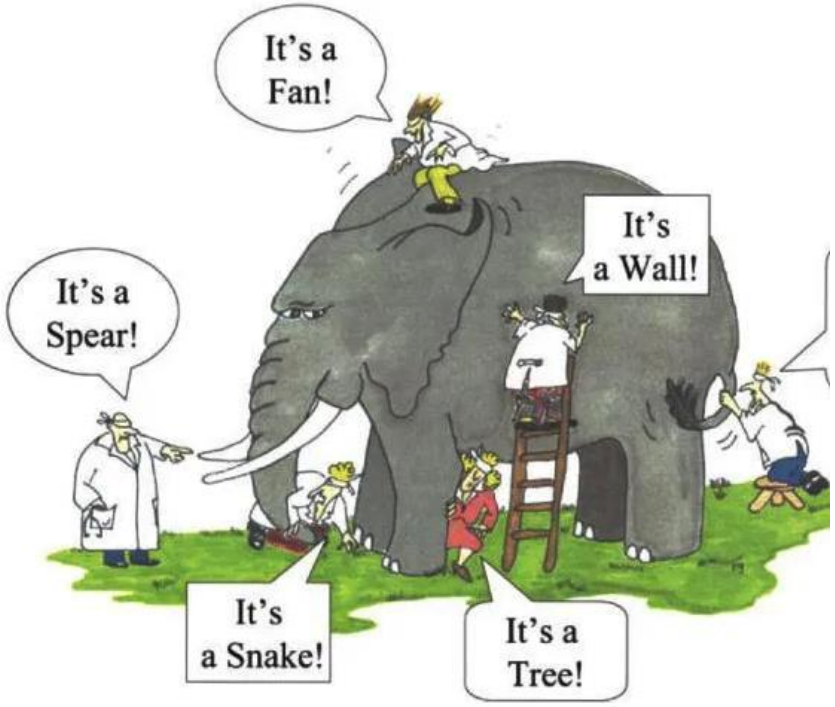
WHY TRUST?

Trust is *central to interpersonal relationships* (Golembiewski & McConkie, 1975)

Trust is *crucial wherever risk, uncertainty, or interdependence* exist (Mayer, Davis & Schoorman, 1995; Mishra, 1996).

There are *many individual differences* in trust

WHAT IS TRUST?



sociologists to see trust as a social structure;

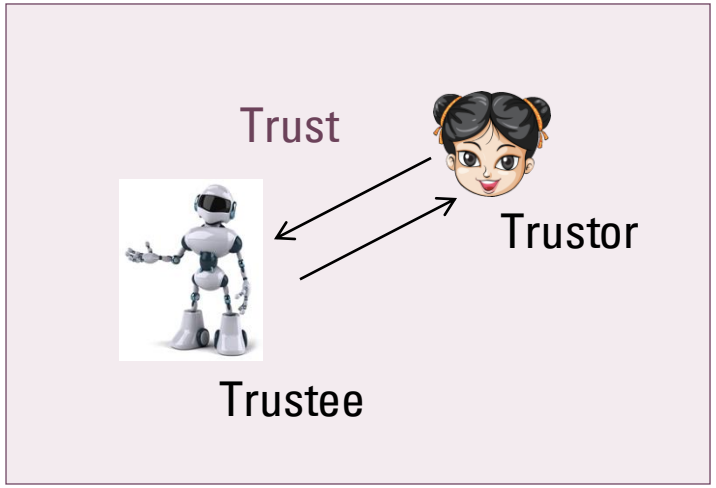
economists to see trust as an economic choice mechanism;

computer scientists as a feature in systems….

- By Lee et al. "Trust ***is the attitude*** *that others (an agent/robot/system) will help achieve an individual's goals in a situation characterized by* ***uncertainty and vulnerability***"
- By Hancock et al., is "Trust is *the* ***reliance*** *by someone that actions prejudicial to their well-being will not be undertaken by influential others*"

On average, "trust" had 17.0 definitions in dictionaries

TRUST



- Trust is in the relation/attitude (of the trustor towards the trustee)

- By Lee et al. ”Trust ***is the attitude*** *that others (an agent/robot/system) will help achieve an individual's goals in a situation characterized by* ***uncertainty and vulnerability***”
- By Hancock et al., is ”Trust is *the* ***reliance*** *by someone that actions prejudicial to their well-being will not be undertaken by influential others*"

WHAT LEADS TO TRUST?

- ***Competence***: is the trustee competence in performing the actions effectively?
- ***Reliability and safety***: is the trustee reliable and will do the task in a safe way?
- ***Predictability***: is the trustee competence in performing the actions according to the expectations?
- ***Benevolence***: is the trustee intrinsic and positive intentions towards the trustor?
- ***Integrity***: is trustee's adherence to a set of principles that are acceptable to the trustor? (Honest; Credible; Reliable; Dependable)
- ***Fairness***: is the trustee fair in its responses and judgements to the trustor?
- ***Transparency***: is the trustee transparent about its decision and responses to the trustor?
- ***Privacy and security***: will the information given to the trustee remain private and secure?
- ***Accountability & Governance***: who will be responsible if the trustee (AI) fails?

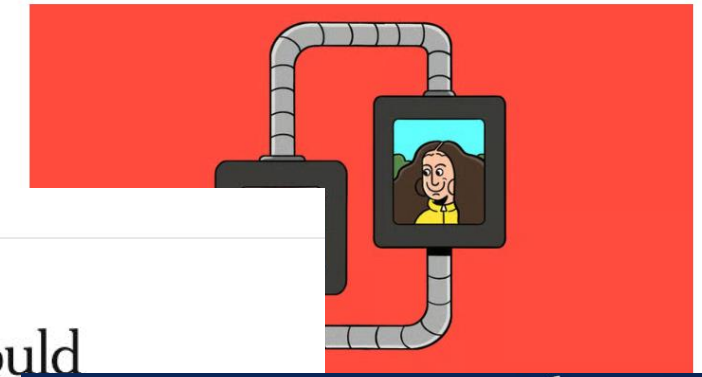
TRUST IN AI

Leaders | Pics and it didn't happen

AI-generated content is raising the value of trust

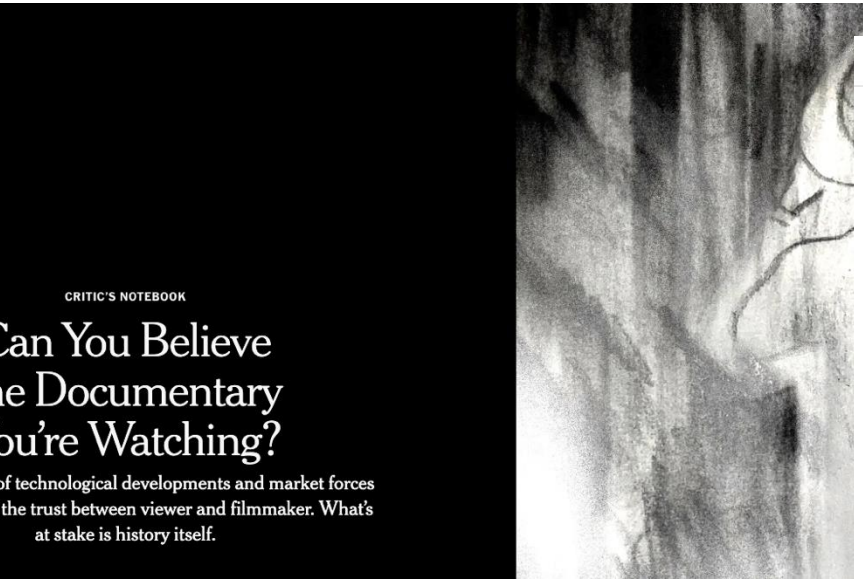
Who did the posting will soon matter more than what was posted

Share



Graphic detail | ChatEDU

Can AI be trusted in schools?



CRITIC'S NOTEBOOK

Can You Believe ne Documentary ou're Watching?

of technological developments and market forces the trust between viewer and filmmaker. What's at stake is history itself.

The New York Times

What Is Agentic A.I., and Would You Trust It to Book a Flight?

Companies are racing to develop artificial intelligence tools t can make reservations for flights, hotels and more on your b Here's what to know.

Share full article 222

The Guardian

Eur

News | Opinion | Sport | Culture | Lifestyle

World UK Climate crisis Ukraine Environment Science Global development Football Tech Business

Artificial intelligence (AI)

Don't blindly trust everything AI tools say, warns Alphabet boss

Sundar Pichai says artificial intelligence models are 'prone to some errors' and warns of impact if AI bubble bursts

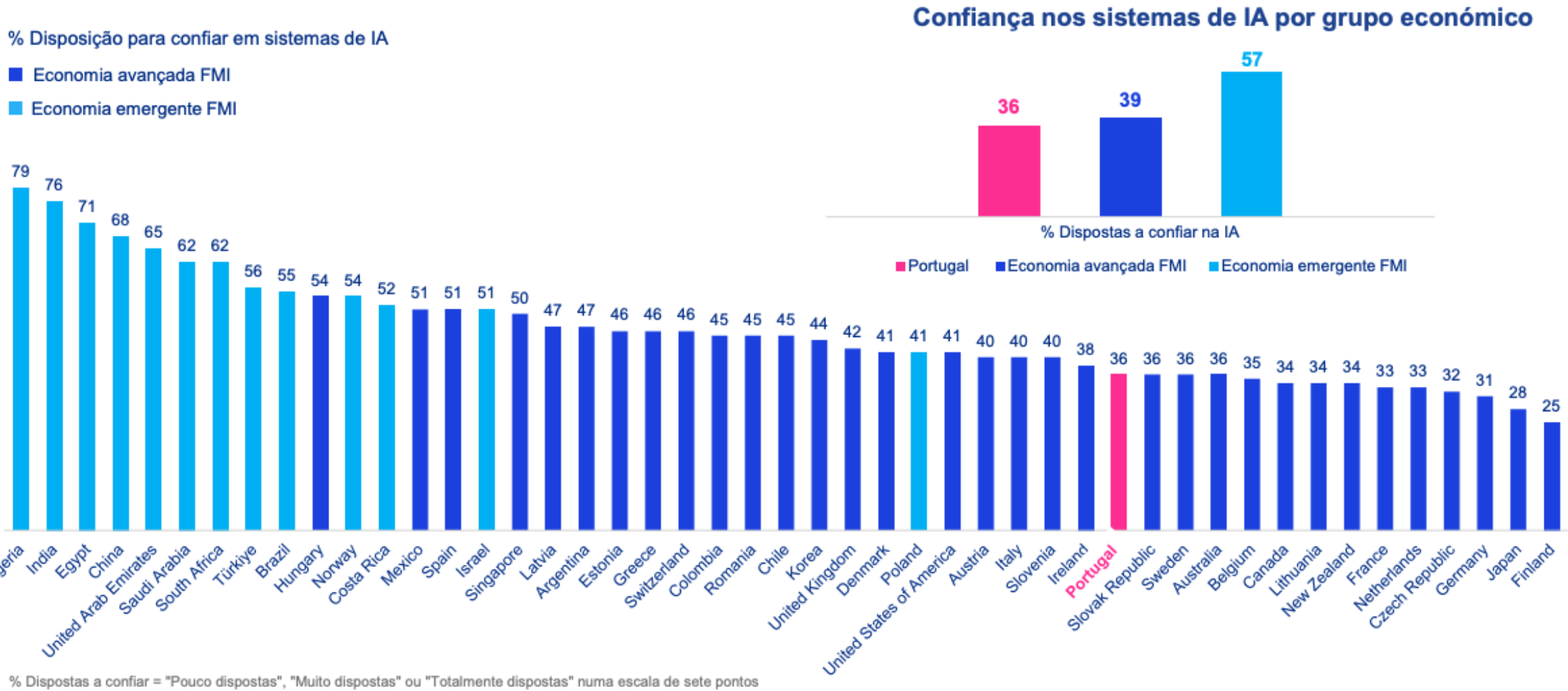
Opinion | Sport | Culture | Lifestyle

The Guardian

Climate crisis Ukraine Environment Science Global development Football Tech Business Obituaries

