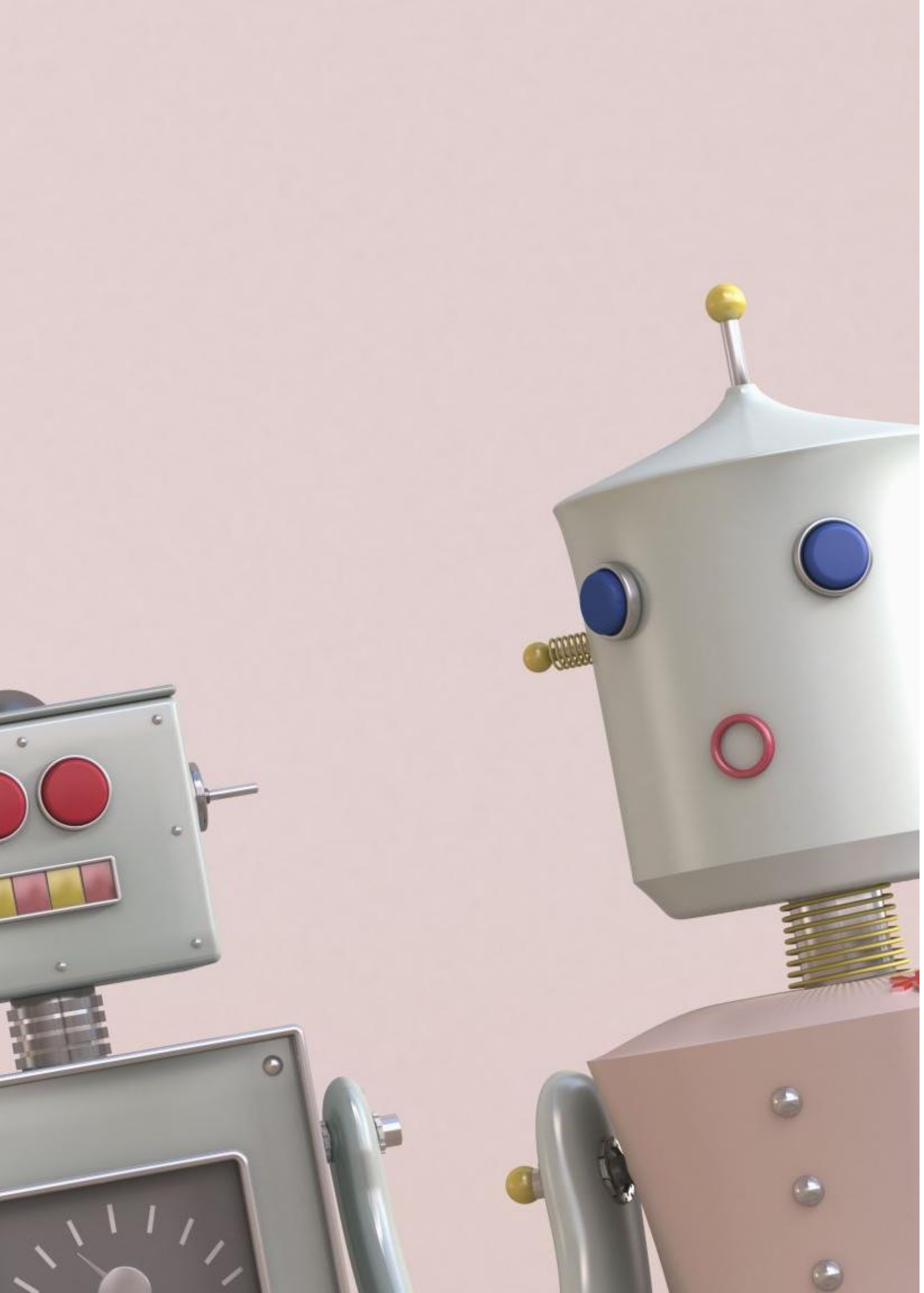


SOCIAL ROBOTS AND HUMAN-ROBOT INTERACTION

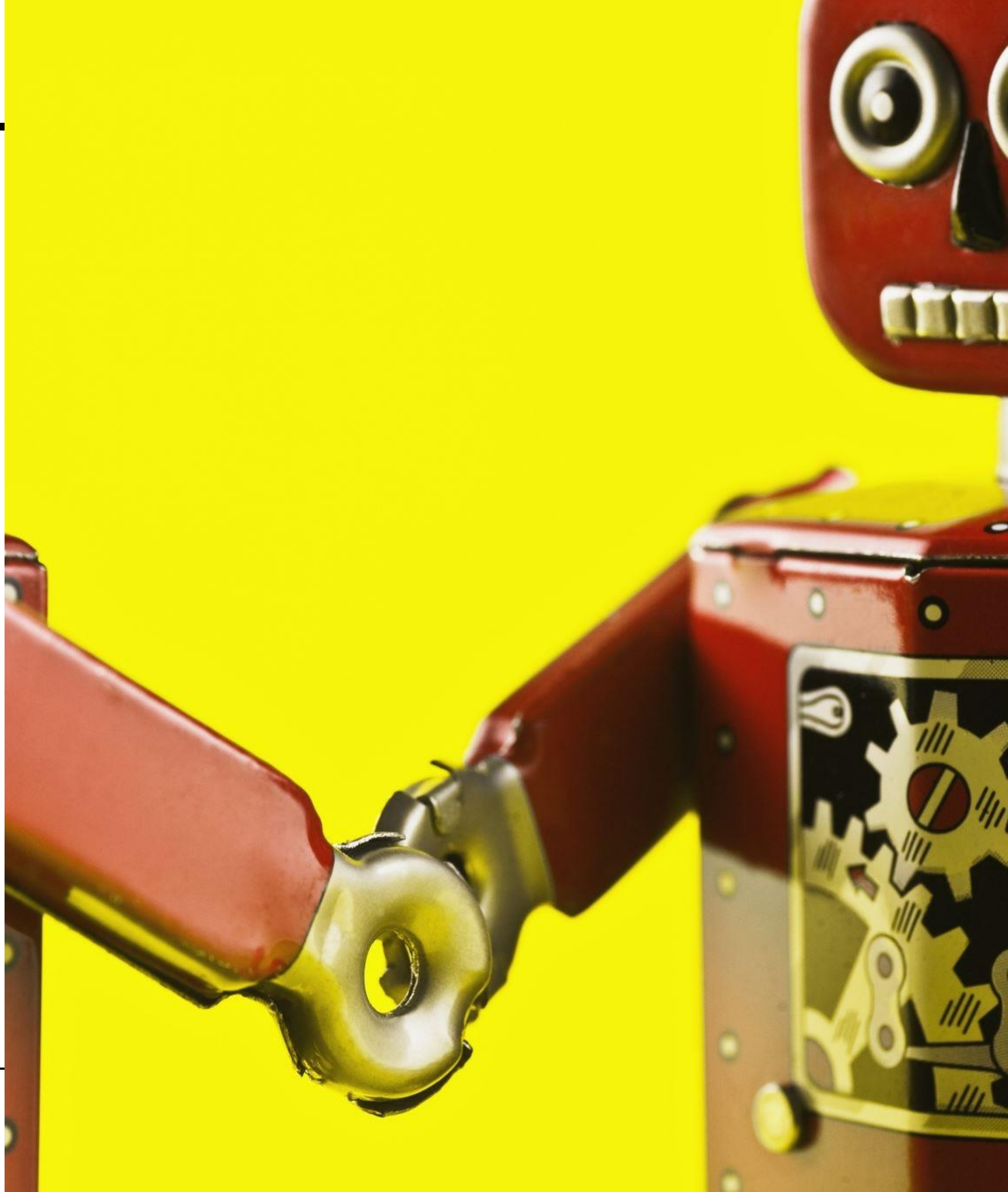
Day 3, Designing Experiments in HRI

Ana Paiva

P2
2025/2026

OVERVIEW OF LECTURE

- Experimental design
  - Research questions
  - Hypotheses
  - Variables
  - Type of experiments
- Typical measures in HRI
- Situating the experiment in the scientific world
- Running an experiment



GENERAL ASPECTS ON DESIGNING EXPERIMENTS FOR HRI

Three **types of studies** we may conduct:

1. Study to collect data to use in the construction of a social robot;
2. Investigate the responses people give to social robots (even if the capabilities and features of such social robots do not exist yet);
3. Show that a technique built for a social robot has the desired effect in users.