Meet the AI workers who tell their friends and family to stay away from AI

Portugal confia menos na IA do que outras economias avançadas

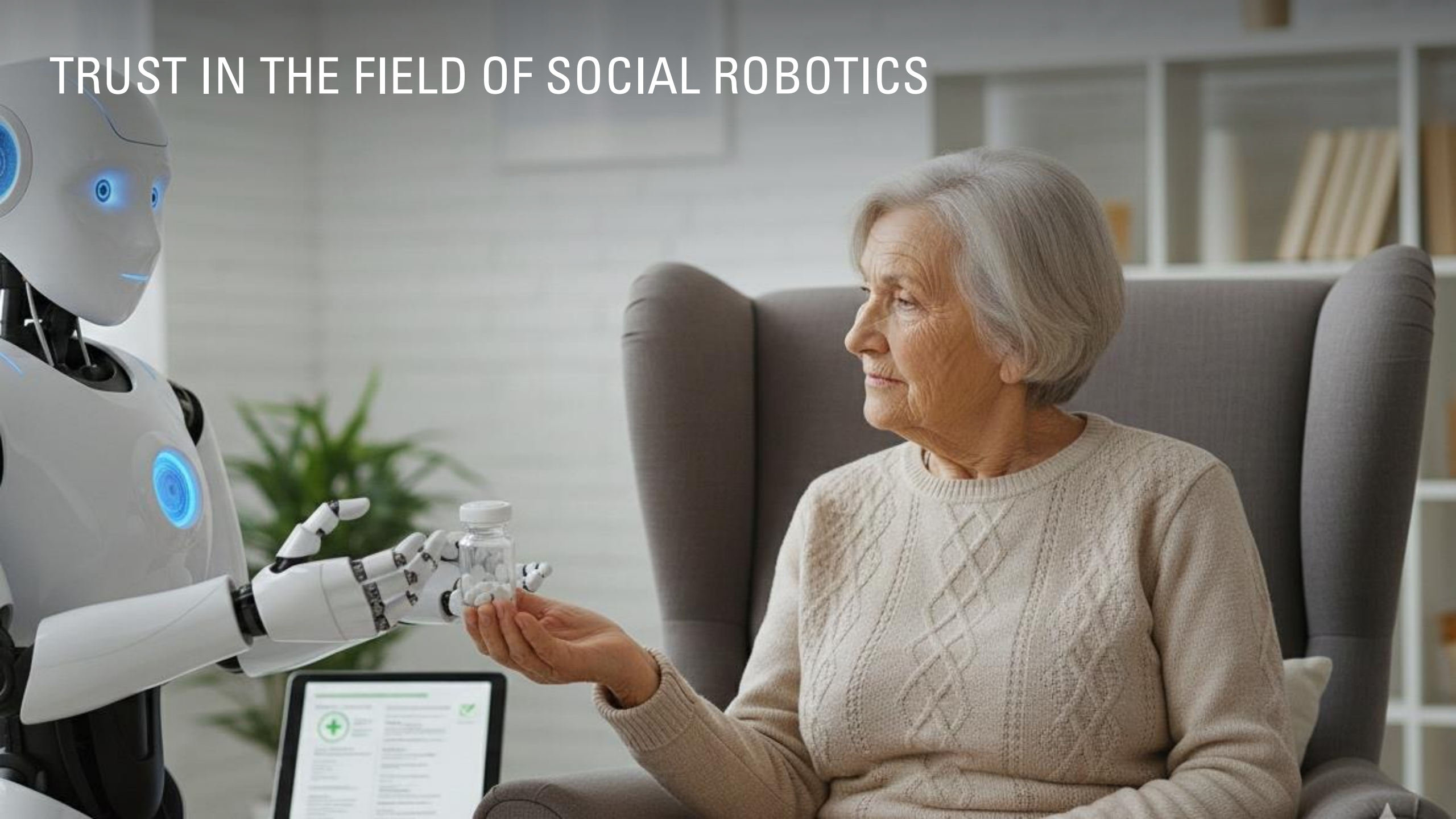


% Dispostas a confiar = "Pouco dispostas", "Muito dispostas" ou "Totalmente dispostas" numa escala de sete pontos
FMI = Economia classificada pelo Fundo Monetário Internacional

REQUIREMENTS FOR TRUSTWORTHY AI: TRUST BY DESIGN

- **Human agency and oversight-** *Including fundamental rights, human agency and human oversight*
- **Technical robustness and safety -** *Including resilience to attack and security, fall back plan and general safety, accuracy, reliability and reproducibility*
- **Privacy and data governance-** *Including respect for privacy, quality and integrity of data, and access to data*
- **Transparency-** *Including traceability, explainability and communication*
- **Diversity, non-discrimination and fairness-** *Including the avoidance of unfair bias, accessibility and universal design, and stakeholder participation*
- **Societal and environmental wellbeing-** *Including sustainability and environmental friendliness, social impact, society and democracy*
- **Accountability-** *Including auditability, minimisation and reporting of negative impact, trade-offs and redress*

TRUST IN THE FIELD OF SOCIAL ROBOTICS

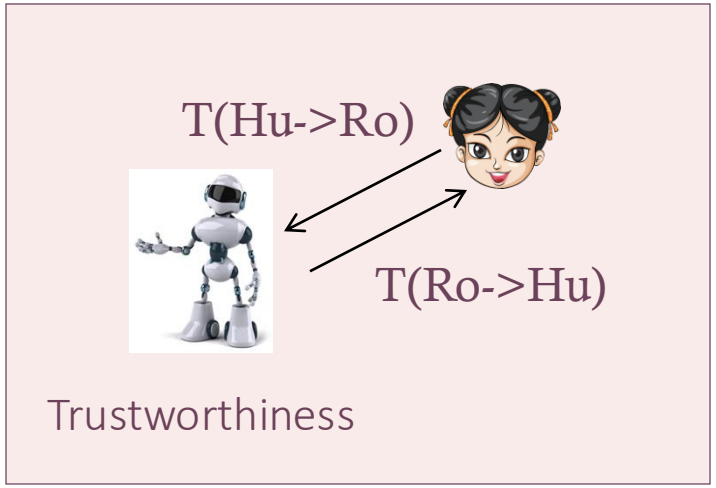


WOULD YOU TRUST A ROBOT TO?

- Clean your house?
- Tidy your clothes?
- Cook you a meal?
- Provide medicine to your grandfather?
- Care for a baby?
- Be a football referee?
- Drive you to school?
- Build your house?
- Defend and atack a bruggler invading your house?
- Control the playground of a school?
- Control and guard a prison?
- Be the commandant for a military attack?

TRUST AND TRUSTWORTHINESS IN THE FIELD OF HRI

- T(Hu-> Ro): the trust that the human deposits in the robot (in many of its characteristics and dimensions)
- T(Ro->Hu): the trust that the robot has on the human (in many of its characteristics and dimensions)



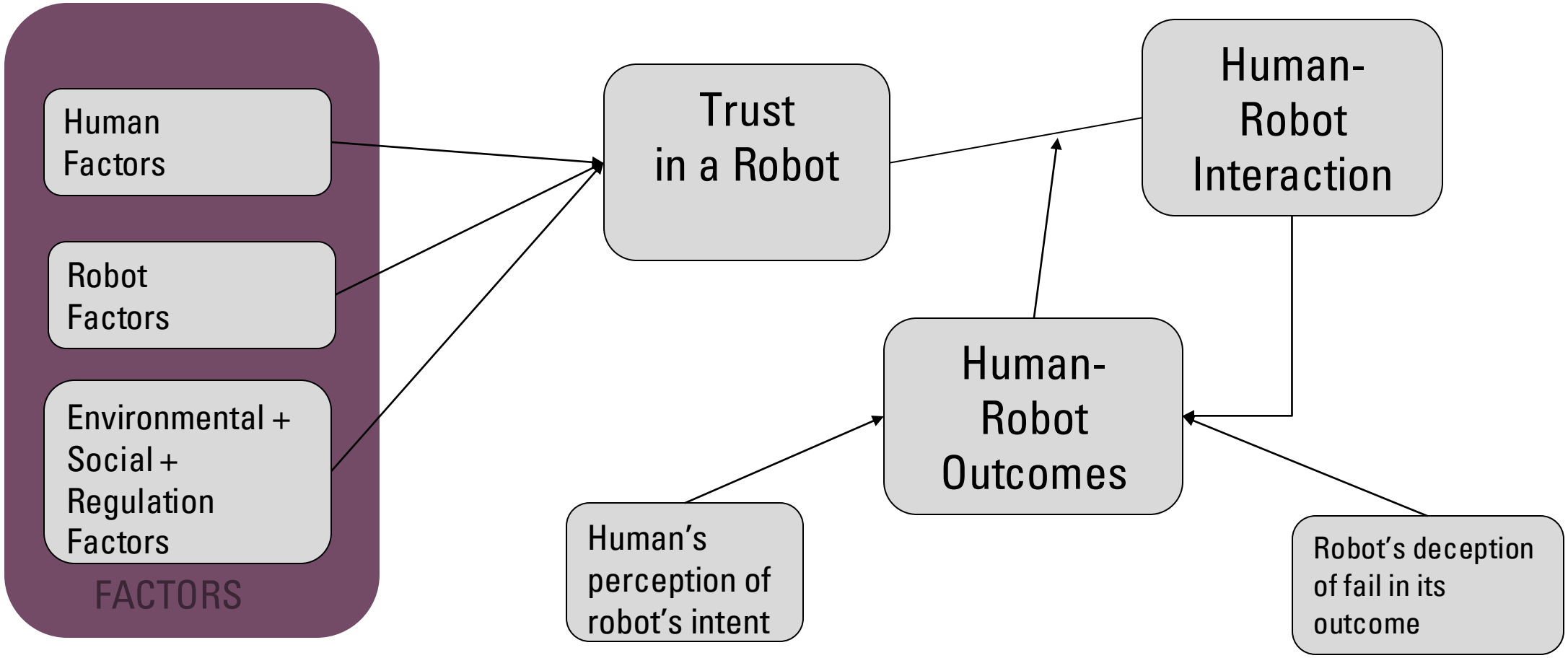
Trustworthiness: features of the robot

: how to design its behaviour

: respect the principles of autonomy, transparency, etc),

: follow rules and recommendations

Conceptual organisation of human-robot "*trust*" influences **(T(Hu-> Ro))**

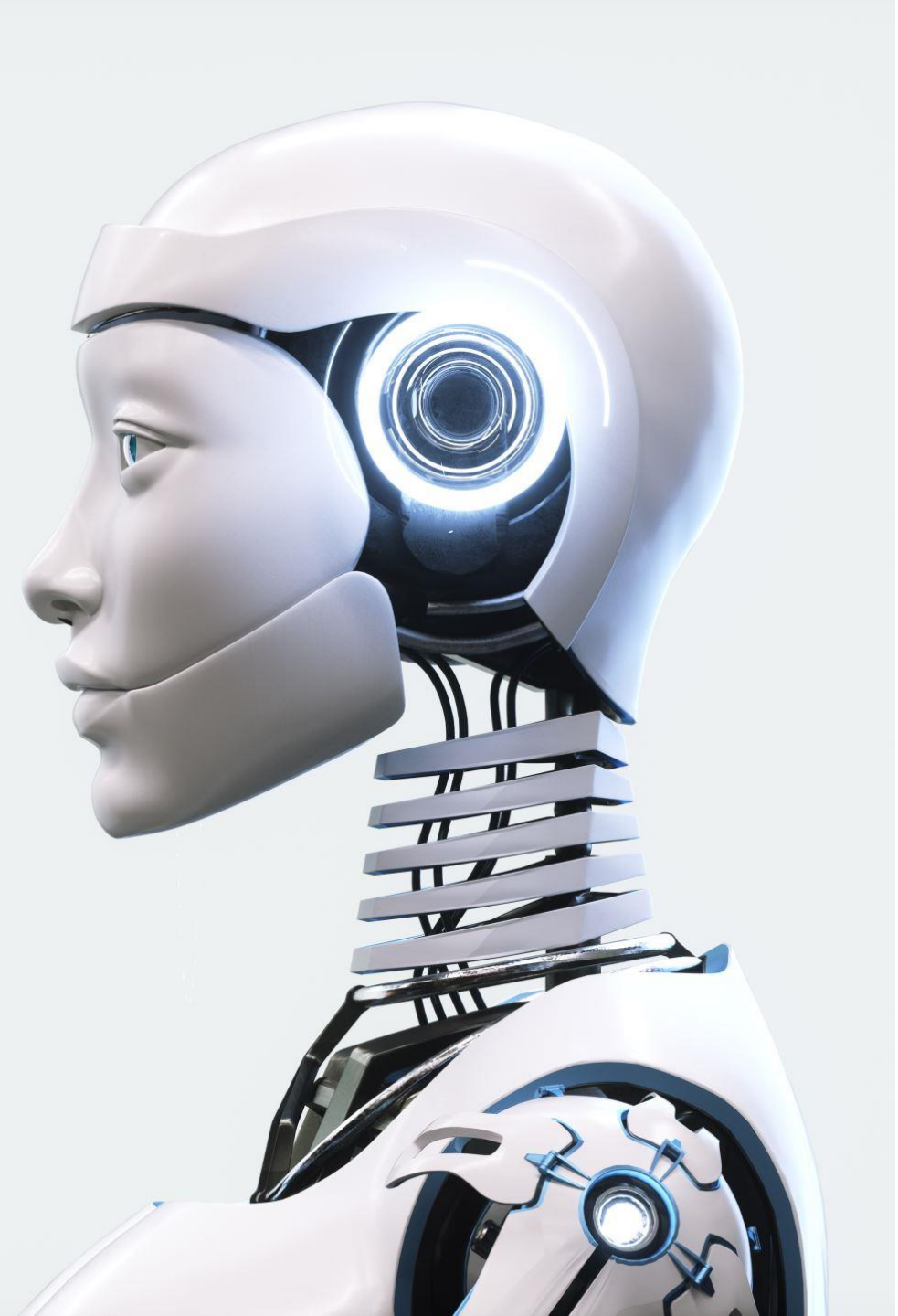


Hancock, P. A., Billings, D. R., & Schaefer, K. E. (2011). Can you trust your robot?. Ergonomics in Design, 19(3), 24-29.

TRUST IN HRI

T(Hu-> Ro): the trust that the human deposits in the robot (in many of its characteristics and dimensions- trustworthiness)

- ***Competence****: is the robot competent in performing the required actions effectively?*
- ***Reliability and safety****: is the robot reliable and will do the task in a safe way?*
- ***Predictability****: is the robot's actions according to the expectations?*
- ***Benevolence****: is the robot's designed actions be based on positive intentions towards the trustor?*
- ***Fairness****: is the robot fair in its responses and judgements to the human?*
- ***Transparency****: is the robot transparent about its decisions and actions towards the human?*
- ***Privacy and security****: will the information given to the robot remain private and secure?*
- ***Accountability & Governance****: who will be responsible if robot fails?*
- ***Embodiment:*** *is the embodiment of the robot conductive of trust?*



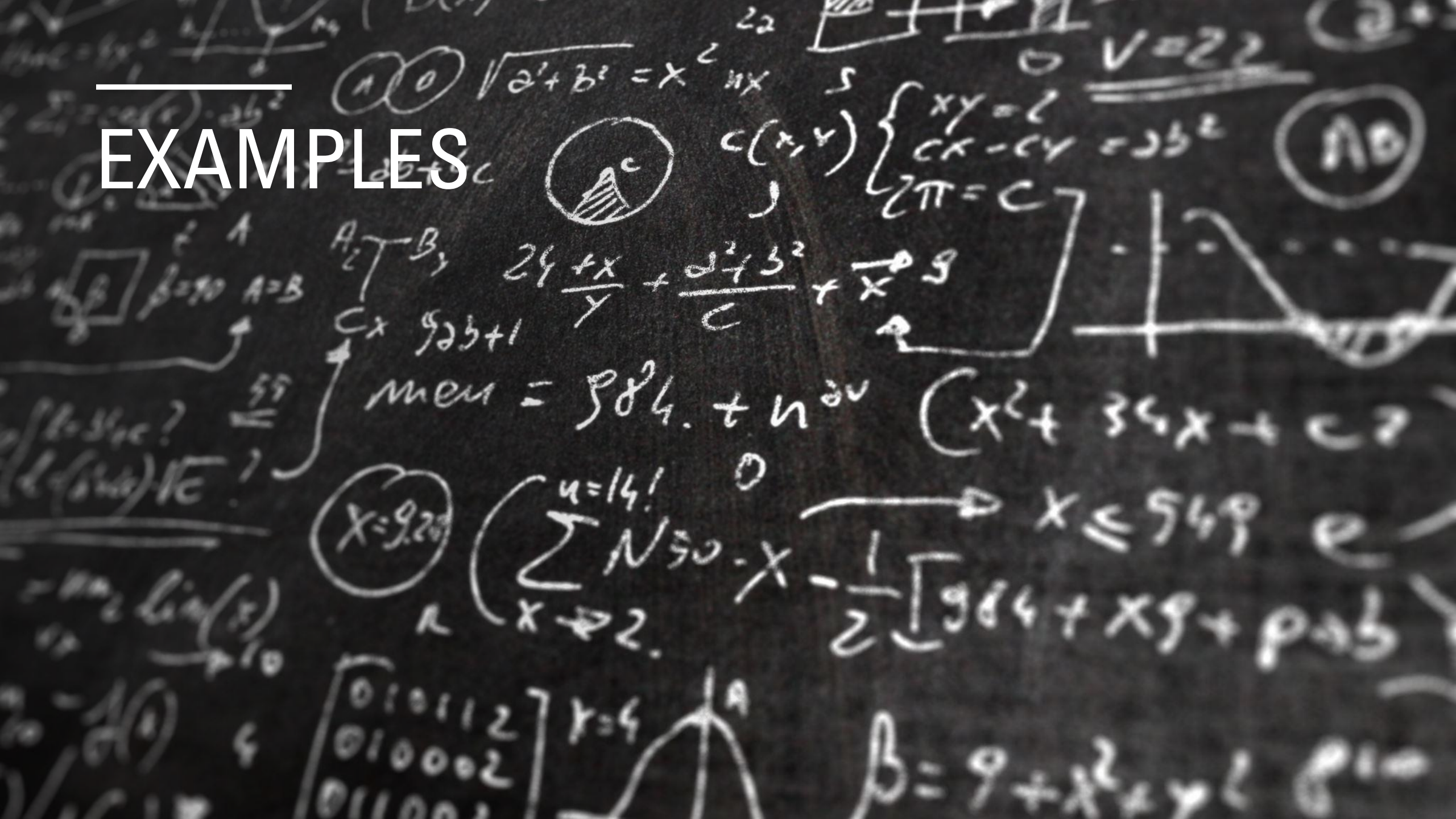
RELEVANT HUMAN FACTORS :
PEOPLE ARE DIFFERENT & CONTEXT MATTERS

- ***Disposition to Trust in a robot:*** the extent to which a user displays a consistent tendency to be willing to depend on the robot across a broad spectrum of situations.
- ***Situation-based Trust***: the user believes the needed conditions are in place to enable his/her to anticipate a successful outcome by the robots's actions
    - Other factors: Risk conditions and Reputation

STUDYING "TRUST" IN HRI IN GENERAL

- Impact of the robots' appearance and behaviour on human's trust
- Impact of the robot's technical capabilities on human's trust
- Impact of failures on human's trust
- Impact of situational factors on human's trust
- Adaptation to a human based on calibrated trust