Experimental Design: finding the research question

Motivation And Interest

Facial features

Bibliography on the topic

Research on Facial features & Robots

Other similar research: faces in robots

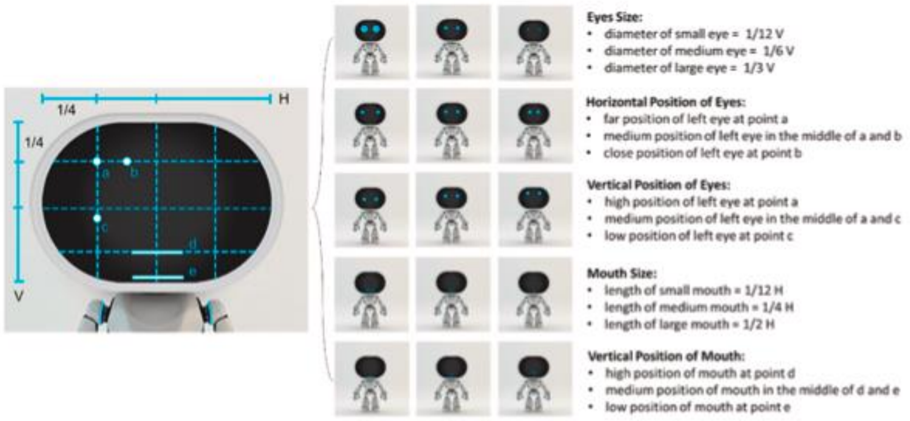
No research was done investigating "facial features" impact on trust in social robots

Research question

Example: Designing an experiment: research question

Example of "generic" problems/questions:

- Do specific features in robotic's faces increase/decrease the trust in that robot?



Design Study

Song, Y., Luximon, A., & Luximon, Y. (2021). The effect of facial features on facial anthropomorphic trustworthiness in social robots. Applied Ergonomics, 94, 103420.

Experimental Design: finding the research question

Motivation & interest → Exploring faults/errors in robots

Bibliography on the topic → Research on the effect of faults in robots

Other similar research: both algorithms and studies → No research was done investigating how faults impact likeability of robots

Research question: Do errors performed by a robot during its interaction with humans influence how likeable that robot is?

Experimental Design: finding the research question

Motivation & interest

A new algorithm for task execution **together** with a human

Bibliography on the topic

Research on **collaboration** with humans

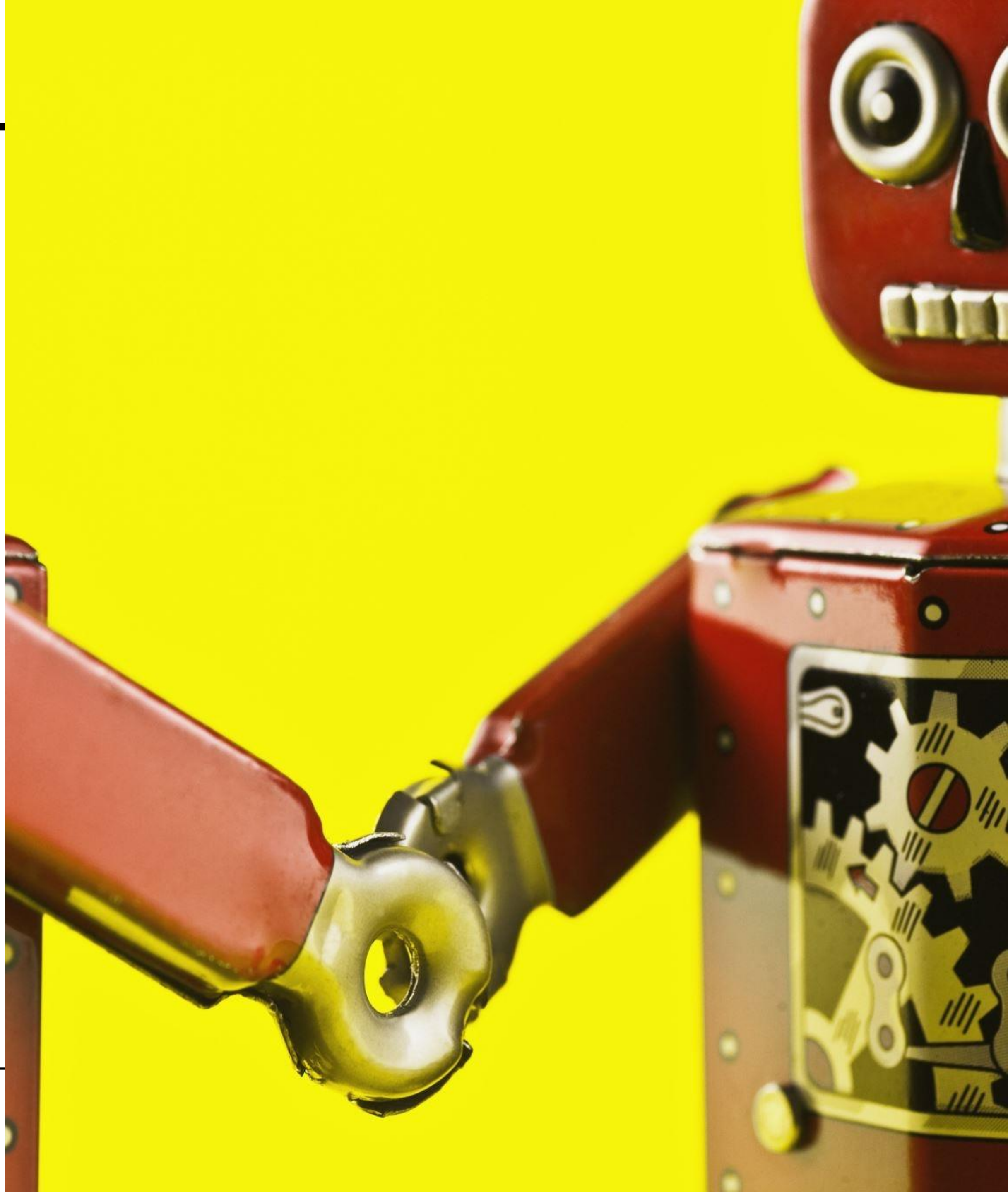
Other similar research: both algorithms and studies

No research was done investigating how the new algorithm for collaborative task execution impacts collaboration

Research question: does algorithm X for task execution with a robot impacts collaboration?

OVERVIEW OF LECTURE

- Experimental design
  - Research questions
  - Hypotheses
  - Variables
  - Type of experiments
- Typical measures in HRI
- Situating the experiment in the scientific world
- Running an experiment



HYPOTHESES

What is a hypothesis?

Makes a prediction about the behavior of a measurable outcome of an experiment (related to a research question)

- **Can't be an educated guess**
- Should be based on literature and theory
- Can allow to understand what is the best **design** for a social robot

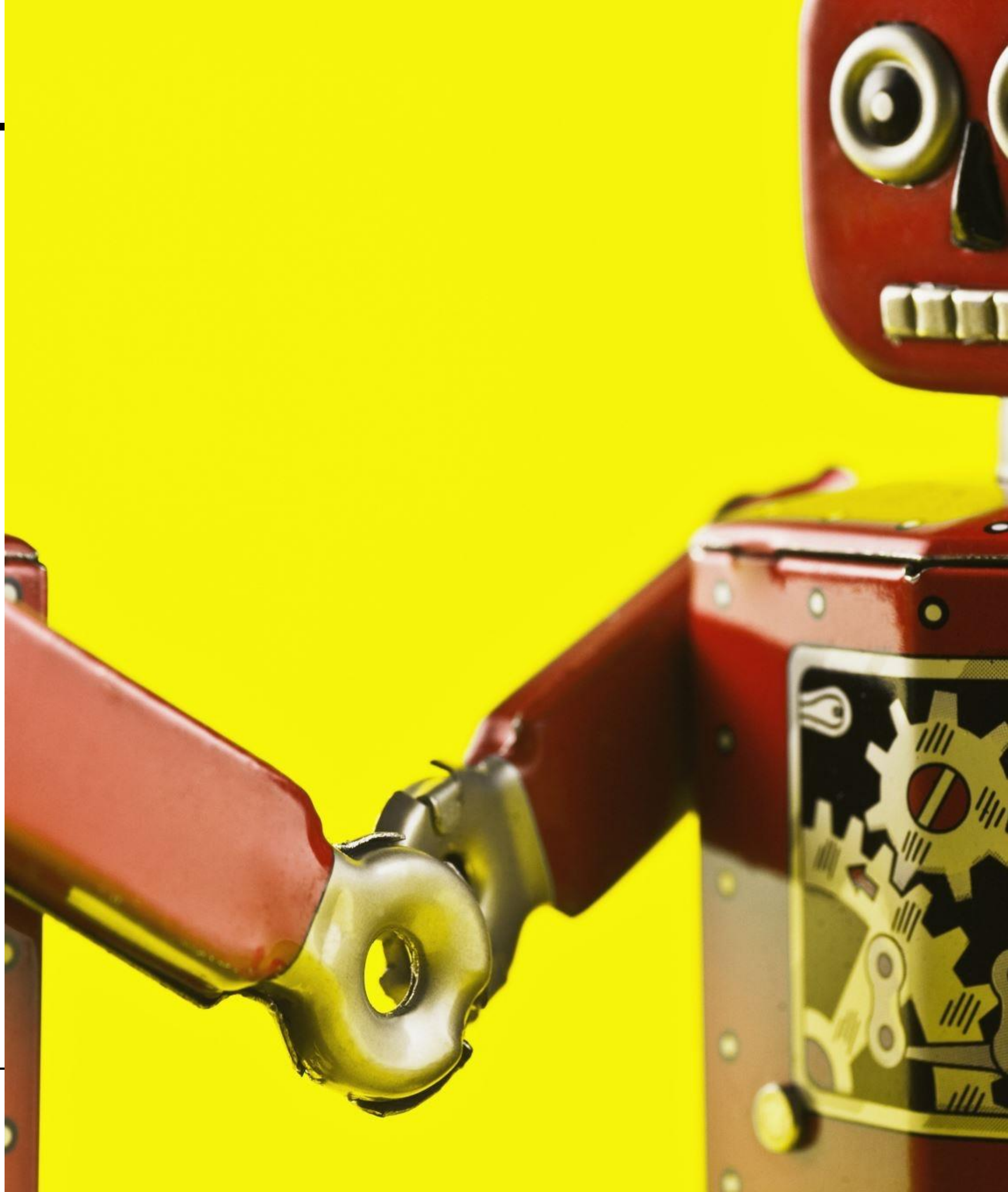
H1. Social robots with large (vs. small) eyes are perceived as more trustworthy.