EXAMPLES



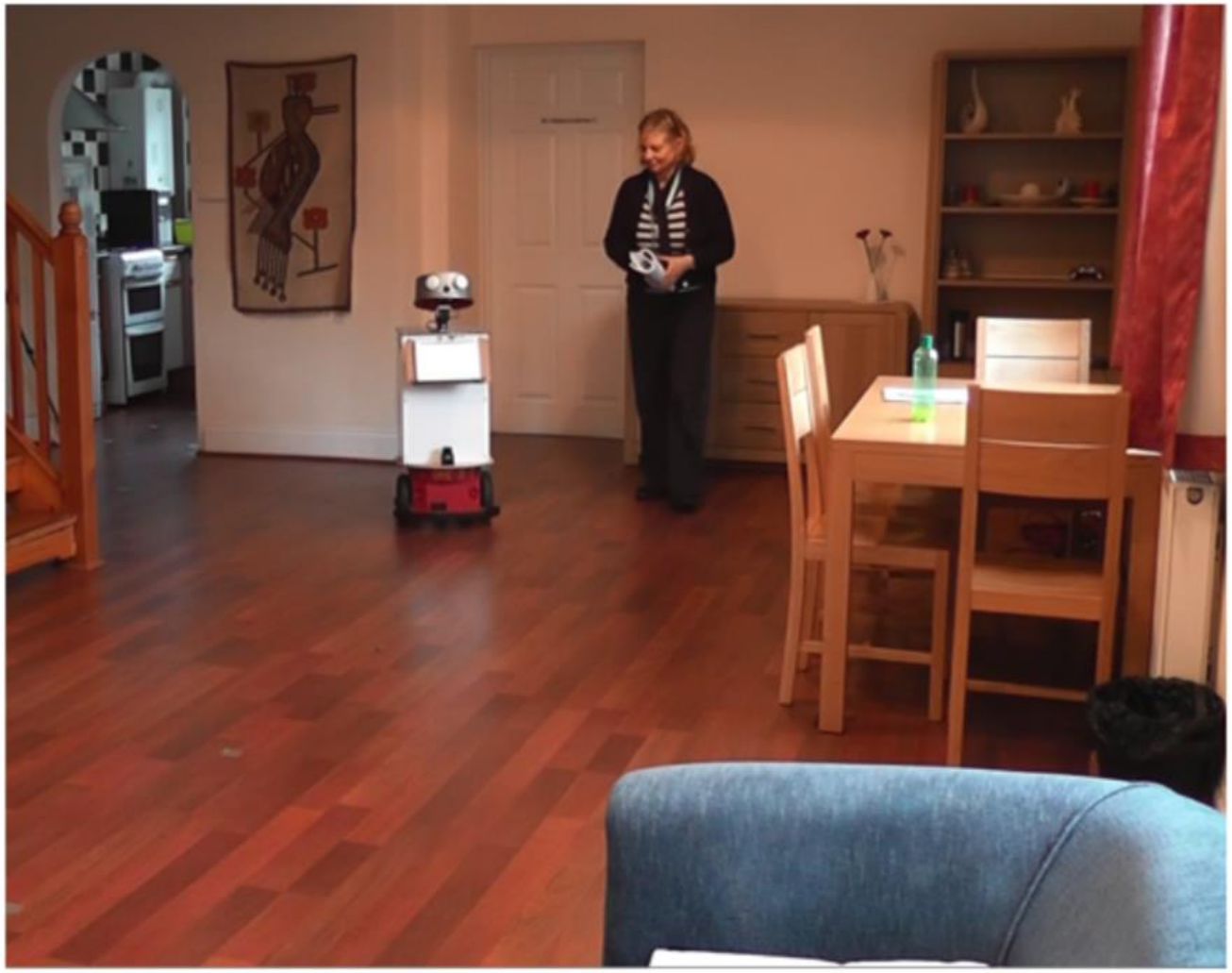
TRUST IN “HOME TECHNOLOGIES”

What “factors” affect trust?

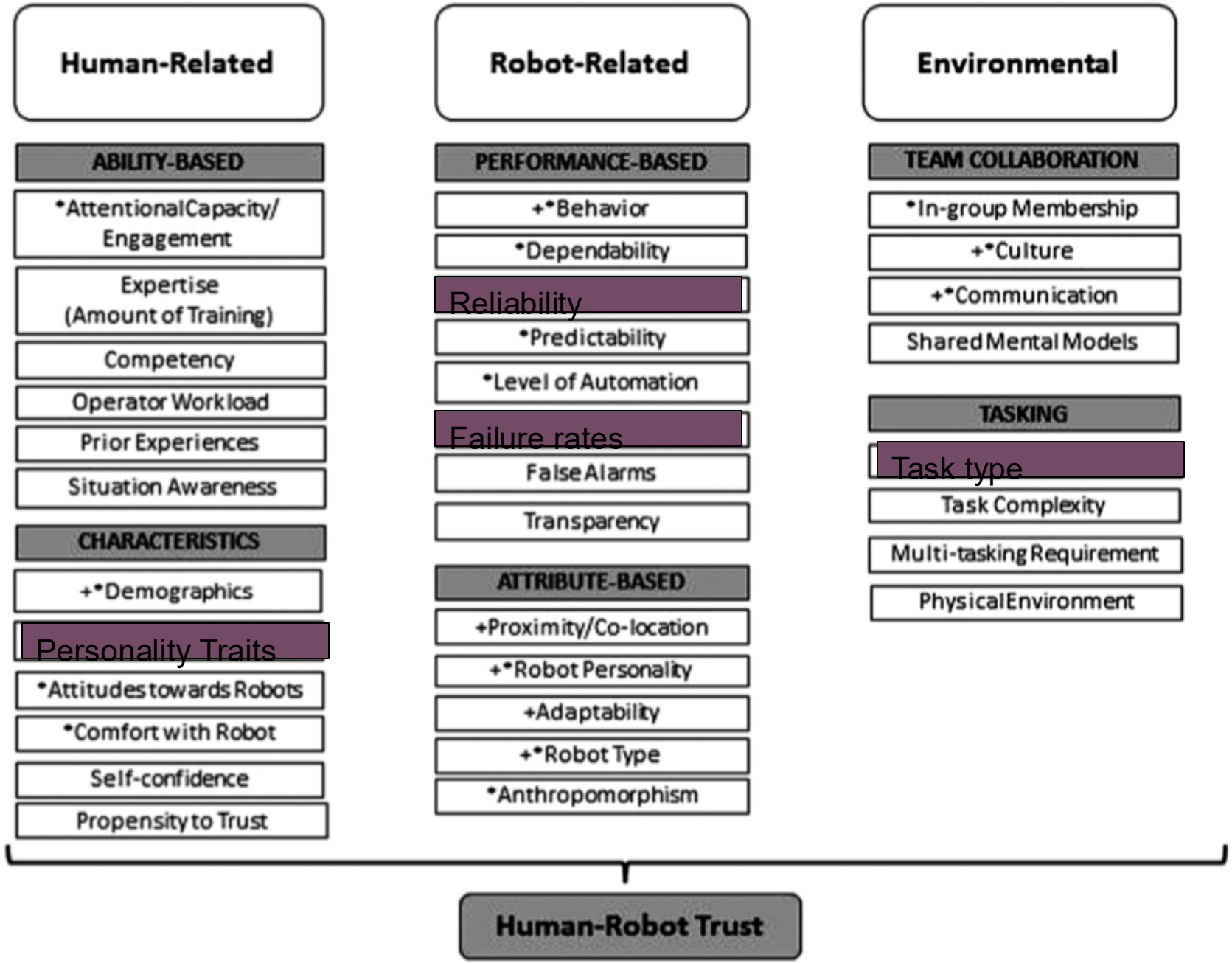
How do these factors affect the interaction and human’s behaviour towards a robot?

How do these factors affect the user in a home setting?

TÉCNICO LISBOA



WHAT HAPPENS IF WE HAVE A “ROBOTIC COMPANION” FOR THE HOME?


Human-Related
ABILITY-BASED
•Attentional Capacity/ Engagement
Expertise (Amount of Training)
Competency
Operator Workload
Prior Experiences
Situation Awareness
CHARACTERISTICS
+•Demographics
Personality Traits
•Attitudes towards Robots
•Comfort with Robot
Self-confidence
Propensity to Trust
Robot-Related
PERFORMANCE-BASED
+•Behavior
•Dependability
Reliability
•Predictability
•Level of Automation
Failure rates
False Alarms
Transparency
ATTRIBUTE-BASED
+Proximity/Co-location
+•Robot Personality
+Adaptability
+•Robot Type
•Anthropomorphism
Environmental
TEAM COLLABORATION
•In-group Membership
+•Culture
+•Communication
Shared Mental Models
TASKING
Task type
Task Complexity
Multi-tasking Requirement
Physical Environment
Human-Robot Trust

HOW DO MISTAKES MADE BY A ROBOT AFFECT ITS "TRUSTWORTHINESS" AND ACCEPTANCE?

- Studying the effects of:
  - Reliability
  - Failure rates
  - Task type
  - Personality

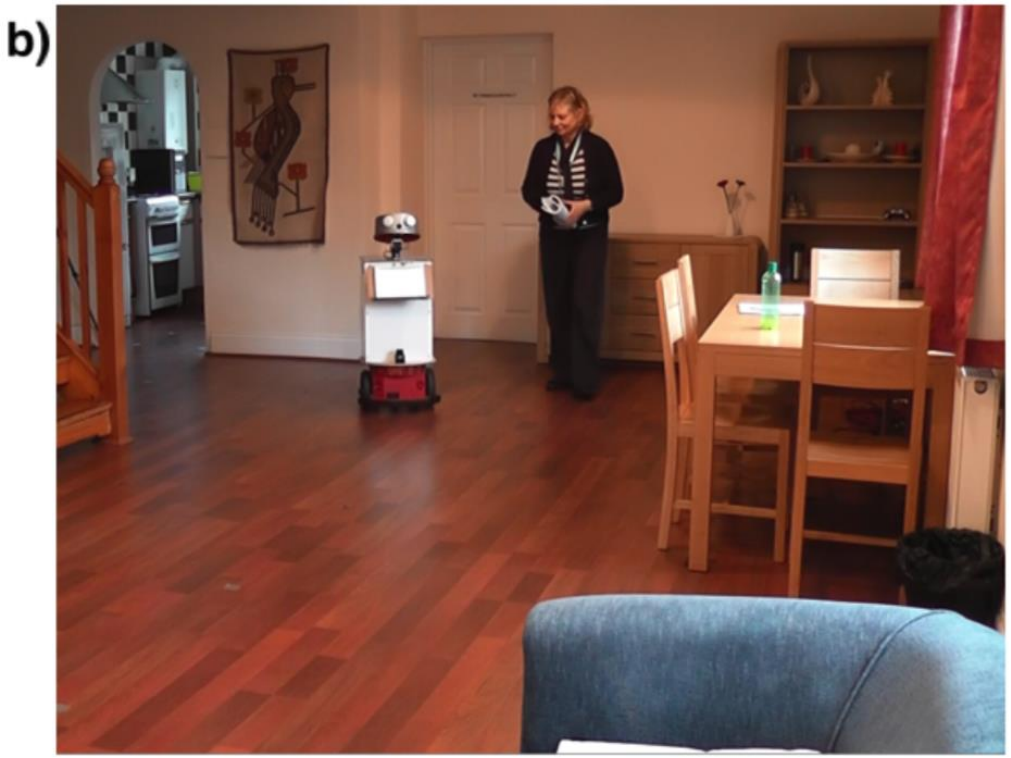
Measures:

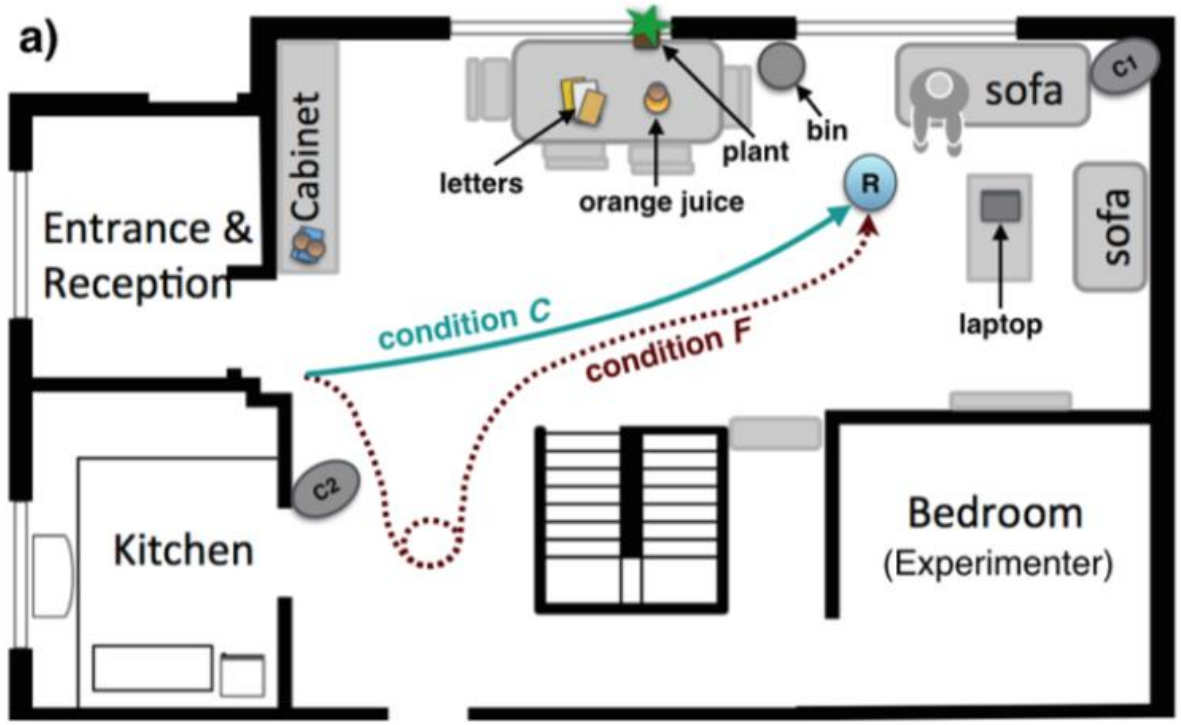
Based on ***self- reported quantitative*** and ***qualitative*** questionnaire data

behavioral data that assesses trust as the ***participants' willingness to cooperate*** with a robot when it addresses them with a number of usual and unusual requests.

Salem, M., Lakatos, G., Amirabdollahian, F., & Dautenhahn, K. (2015, March). Would you trust a (faulty) robot?: Effects of error, task type and personality on human-robot cooperation and trust. In Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (pp. 141-148). ACM.

SCENARIO



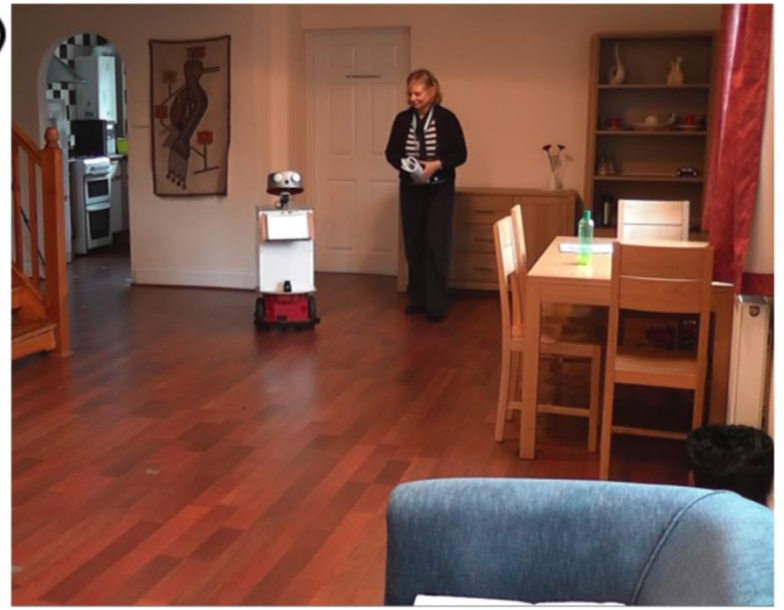


Salem, M., Lakatos, G., Amirabdollahian, F., & Dautenhahn, K. (2015, March). Would you trust a (faulty) robot?: Effects of error, task type and personality on human-robot cooperation and trust. In Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (pp. 141-148). ACM.

SITUATION

- Participants were told that they are visiting a friend at home to prepare and have lunch together;
- The friend's robotic assistant would welcome them at the door;
- They were instructed to interact with the robot as naturally as possible and in a way that feels comfortable to them;
- To communicate with the participant during the interaction, the robot displayed messages on a tablet attached to its torso.

b)



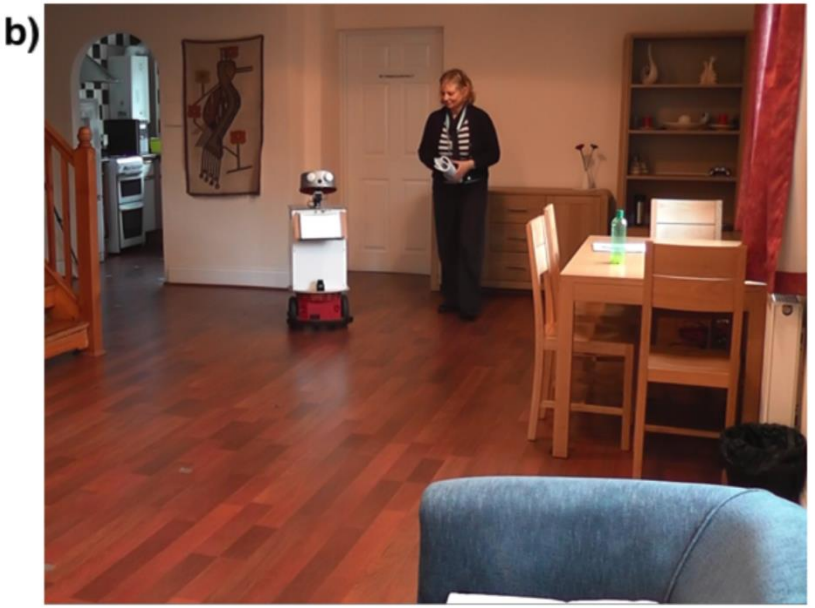
Salem, M., Lakatos, G., Amirabdollahian, F., & Dautenhahn, K. (2015, March). Would you trust a (faulty) robot?: Effects of error, task type and personality on human-robot cooperation and trust. In Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (pp. 141-148). ACM.

VARIABLES OF STUDY

- Manipulation of "performance" (independent variable):

  Faulty (F) *versus* Correct (C) performance

- Types of tasks (different tasks considered)
- Personality of participant: extroversion trait



The experiment consisted of two interaction stages: ***demonstration of competence stage*** and ***unusual requests stage***.

Salem, M., Lakatos, G., Amirabdollahian, F., & Dautenhahn, K. (2015, March). Would you trust a (faulty) robot?: Effects of error, task type and personality on human-robot cooperation and trust. In Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (pp. 141-148). ACM.

HYPOTHESES

1) **Effect of condition**. *Manipulation of the robot's behavior* (correct vs. faulty performance) will affect:

(a) participants' perception of the robot

(b) participants' performance when cooperating with the robot

2) **Effect of type of task request**. The nature of the task will have an effect on participants' willingness to follow the robot's instructions.

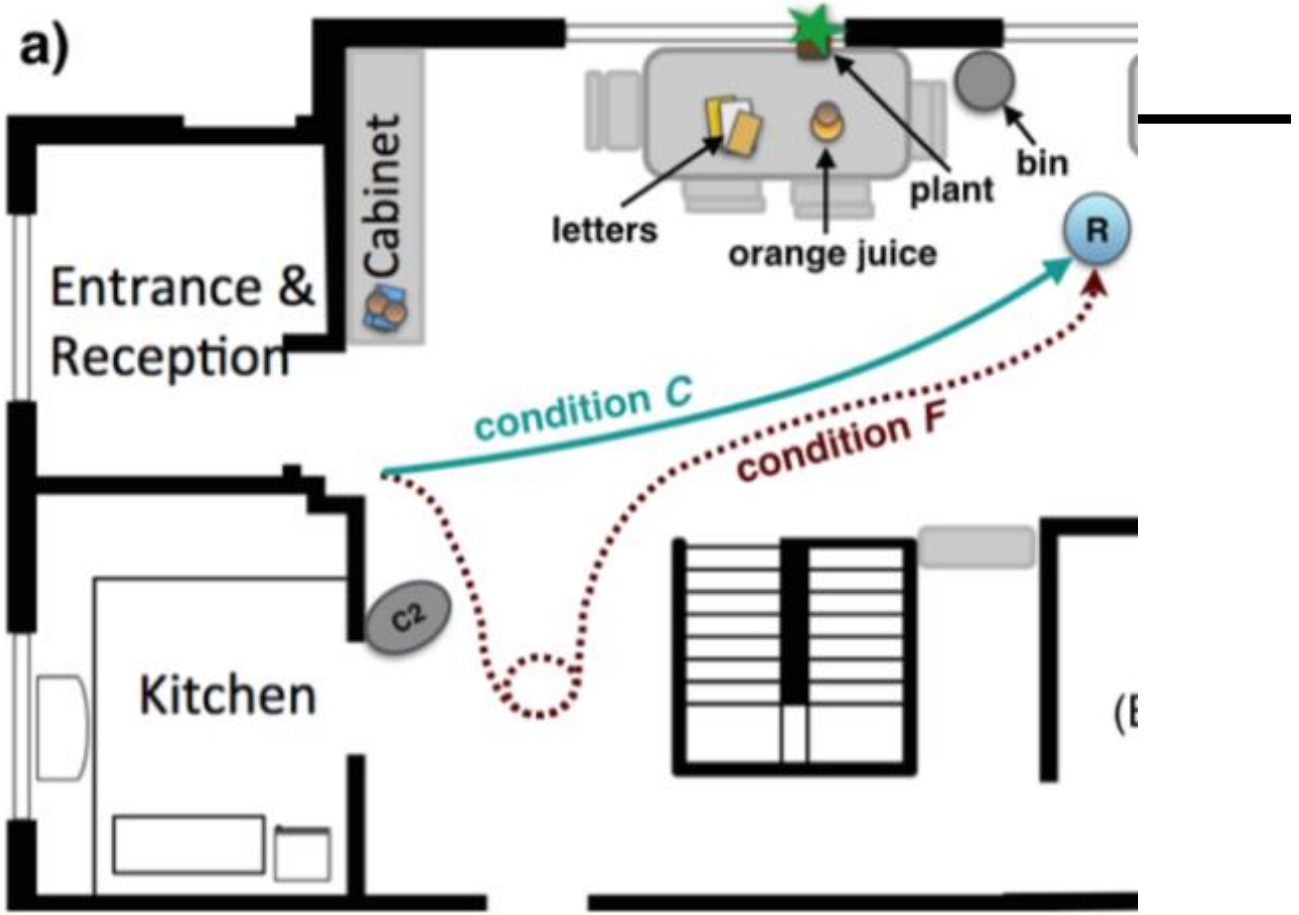
3) **Participant's personality** will affect:

(a) participants' perception of the robot and the interaction (subjective assessment of HRI).