H2. A robot that commits errors during its interaction with humans, is perceived as more likeable than a robot that performs flawlessly.

H3. The animation of a robot (through algorithm X) will positively affect the participants' perception of the collaboration.

OVERVIEW OF LECTURE

- Experimental design
  - Research questions
  - Hypotheses
  - Variables
  - Type of experiments
- Typical measures in HRI
- Situating the experiment in the scientific world
- Running an experiment



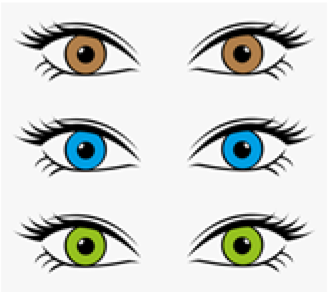
Designing an experiment : Variables

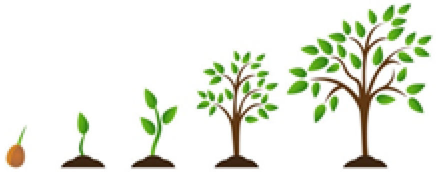
To draw conclusions and answer our research questions we need to have **variables** which should be **measured** in the study

- Numeric variables:
  - Discrete / count (e.g. number of energy bars picked up by participants)
  - Continuous (e.g. distance kept from a robot)
- Categorical variables:
  - Logistic / binary (e.g. did people reveal a secret to the robot or not?)
  - Ordinal (e.g. Likert scale)
  - Nominal (e.g. type of robot people interacted with)

TYPES OF VARIABLES: WHAT TO MEASURE?

A variable is any characteristics, number, or quantity that can be measured or counted in an experiment





IN EXPERIMENTAL DESIGN: TWO TYPES OF VARIABLES

INDEPENDENT VARIABLES

The manipulation

The factor

The predictor

The ***Independent variable*** is what we intentionally change

DEPENDENT VARIABLES

The effect

The response variable

The outcome

The ***Dependent variable*** changes in response to the independent variable

Dependent versus Independent Variables

So, in a study we have the Independent variables and Dependent variables:

The Independent variable (the "causes") is the variable that is purposely "changed" by us. It is the manipulated variable.

The *Dependent variable* (the "effects") changes in response to the independent variable. It is the responding variable.

HYPOTHESES AND VARIABLES

H1. A robot that commits errors during its interaction with humans, is perceived as more likeable than a robot that performs flawlessly.

Independent variable

Robot commits errors (or not/ number)

Dependent variable

Measure: degree of likeability

How to manipulate the Independent Variable?

It's not easy….

Examples:

1) People trust more a robot that addresses the user by his or her name ;

2) Children learning more in a simple math task together with a robot, when it has a social learning mechanism to make the robot adapt to the child.

3) Humanoid robots with a face are more friendly

Independent Variable

1) Calling the user by the name

2) Social learning mechanism for personalisation and adaptation

3) Have a face or not (in a humanoid robot)

Dependent Variable

1) Trust the robot

2) Learn more

3) Perception of friendliness

Choice of the Dependent Variable

Depends on what our research question (and hypothesis) is.

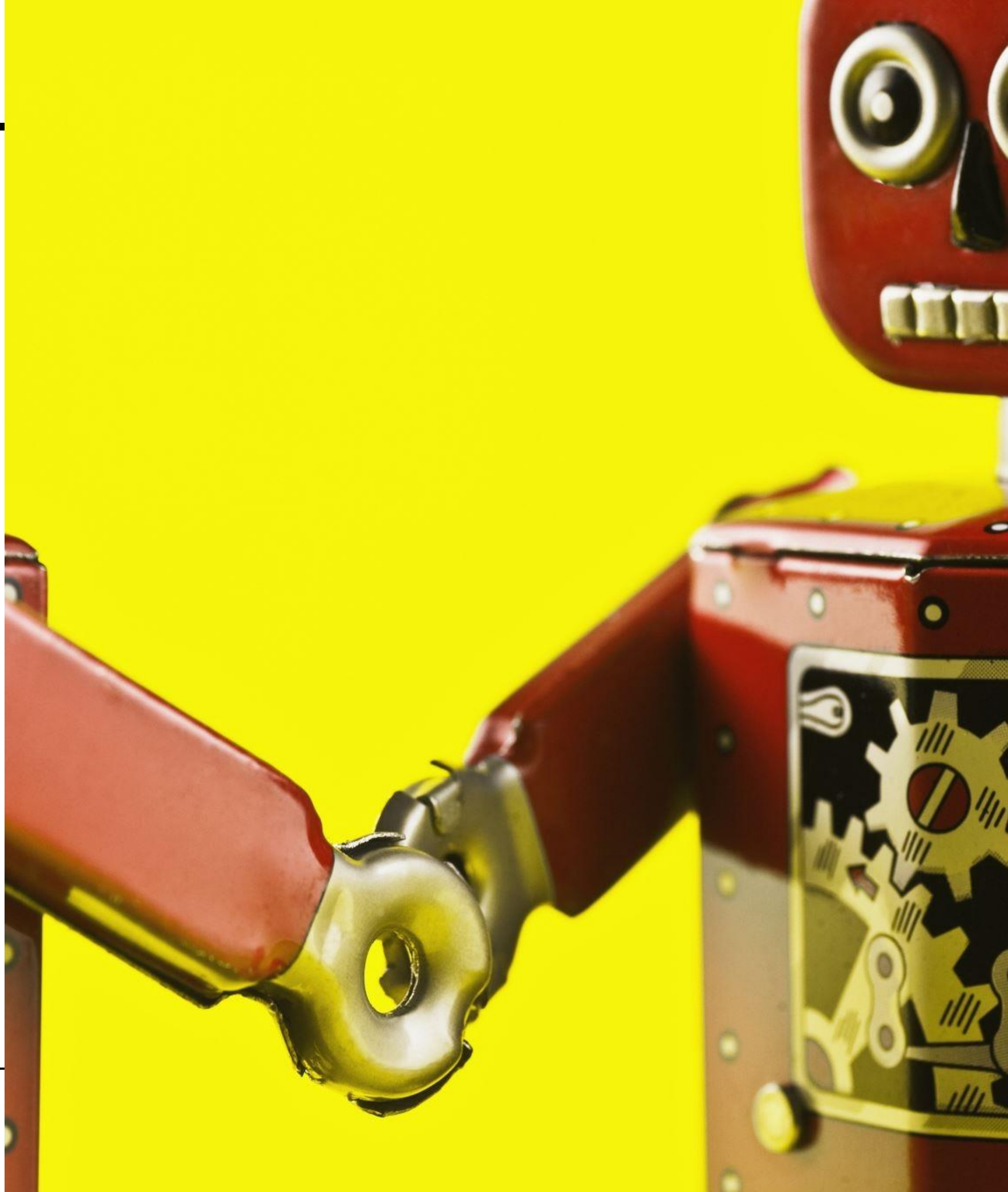
What do we want to measure?

- “trust” in a robot
- speed of task execution (human and robot)
- emotional reaction towards a robot
- social proximity with the robot
- likeability of the robot

Often to measure these “attitudes”, we need to rely on variables that we can measure (through questionnaires, behaviour analysis, physiological measures, etc).

OVERVIEW OF LECTURE

- Experimental design
  - Research questions
  - Hypotheses
  - Variables
  - Type of experiments
- Typical measures in HRI
- Situating the experiment in the scientific world
- Running an experiment



FROM RESEARCH QUESTION TO EXPERIMENT

We need to compare the manipulation to something

- **Treatment vs. control**
  - E.g. a robot that makes errors vs. a robot that doesn't make errors

(Condition 1 vs. condition 2)

"I couldn't afford a control group so I decided to go with an out-of-control group."