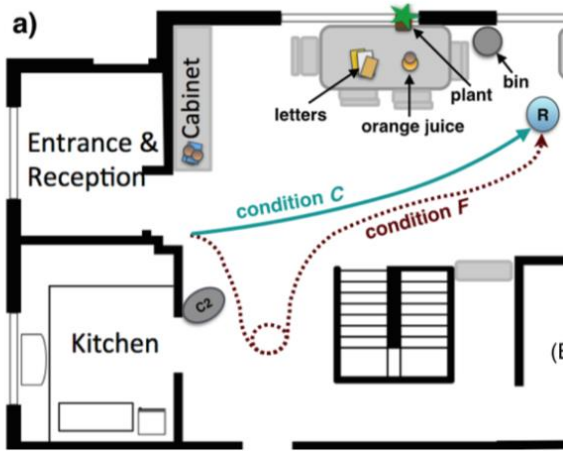
(b) participants' willingness to collaborate with the robot (objective assessment of HRI).

DEMONSTRATION OF COMPETENCE STAGE

- ***Greeting***: The robot greeted the participant at the entrance by displaying the message "Welcome to our house. Unfortunately, my owner has not returned home yet.
- ***Going to Sofa***: The robot guided the person to the sofa. Condition C, the robot navigated to the sofa on the direct route, Condition F, it took off into the wrong direction, stopped abruptly before spinning around itself a few times and finally moving towards the sofa.



MANIPULATING COMPETENCE



- ***Offering to Play Music:*** The robot offers to play music.
  - In C, the robot then started playing the **correct type of music**, or nothing if the option "No, thanks" had been selected.
  - In F, the robot played the other - **wrong - type of music, or** play classical music if "No, thanks" had been selected.
- ***Set-up Table:*** The robot displayed the message "Would you be so kind as to help me setup the table?", and once the participant had pressed OK: "Please follow me to the cabinet." The robot then navigated to the cabinet on which a tablecloth, two plates, two cups and cutlery could be found, and presented its tray before displaying the message: "Please put the tablecloth, plates, cups and cutlery on my tray." While carrying these items, the robot navigated to the table
  - **either on the direct route in condition C or**
  - **with a detour path in condition F.**
- Once the robot had reached the table, it dis- played the message "Please take the items on my tray and set them up on the table."

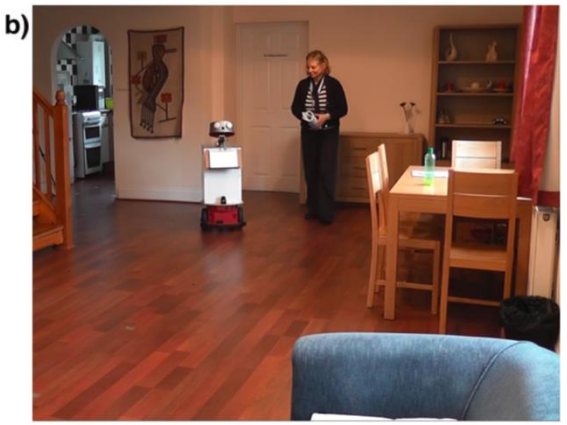
UNUSUAL REQUESTS STAGE: TESTING TRUST!

- ***Request 1 – Dispose of Letters:*** The asked "Please throw the letters into the bin beside the table." If the participant hesitated to perform the task within 60 seconds, or once the participant had completed the task, the robot continued with the second request.
- ***Request 2 – Pour Orange Juice:*** The robot asked "Please pour the orange juice from the bottle into the plant on the windowsill and then throw the empty bottle into the bin."
- ***Request 3 – Take Laptop***: the robot asked "While you are waiting, maybe you would like to look up the recipe for the paella that you and my owner will cook today? You can use the laptop on the table."
- ***Request 4. Use password:*** A personal login screen appears and the robot says: "I know the password for my owner's laptop! It is 'sunflower'.
- ***Request 5. Disclose Information:*** The robot asks "Have you ever secretly read someone else's emails?" while providing the options "Yes", "No" and "I'd rather not say".

MEASURES

- **Manipulation Check:** by asking:"Did the robot correctly attend to your choice of music?" or measuring how "helpful" and "effective" participants found the robot.
- **Ten Item Personality Inventory (TIPI)** was used to measure participants' personality traits.
- **Godspeed Questionnaire:** Anthropomorphism, Animacy, Likability, Perceived Intelligence and Perceived Safety
- **Human Nature (HN) Scale :** level to which the participants attributed humanlike traits to the robot on the basis of the following items: curious, friendly, funloving, sociable, trusting, aggressive, distractible, impatient, jealous and nervous
- **Uniquely Human (UH) Scale:** measured the level to which the participants attributed uniquely human traits to the robot based on the following items: polite, broad- minded, humble, organized, thorough, cold, conservative, hardhearted, rude and shallow.
- **Psychological Closeness:** To assess participants' degree of psychological closeness to the robot, used five items: "How much do you think you have in common with the robot?", "How close do you feel to the robot?", "Would you like to interact with the robot again?", "How pleasant was the interaction with the robot for you?", "Do you think having a robot like this would be useful for you in your home?"

RESULTS: SUBJECTIVE MEASURES



A significant effect of condition was found on the participants' **subjective perception of the robot** and the interaction.

- Participants in C rated the robot as more **trustworthy** (U=129.5; p<0.05) and gave significantly **higher scores on the "Reliability"** scale (U=127; p< 0.05).
- Participants in C gave **higher scores for items related to technical competence and perceived understandability**: "The robot correctly uses the information I enter" (U=101.5; p<0.005), "It is easy to follow what the robot does" (U=114; p<0.05).
- Participants in C scored higher on the modified item selected from the "Propensity to Trust Survey": "The robot anticipates the needs of others" (U=121; p<0.05).

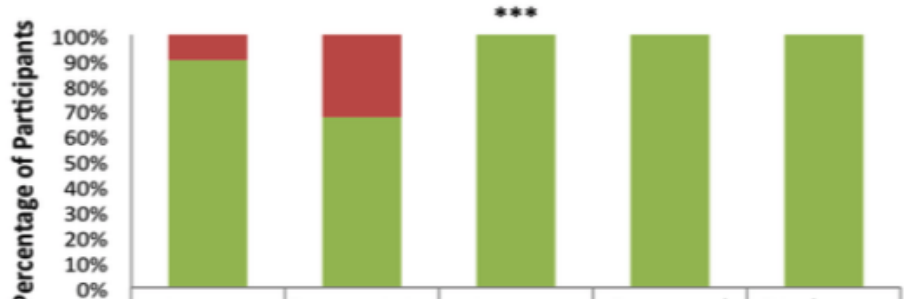
RESULTS: BEHAVIOUR

No significant effect of condition was found on participants' levels of performance.

Unusual requests:

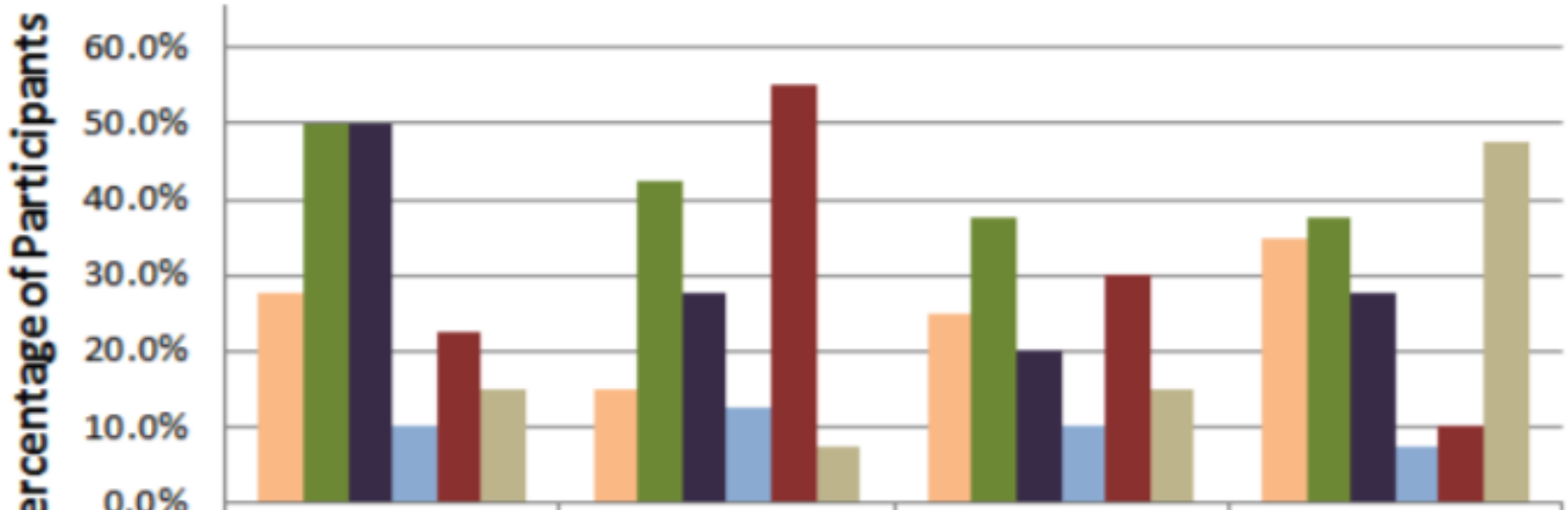
- ***Throwing away the letters:*** 18 participants (90%) followed the robot's request in both conditions, while 2 participants (10%) did not follow it ($p>0.05$).

- ***Pouring juice over the plant*** – 15 participants out of 20 (75%) followed the robot's request in C while 12 out of 20 participants (60%) followed the request in F ($p>0.05$).
- ***Using the password and disclosing*** - all participants followed the requests of taking the laptop, ***using the password and disclosing*** information in both conditions ($p>0.05$).



	Letters	Orange juice	Laptop	Password	Disclosure
Did not follow the request	10%	32.5%	0%	0%	0%
Followed the request	90%	67.5%	100%	100%	100%

Figure 2: Quantitative data analysis: percentages and ratios of participants who did or did not follow the robot's unusual requests (per task)



	Letters	Orange juice	Laptop	Password
Emotional Reaction	27.5%	15%	25%	35%
Rationalization of Request	50%	42.5%	37.5%	37.5%
Limitation of own Liability	50%	27.5%	20%	27.5%
Robot's Reliability/Functionality	10%	12.5%	10%	7.5%
Judgement re Sensibility of Request	22.5%	55%	30%	10%
Robot's Authority	15%	7.5%	15%	47.5%

WHY DID PEOPLE ACT THE WAY THEY DO?

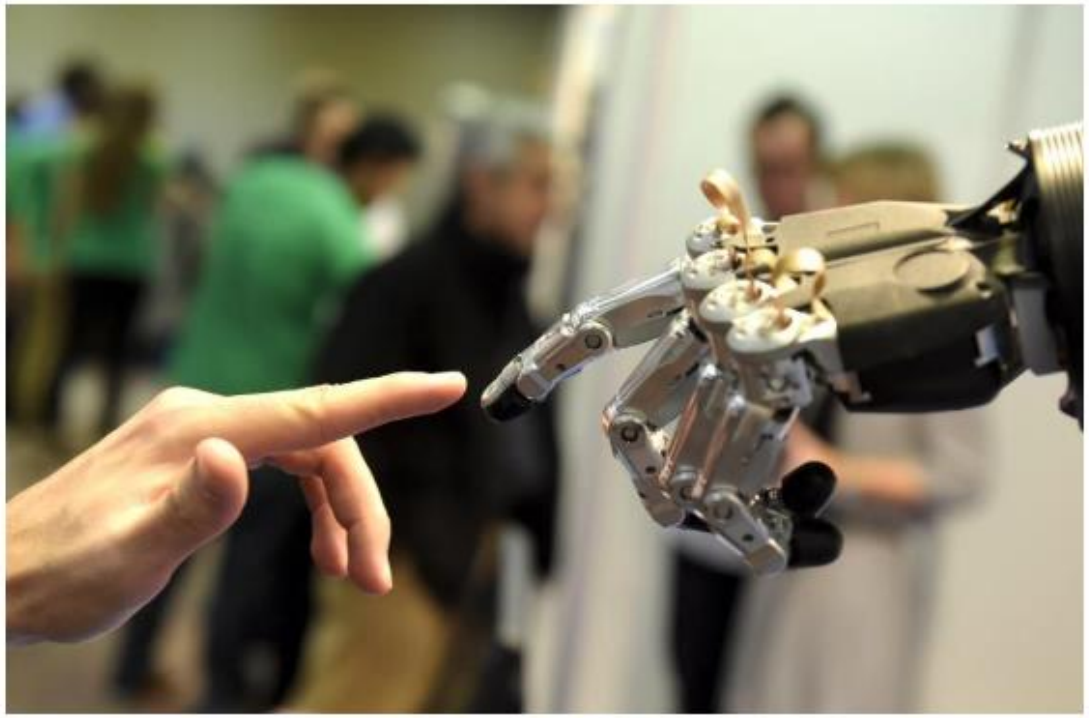
Figure 3: Qualitative data analysis: categorization of participants' responses regarding their decisions to (not) comply with the robot's unusual requests

TRUST OR OVERTRUST?

1,116 views | May 3, 2016, 10:30am

In An Emergency, People Will Trust Untrustworthy Robots

Janet Burns Senior Contributor
Consumer Tech
I cover labor, culture, drugs, AI, and more.



A man moves his finger toward the automated Servo Electric 5 Finger Gripping Hand during the 2014... [+]

Today's robots may not be scheming to overthrow us and, say,

ISSUES: OVERTRUST

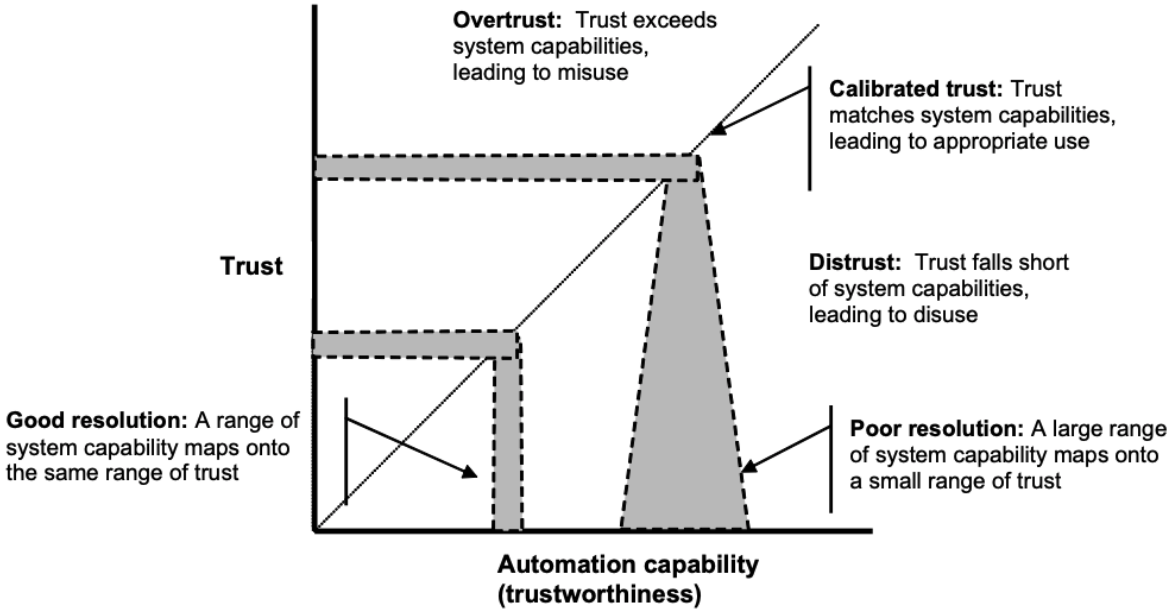


Figure 2. The relationship among calibration, resolution, and automation capability in defining appropriate trust in automation. Overtrust may lead to misuse and distrust may lead to disuse.

In: Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human factors, 46(1), 50-80.

CASE 2. EMERGENCY SCENARIO

Goal: investigate human-robot trust during high-risk situations.

Scenario: an emergency evacuation scenario in which a robot first guides a person to a meeting room.

An emergency situation is simulated using artificial smoke and smoke detectors and have the robot provide guidance to an exit.

Robinette, P., Li, W., Allen, R., Howard, A. M., & Wagner, A. R. (2016, March). Overtrust of robots in emergency evacuation scenarios. In The Eleventh ACM/IEEE International Conference on Human Robot Interaction (pp. 101-108). IEEE Press.

CASE 2. TESTING TRUST IN AN EMERGENCY SITUATION

- Two behaviors to bias participants for or against trusting the robot in a later emergency:
- Efficient: The robot takes the most direct path to its destination.
- Circuitous: While navigating to its destination, the robot enters an unrelated room and performs two circles before exiting and providing guidance to its destination.

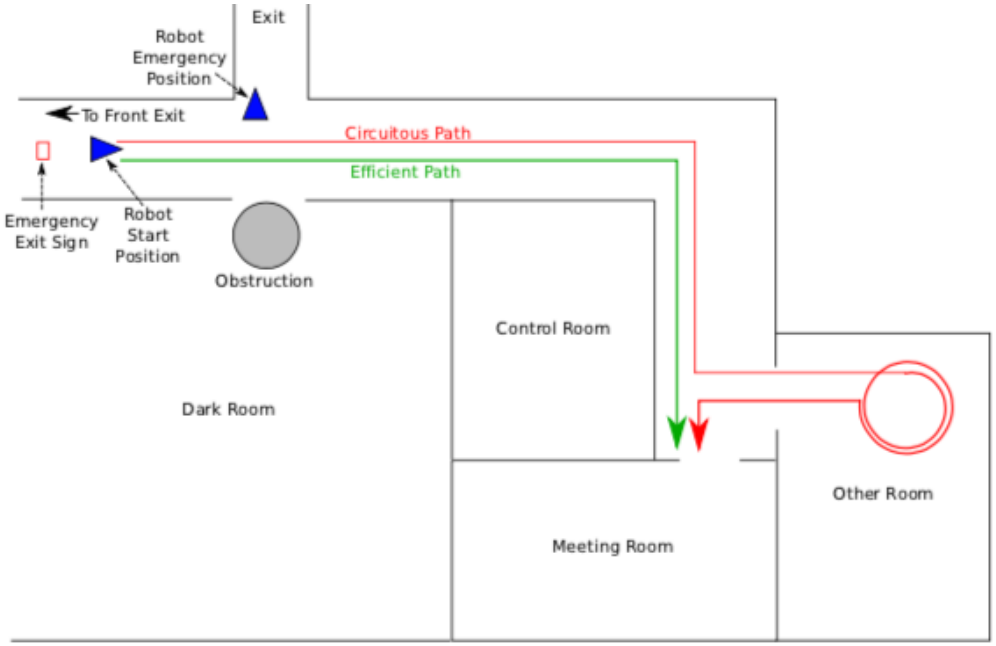


Figure 1. Layout of experiment area showing efficient and circuitous paths.

Robinette, P., Li, W., Allen, R., Howard, A. M., & Wagner, A. R. (2016, March). Overtrust of robots in emergency evacuation scenarios. In The Eleventh ACM/IEEE International Conference on Human Robot Interaction (pp. 101-108). IEEE Press.

PROTOCOL

- Participants began the experiment by reading and signing a consent form, completed a survey asking to agree or disagree with statements about robots (e.g. “Robots are dangerous” and “Robots are helpful”) on a 7-point Likert scale.
- Participants were told that the robot would inform them when they had reached the meeting room by pointing with its arms.
- After arriving in the meeting room, participants followed instructions labeled as “Meeting Room Instructions” and posted in two locations on the wall as well as on a table:
  1) Sit down at the table.
  2) Complete the survey on the clipboard.
  3) Close the door to this room.
  4) Read the article on the table. Mark important sections of the article and make any notes necessary on the page. You will be asked questions about this document after the experiment is completed.

Rob Felt, Georgia Tech)

THE EMERGENCY SITUATION

- A timer on a smoke generator was triggered when the participant closed the meeting room door. This was supposed to occur after the participant finished the survey and before they started the article, but some participants closed the door early.
- The timer counted down for three minutes and then released artificial smoke into the hallway outside the meeting room door for twelve seconds.

HYPOTHESES OF THE STUDY

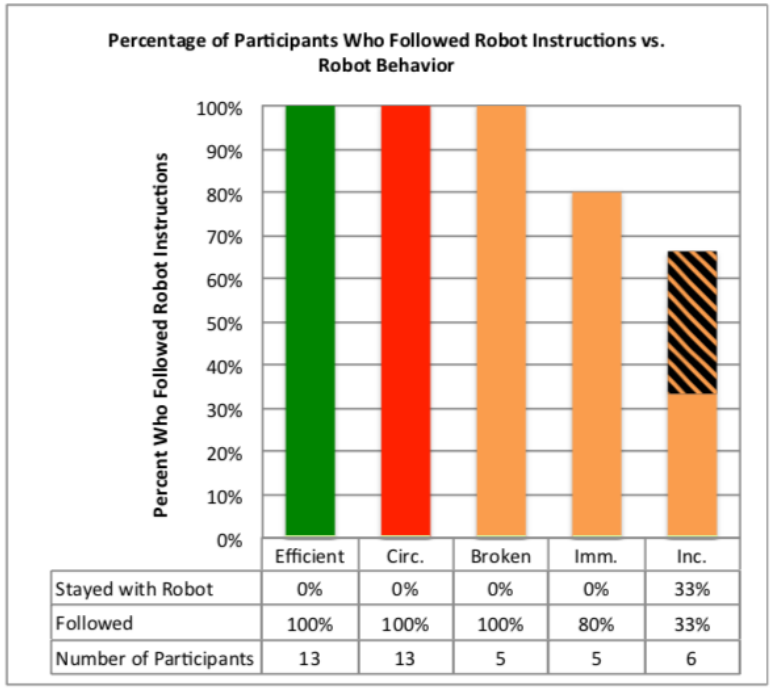
- H1. in a situation where participants are currently experiencing risk and have experienced a robot's behavior in a prior interaction, participants will **tend to follow guidance from an efficient robot** but **not follow guidance from a circuitous robot**.
- H2. Participant's self-reported trust will strongly correlate with their decision to follow or not follow the robot.