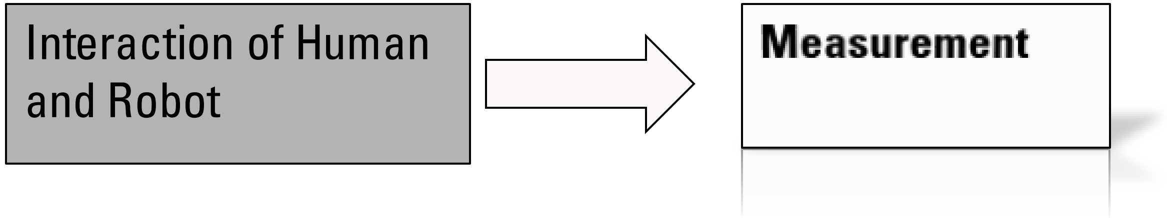
Experimental designs

The typical Com… Scie… xperimental

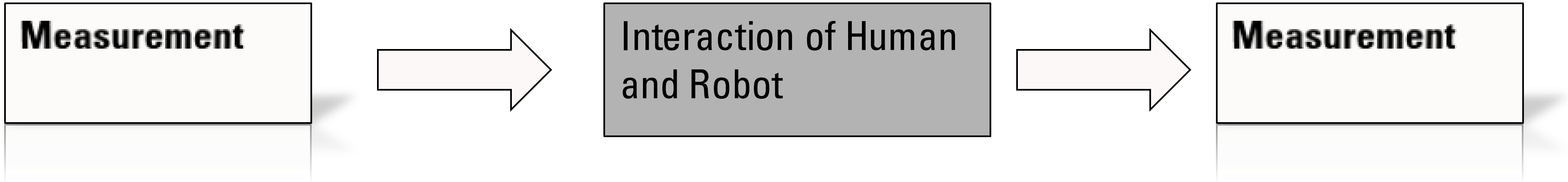
Ask a few friends to interact with a Robot → Ask a few friends if they liked it

Quasi-experimental designs

One group post-test design (the typical CS student design)



One group pre-test/post-test design



Problems of Quasi-experimental designs

- It does not control the variables (there is only one variable)
- It does not give relevance to the effect of what was done (except as a general impact of the system)
- Used for things like: does the robot lead to learning?
  - Measure “learning” pre- and post- interaction and see if there is a positive result.
    - However, we do not know if there wasn’t something else contributing to learning

So… how to do it?

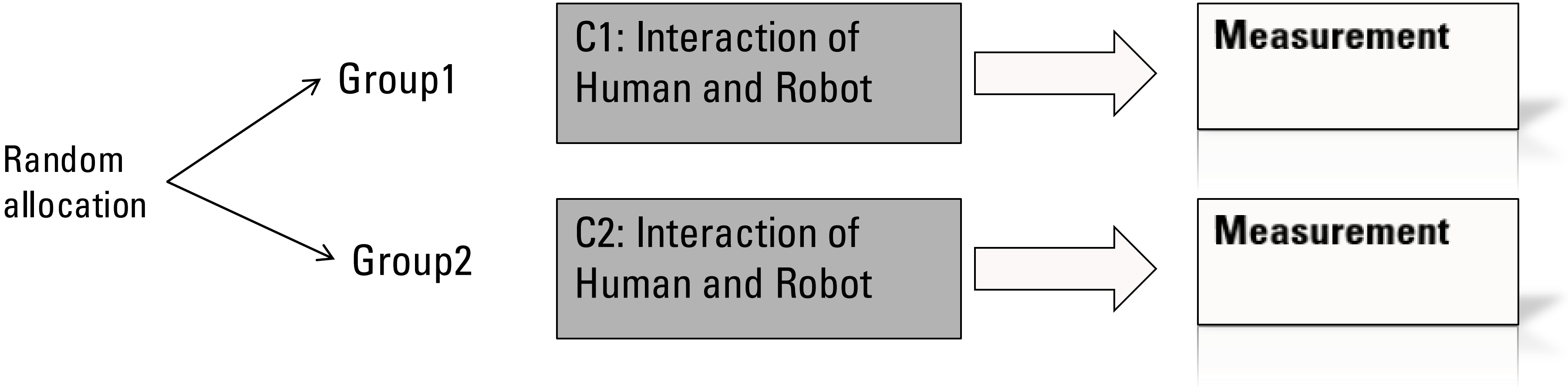
Test ”conditions (treatments)” separately and measure the effects….

Between-Groups Experimental Designs

Between groups (or independent measures)

- Use separate groups of users (participants) for each of the different conditions (associated with the manipulation done).
- Each participant is tested only once.

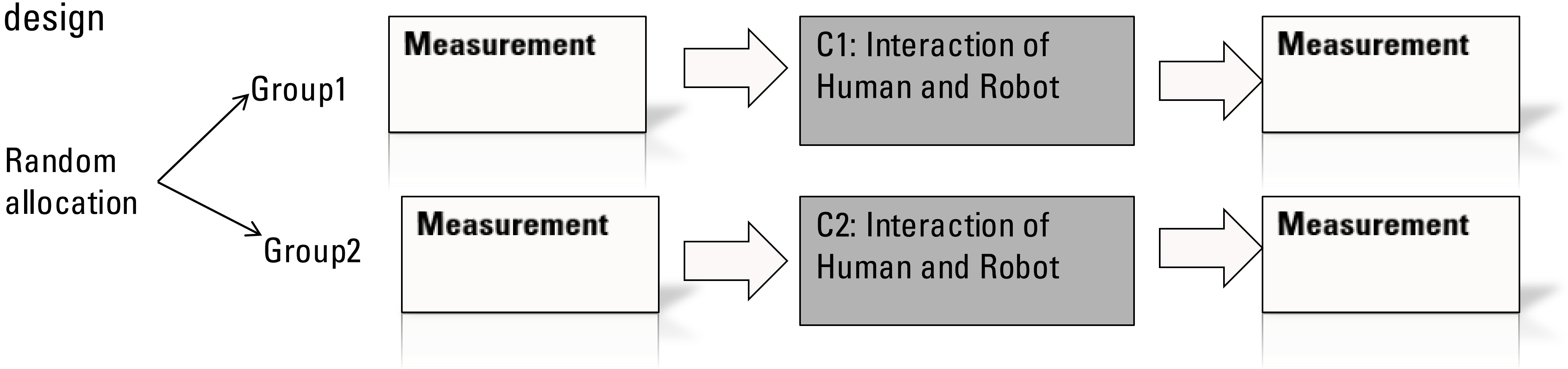
A post-test only/control group design



Between groups (or independent measures)
It is essential that we allocate participants "randomly" to our experimental conditions

Between-Groups experimental design

A pre-test post-test control group design



Between groups (or independent measures)
It is essential that we allocate participants “randomly” to our experimental conditions

BETWEEN-GROUPS EXPERIMENTAL DESIGN

Advantages

- Simplicity
- Less chance of practice and fatigue effects
- Useful when it is impossible for a user to participate in all experimental conditions

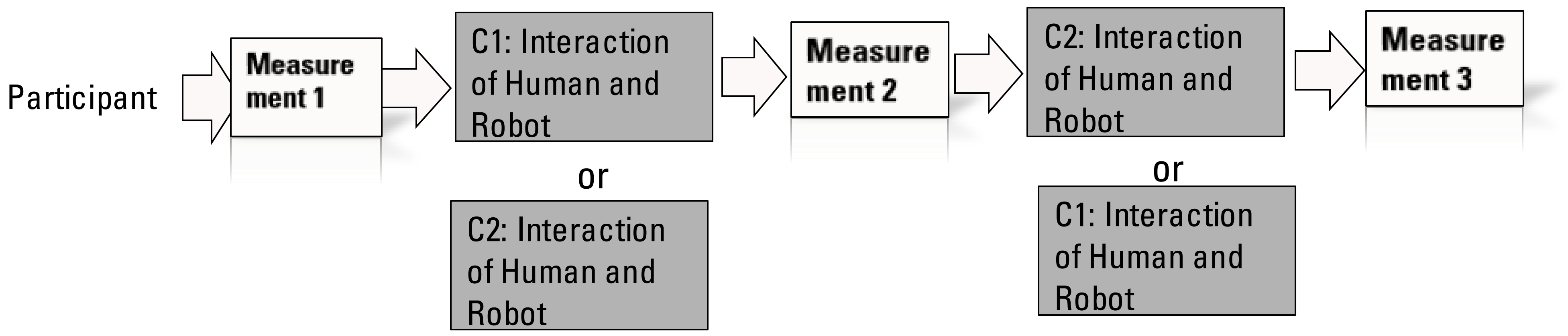
Disadvantages

- Expense in time, effort and participants
- Insensitivity to experimental manipulations

Within Subjects Experimental Designs

Within-subjects (or repeated measures)

- Participants are exposed to the different conditions of the experiment.



Randomization/counterbalancing

Within-subjects (or repeated measures)

- Participants can all experience our experimental conditions in the same order (if we are interested in a learning effect or similar)
- Half of the participants experience one condition first, the other half experience the other condition first (counterbalancing or cross over design)
- The more conditions you have, the more cells you need to fit in your counterbalancing, the more participants you need…

WITHIN SUBJECTS EXPERIMENTAL DESIGN (REPEATED MEASURES)

In a within-subjects, all the users pass through the all the interactions.