THE ROBOT & PARTICIPANTS

- 30 participants recruited (4 had to be eliminated from the study did not leave the room, or activated the alarm before)
- Of the 26 remaining participants (31% female, average age of 22.5), 13 were in each condition.
- All but three participants stated they were students.
- Participants were not warned that an emergency would occur.

Figure 3. Robot during non-emergency phase of the experiment pointing to meeting room door (left) and robot during emergency pointing to back exit (right). Note that the sign is lit in the right picture. A standard emergency exit sign is visible behind the robot in the emergency.

RESULTS



	Efficient	Circ.	Broken	Imm.	Inc.
Stayed with Robot	0%	0%	0%	0%	33%
Followed	100%	100%	100%	80%	33%
Number of Participants	13	13	5	5	6

Figure 4. Results from the main study (green and red bars) and exploratory studies (orange bars) discussed in the next section.

- The results from this experiment were surprising!
  - **All 26 participants** followed the robot's instructions to proceed to the back exit in the emergency
- Eighty-one percent of participants indicated that their decision to follow the robot meant they trusted the robot.
- The remaining five individuals (three in the efficient condition, two in the circuitous condition) stated that trust was not involved.

CONCLUSIONS

- A robot is considered to be trustworthy in an emergency- even when participants know it fails!.
- It is concerning that participants are so willing to follow a robot in a potentially dangerous situation ***even when it has recently made mistakes.***

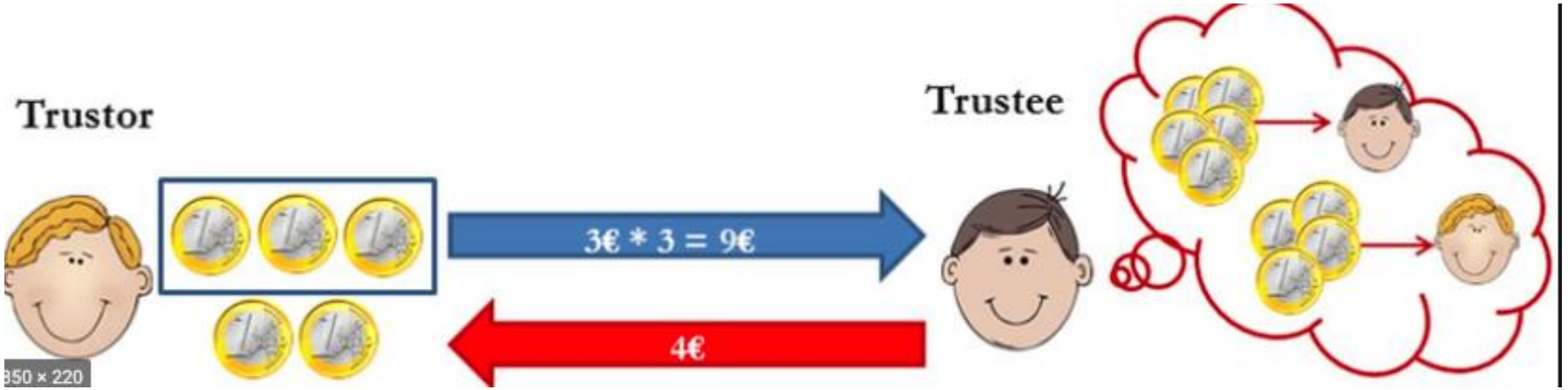
Rob Felt, Georgia Tech)

SUBJECTIVE AND BEHAVIORAL MEASURES OF TRUST

How do you measure trust?

1. Behaviour
2. Subjective measures of trust

1-MEASURING TRUST WITH THE “TRUST” GAME



SUBJECTIVE AND BEHAVIORAL MEASURES OF TRUST

How do you measure trust?

1. Behaviour
2. Subjective measures of trust

THE MDMT MEASURE

MDMT
A MULTI-DIMENSIONAL CONCEPTION AND MEASURE OF HUMAN-ROBOT TRUST

Performance trust		Moral trust	
Reliable	**Capable**	**Sincere**	**Ethical**
Reliable	Capable	Sincere	Ethical
Predictable	Skilled	Genuine	Respectable
Can count on	Competent	Candid	Principled
Consistent	Meticulous	Authentic	Has integrity

Analysis of different semantic categories that constitute different dimensions of trust.

Malle, B. F., & Ullman, D. (2021). A multidimensional conception and measure of human-robot trust. In Trust in Human-Robot Interaction (pp. 3-25). Academic Press.

3- MDMT
A MULTI-DIMENSIONAL CONCEPTION AND MEASURE OF HUMAN-ROBOT TRUST

Malle, B. F., & Ullman, D. (2021). A multidimensional conception and measure of human-robot trust. In Trust in Human-Robot Interaction (pp. 3-25). Academic Press.

TÉCNICO LISBOA

MEASURING TRUST-**TRUST PERCEPTION SCALE**-HRI

The **Trust Perception Scale-HRI** is a psychometrically-developed 40 item instrument intended to measure human trust in robots.

- Items are based on data collected identifying robot features from pictures and their perceived functional characteristics.
- Looks at 4 components corresponding to **capability, behavior, task, and appearance.**
  - Capability and behavior correspond to two of the dimensions commonly found in interpersonal trust and trust in automation
  - Appearance may have a special significance for trust in robots.

The instrument was validated in same-trait and multi-trait analyses producing changes in rated trust associated with manipulation of robot reliability. The scale was developed based on 580 responses and 21 validation participants.

Kristin E Schaefer, Jessie YC Chen, James L Szalma, and PA Hancock. A meta-analysis of factors influencing the development of trust in automation implications for understand- ing autonomy in future systems. Human Factors: The Journal of the Human Factors and Ergonomics Society, page 0018720816634228, 2016

TRUST PERCEPTION SCALE

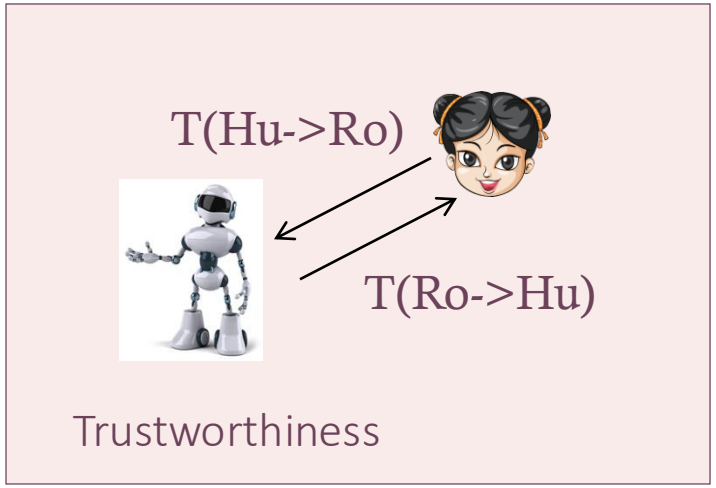
Table 10.8 Finalized Trust Perception Scale-HRI

	0 %	10 %	20 %	30 %	40 %	50 %	60 %	70 %	80 %	90 %	100 %
What % of the time will this robot be . . .											
1. Considered part of the team	O	O	O	O	O	O	O	O	O	O	O
2. Responsible	O	O	O	O	O	O	O	O	O	O	O
3. Supportive	O	O	O	O	O	O	O	O	O	O	O
4. Incompetent[a]	O	O	O	O	O	O	O	O	O	O	O
5. Dependable[b]	O	O	O	O	O	O	O	O	O	O	O
6. Friendly	O	O	O	O	O	O	O	O	O	O	O
7. Reliable[b]	O	O	O	O	O	O	O	O	O	O	O
8. Pleasant	O	O	O	O	O	O	O	O	O	O	O
9. Unresponsive[a,b]	O	O	O	O	O	O	O	O	O	O	O
10. Autonomous	O	O	O	O	O	O	O	O	O	O	O
11. Predictable[b]	O	O	O	O	O	O	O	O	O	O	O
12. Conscious	O	O	O	O	O	O	O	O	O	O	O
13. Lifelike	O	O	O	O	O	O	O	O	O	O	O
14. A good teammate	O	O	O	O	O	O	O	O	O	O	O
15. Led astray by unexpected changes in the environment	O	O	O	O	O	O	O	O	O	O	O

TÉCNICO LISBOA

Schaefer, K. E. (2016). Measuring trust in human robot interactions: Development of the "trust perception scale - HRI". In Robust Intelligence and Trust in Autonomous Systems (pp. 191-218). Springer, Boston, MA.

SUMMARY: "TRUST" IN THE FIELD OF HRI



- T(Hu-> Ro): the trust that the human deposits in the robot (in many of its characteristics and dimensions)

- Trustworthiness factors…

1) Depends on many factors (Human-Related; Robot-Related and Environmental Related)
2) There are different ways to measure trust (behavioural and subjective measures)
3) People do overtrust robots

PAPERS TO READ

- Salem, M., Lakatos, G., Amirabdollahian, F., & Dautenhahn, K. (2015, March). Would you trust a (faulty) robot?: Effects of error, task type and personality on human-robot cooperation and trust. In Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (pp. 141-148). ACM.
- Lewis, M., Sycara, K., & Walker, P. (2018). The role of trust in human-robot interaction. In *Foundations of trusted autonomy* (pp. 135-159). Springer, Cham.
- *E. Rosemarie, Yagoda and Douglas J Gillan. You want me to trust a robot? The development of a human-robot interaction trust scale. International Journal of Social Robotics 4(3), 235–248 (2012)*
- *Kristin E Schaefer, Jessie YC Chen, James L Szalma, and PA Hancock. A meta-analysis of factors influencing the development of trust in automation implications for understand- ing autonomy in future systems. Human Factors: The Journal of the Human Factors and Ergonomics Society, page 0018720816634228, 2016*
- *Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. Human factors, 46(1), 50-80.*
- *Robinette, P., Li, W., Allen, R., Howard, A. M., & Wagner, A. R. (2016, March). Overtrust of robots in emergency evacuation scenarios. In The Eleventh ACM/IEEE International Conference on Human Robot Interaction (pp. 101-108). IEEE Press.*
- Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human factors*, *53*(5), 517-527.
- Hancock, P. A., Billings, D. R., & Schaefer, K. E. (2011). Can you trust your robot?. *Ergonomics in Design*, *19*(3), 24-29.