Advantages

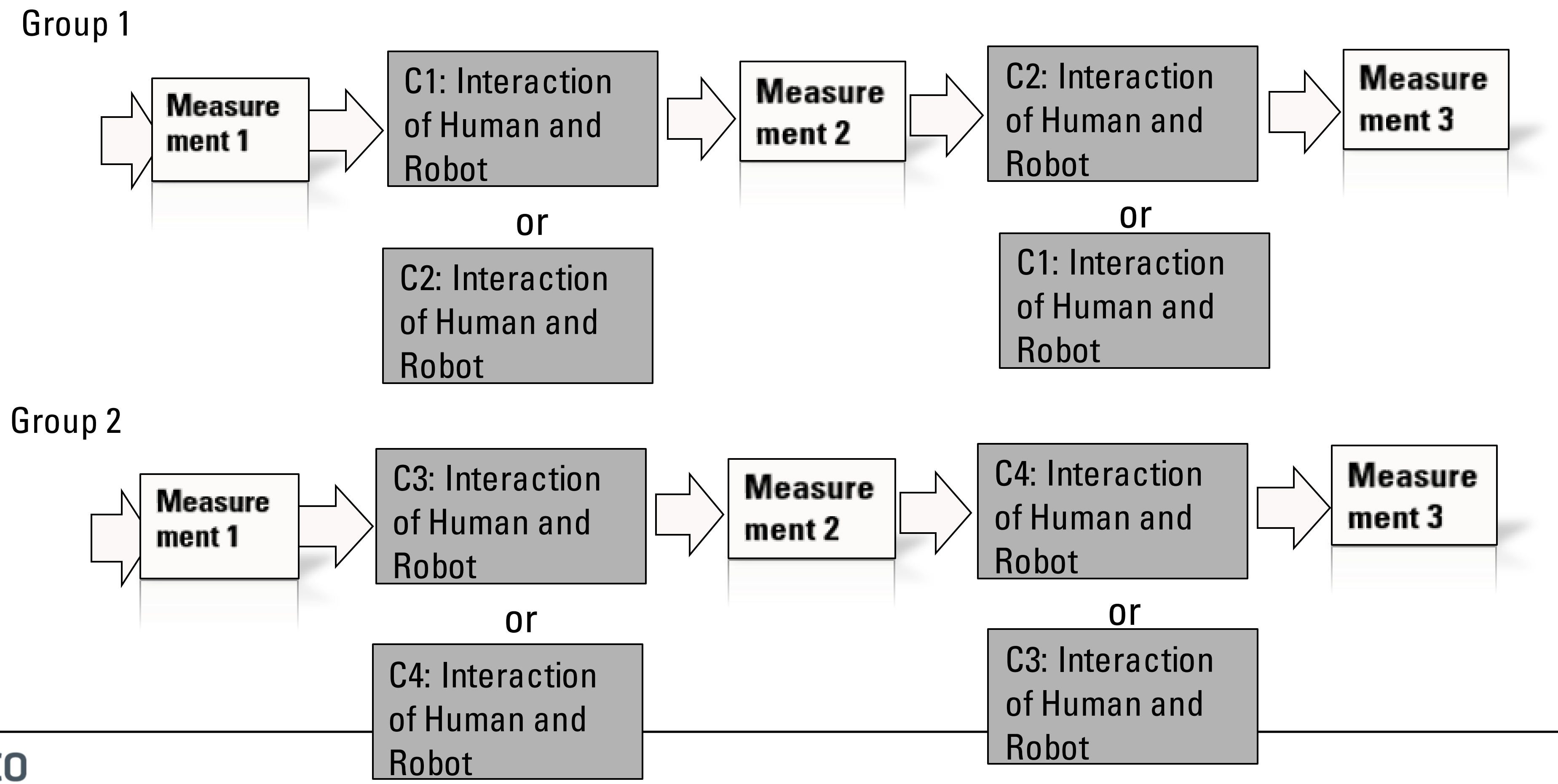
- economical: in terms of time, effort as we are using the same users
- sensitivity: sometimes the differences are hard to perceive and there is random noise

Disadvantages

- Carry over effects from one condition to another
- Need for the conditions to be reversible (due to the randomization of the order of the conditions)

Mixed methods Experimental Designs

Mixed methods: Combines both between subjects (in one variable) and repeated measures (in other)



Summary Experimental Designs

Between groups (or independent measures)

- Use separate groups of users (participants) for each of the different conditions (associated with the manipulation done).
- Each participant is tested only once.

Within-subjects (or repeated measures)

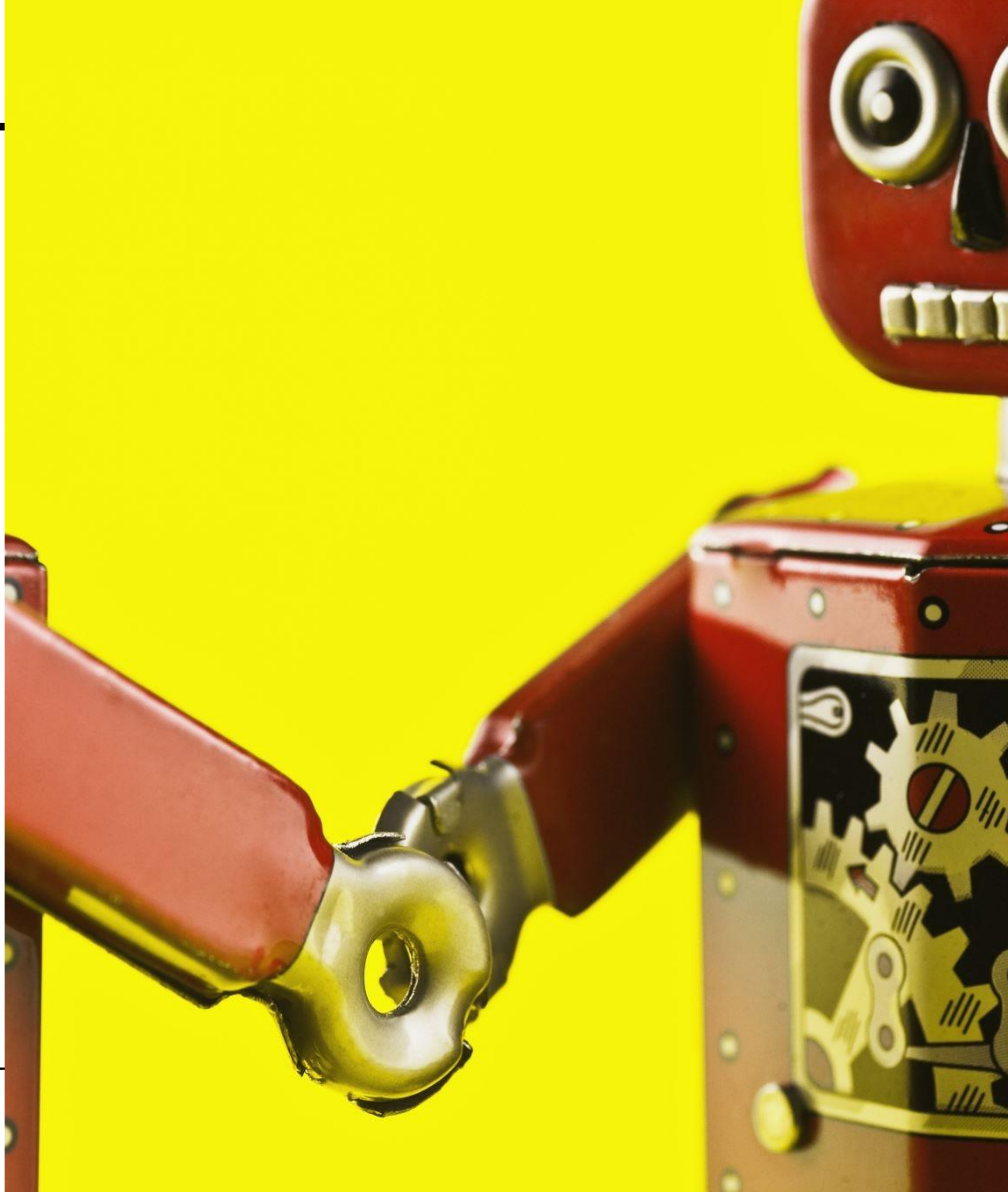
- Participants are exposed to the different conditions of the experiment.

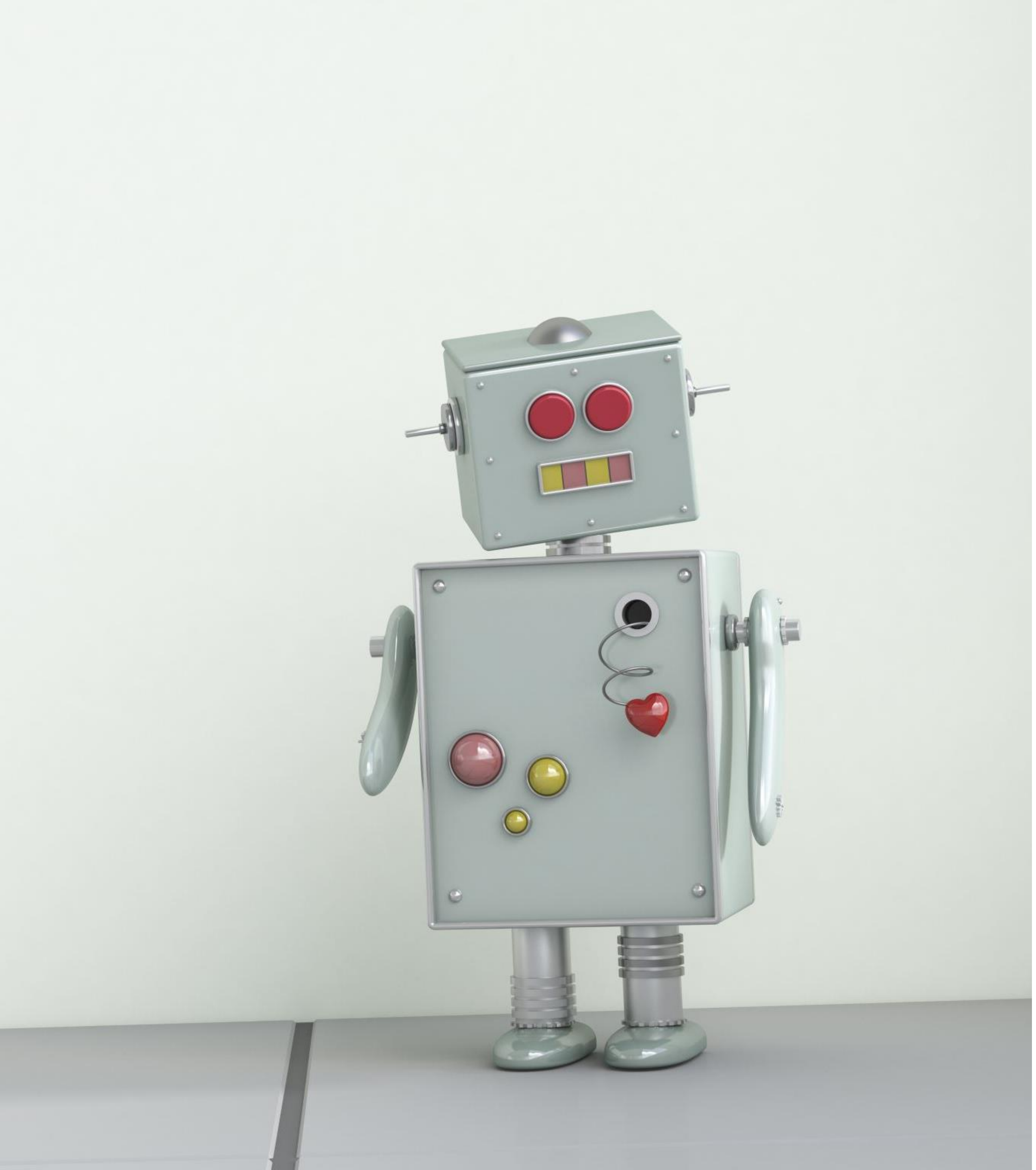
Mixed methods

- Designs that involve a combination of a between groups and within subjects variables.

OVERVIEW OF LECTURE

- Experimental design
  - Research questions
  - Hypotheses
  - Variables
  - Type of experiments
- Typical measures in HRI
- Situating the experiment in the scientific world
- Running an experiment





DIFFERENT TYPES OF MEASURES

- Task Measures
  - Robot task metrics
  - Human task metrics
- Physiological Measures of participants
- Behavioural Measures
  - Human responses to the robot (usually obtained by observing the human), like gaze, gestures, etc.
- Subjective Measures (self reported)
  - Perception of the robot or task by the human, and measures by the use of concrete questionnaires

EXPLICIT VS. IMPLICIT MEASURES

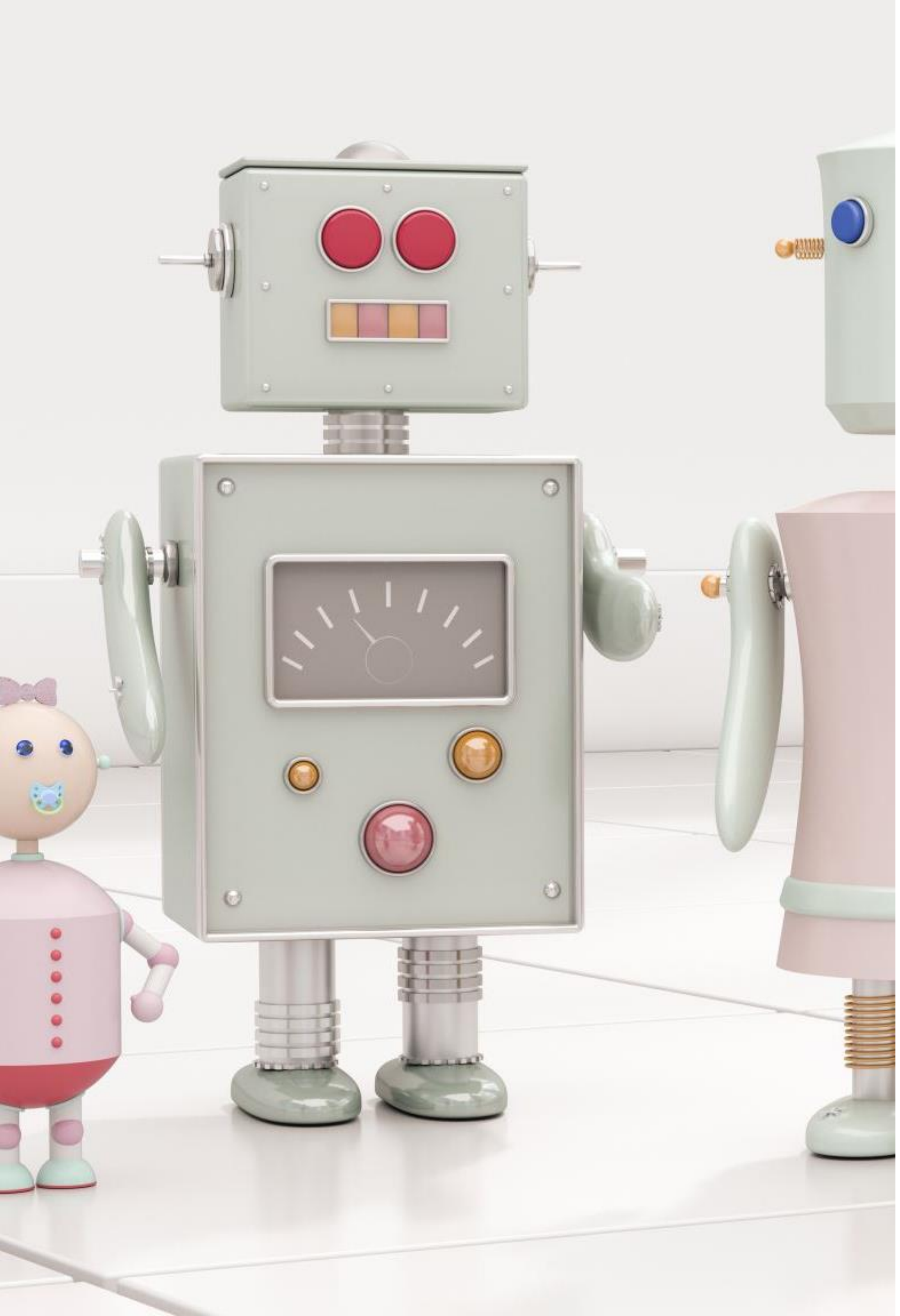
- Explicit measures relate to conscious impressions that people typically have time to reflect on
  - E.g. questionnaire: “From 1 to 7, how safe did you feel…?”
  - E.g. interview: “Why did you feel safe…?”
  - E.g. informal brainstorming: “Would you or your grandmother use this feature if we added it to the robot?”
- Implicit measures refer to unconscious attitudes
  - E.g. reaction times: How quickly does someone click on the word “safe” after seeing a picture of a robot? (IAT)
  - E.g. behaviour: Are people smiling / looking uncomfortable?
  - E.g. engagement: Are people using a product’s feature at all? Are people willing to interact with the system?

Hofmann, W., et al. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. Personality and Social Psychology Bulletin 31, 1369-1385

EXPLICIT VS. IMPLICIT MEASURES

- Keep in mind that unfortunately attitudes (survey responses) don't always correlate with behavior!
- There is a difference between words and actions.
- People say they would 100% trust a system ≠ people actually trusting the system
- So what to do? A mix of explicit and implicit is good and gives you a bigger picture

Greenwald, A. G., McGhee, D. E., & Schwartz, J. L. (1998). Measuring individual differences in implicit cognition: the implicit association test. Journal of personality and social psychology, 74(6), 1464.
Schwarz, N. (1999). Self-reports: How the questions shape the answers. American Psychologist 54(2), 93-105



TASK METRICS

Often we want to measure the **performance of the robot** and **the user** in a task.

- **Performance of the robot** (must be related to what the user also does). Example: navigation, or if the robot performs better when the user gave the robot certain instructions
- **Performance of the user** (for example, how well the user did when interacting with the robot)
- **Performance of the team** (which includes the user and the robot)

TASK METRICS: EXAMPLE-MEASURING EFFICIENCY

Efficiency measures the time needed to complete the task.

Efficiency measures include:

- Time to complete the task
- User time for completing the task
- Average time for obstacle extraction
- Number of user interventions and requests per unit time. Interactions can be planned or unplanned.
- Ratio of user time to robot time. For example, if the User spends 5 minutes to input a navigation plan that allows the robot to successfully navigate for an hour, we have a 1:12 ratio.

TÉCNICO LISBOA

Dragan, A. D., Bauman, S., Forlizzi, J., & Srinivasa, S. S. (2015, March). Effects of Robot Motion on Human-Robot Collaboration. In Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (pp. 51-58). ACM.

Example: Results

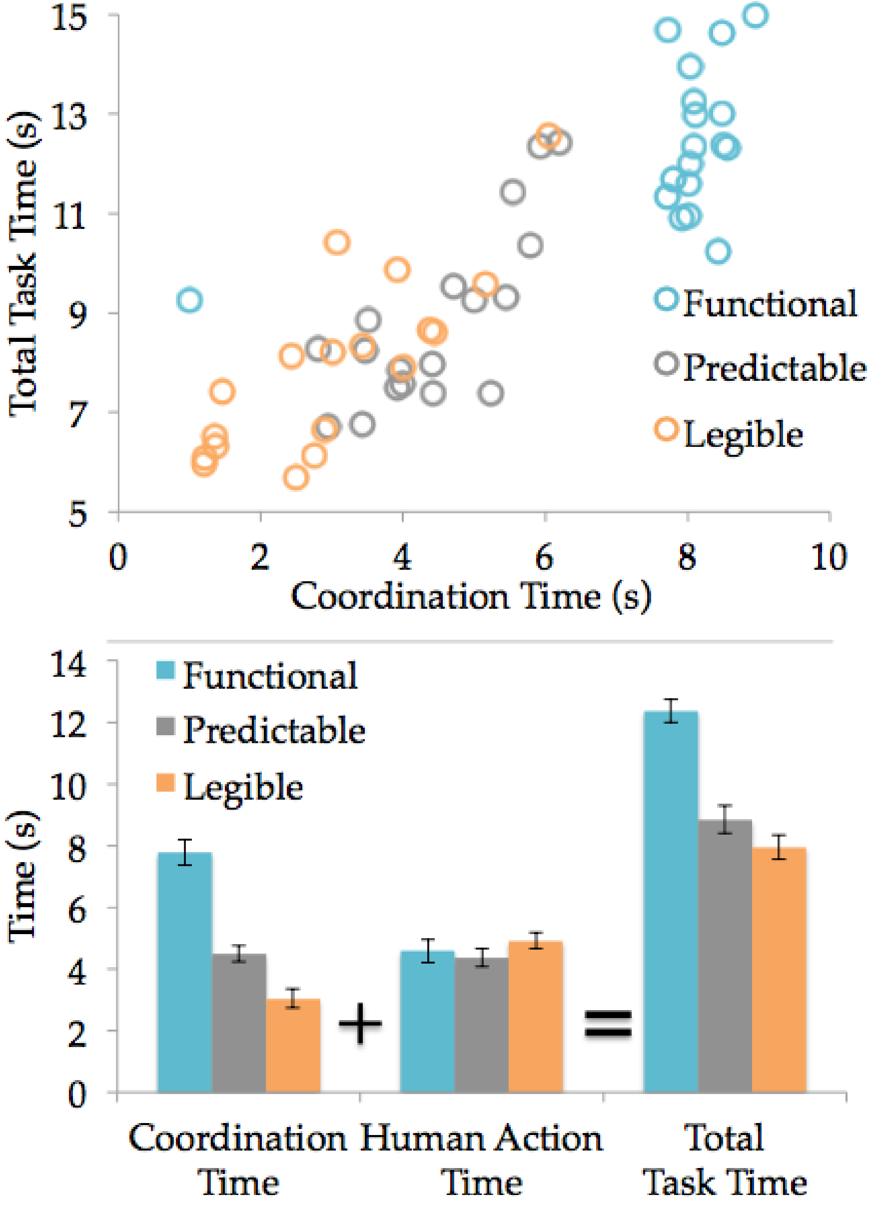


Figure 4: Findings for objective measures.

Dragan, A. D., Bauman, S., Forlizzi, J., & Srinivasa, S. S. (2015, March). Effects of Robot Motion on Human-Robot Collaboration. In Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction (pp. 51-58). ACM.

PHYSIOLOGICAL MEASURES

- Use of heart-rate and blood pressure
- Galvanic Skin response
- Hormone levels
- fMRI

  - Difficulty:
    - Often the need to have expensive equipment
    - The signal processing and analysis is sometimes non trivial
    - The changes detected in the signals may not be a result of the experimental manipulation
    - Difficult to draw conclusions (in terms of HRI).

OTHER BEHAVIORAL MEASURES

- Include the analysis of behavior of users (aspects such as gaze, actions, posture, etc)
- Often relies on the annotation of videos (or use of tools to extract behaviours).

EXAMPLE: GAZE BEHAVIOUR OF USERS INTERACTING WITH ROBOVIE

Fig. 6. Robovie R-1, the robotic platform we used in the implementation and evaluation of the gaze mechanisms studied in this work.

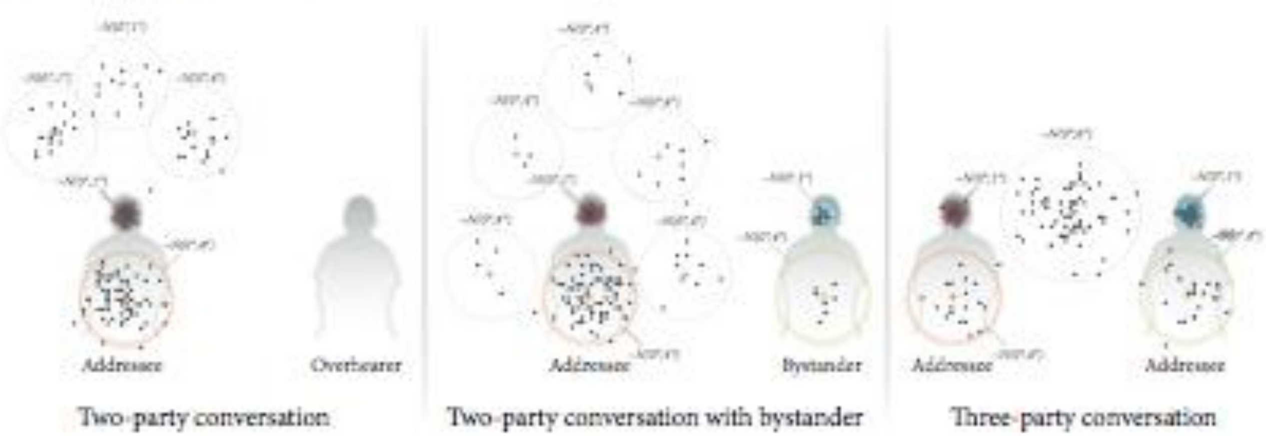
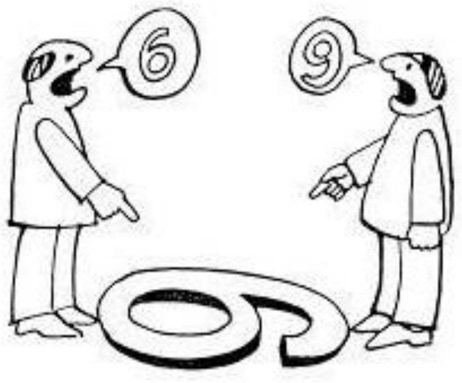


Fig. 7. The gaze targets that the robot produced based on our implementation of the temporal and spatial gaze parameters.

SELF–REPORTED/EXPLICIT MEASURES



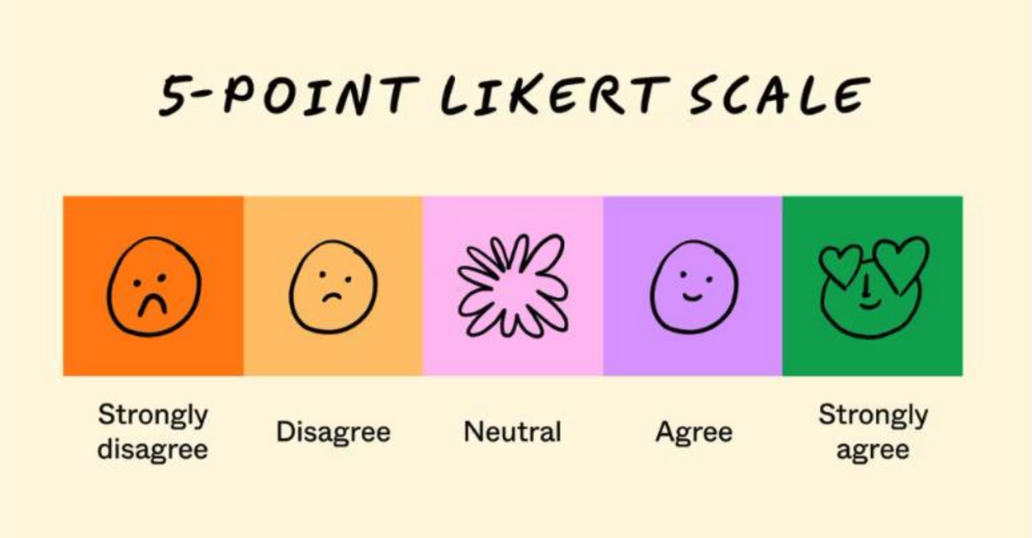
They are subjective!

- Types of Scales:
  - Yes/No scale
  - Likert scales

Do you consider the robot friendly? Yes/No

WHAT IS A LIKERT SCALE?

- It is a psychometric scale commonly involved in research that employs questionnaires.
- It consists of a statement to which the user can express different degrees of agreement.
- It is the most widely used approach to scaling responses in survey research.
- Likert scales are a non-comparative scaling technique and are one-dimensional in nature.



EXAMPLES

Please rate your impression of the robot on these scales:

Machinelike 1 2 3 4 5 6 7 Humanlike
Unpleasant 1 2 3 4 5 6 7 Pleasant

Please the extent to which you agree with the following sentences:

I think this robot looks:

	Strongly Disagree					Strongly agree
	1	2	3	4	5	6
Credible						
Sincere						

If we invent the questions, how do we know we are measuring the same as other people?

HOWEVER...

USEFUL QUESTIONNAIRES

- Need to find validated questionnaires (but many are not)
- Examples of questionnaires typically used:
  - ALMERE
  - Godspeed
  - RoSAS
  - MDMT

EXAMPLE OF QUESTIONNAIRE: ALMERE

ALMERE (Heerink et al.)**:** "[t]he Almere model is a technology acceptance model: it can be used to predict and explain usage of a system by observing the **influences** on the **Intention to Use** this system.

Dimensions (Anxiety; Attitude; Perceived Enjoyment; Perceived sociability; Perceived usefulness; Social influence; Social presence):

ANXIETY

1. If I should use the robot, I would be afraid to make mistakes with it
2. If I should use the robot, I would be afraid to break something
3. I find the robot scary
4. I find the robot intimidating

EXAMPLE OF QUESTIONNAIRE: ALMERE

ATTITUDE

5. I think it's a good idea to use the robot
6. The robot would make my schoolwork more interesting
7. It's good to make use of the robot

PERCEIVED ENJOYMENT

16. I enjoy the robot talking to me
17. I enjoy doing things with the robot
18. I find the robot enjoyable
19. I find the robot fascinating

PERCEIVED SOCIABILITY

24. I consider the robot a nice partner to talk to
25. I find the robot nice to interact with
26. I feel the robot understands me
27. I think the robot is nice

•

PERCEIVED USEFULNESS

28. I think the robot is useful to me
29. It would be convenient for me to work with the robot in school
30. I think the robot can help me with many things in school

SOCIAL INFLUENCE

31. I think the teachers would like me working with the robot
32. I think many people would like me working with the robot

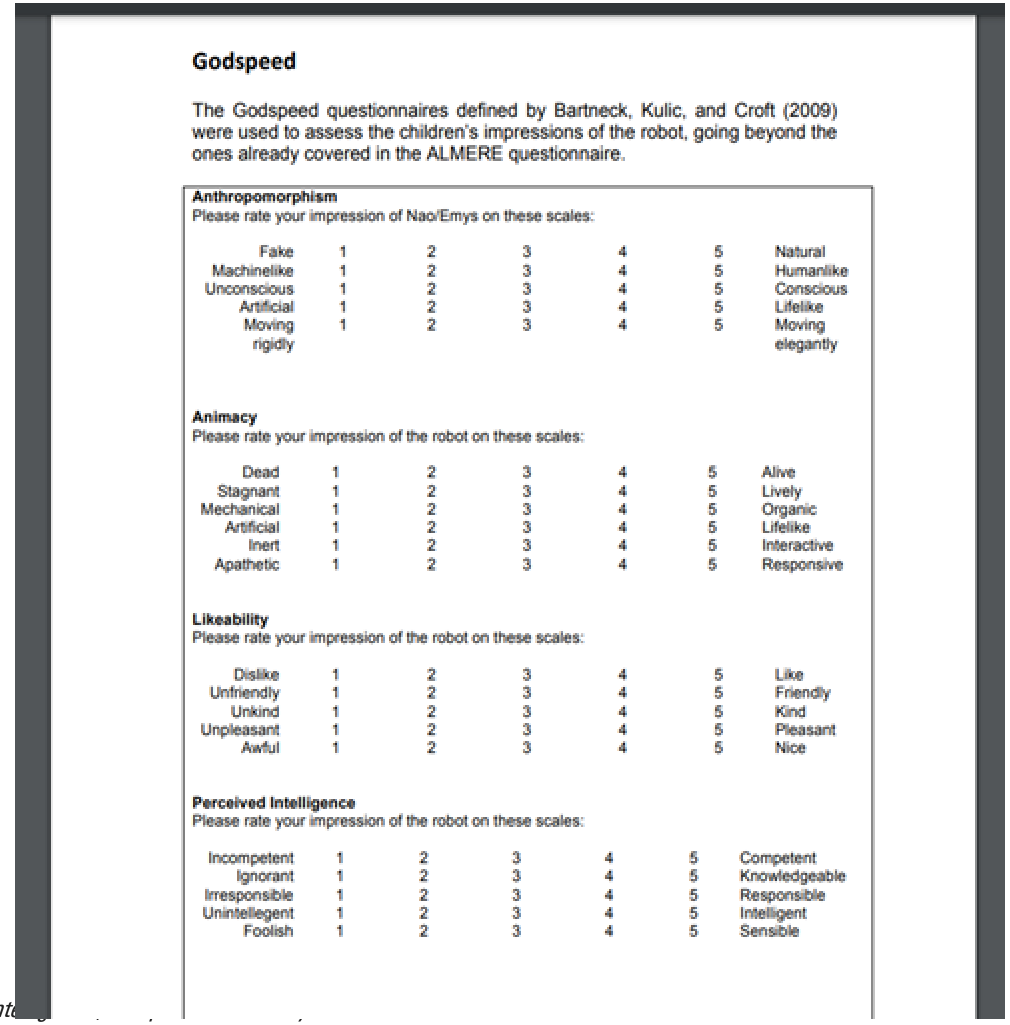
SOCIAL PRESENCE

33. When interacting with the robot I felt like I'm talking to a real person
34. It sometimes felt as if the robot was really looking at me
35. I can imagine the robot to be a living creature
36. I often think the robot is not a real person
37. Sometimes the robot seems to have real feelings

EXAMPLE OF QUESTIONNAIRE: GODSPEED

- A classic in HRI
- Likert scale- based questionnaire used to measure users' overall perceptions of robots
- Made of **subscales**:
  - anthropomorphism
  - animacy
  - likability
  - perceived intelligence
  - perceived safety

http://www.bartneck.de/2008/03/11/the-godspeed-questionnaire-series/



Godspeed

The Godspeed questionnaires defined by Bartneck, Kulic, and Croft (2009) were used to assess the children's impressions of the robot, going beyond the ones already covered in the ALMERE questionnaire.

Anthropomorphism
Please rate your impression of Nao/Emys on these scales:

Fake	1	2	3	4	5	Natural
Machinelike	1	2	3	4	5	Humanlike
Unconscious	1	2	3	4	5	Conscious
Artificial	1	2	3	4	5	Lifelike
Moving rigidly	1	2	3	4	5	Moving elegantly

Animacy
Please rate your impression of the robot on these scales:

Dead	1	2	3	4	5	Alive
Stagnant	1	2	3	4	5	Lively
Mechanical	1	2	3	4	5	Organic
Artificial	1	2	3	4	5	Lifelike
Inert	1	2	3	4	5	Interactive
Apathetic	1	2	3	4	5	Responsive

Likeability
Please rate your impression of the robot on these scales:

Dislike	1	2	3	4	5	Like
Unfriendly	1	2	3	4	5	Friendly
Unkind	1	2	3	4	5	Kind
Unpleasant	1	2	3	4	5	Pleasant
Awful	1	2	3	4	5	Nice

Perceived Intelligence
Please rate your impression of the robot on these scales:

Incompetent	1	2	3	4	5	Competent
Ignorant	1	2	3	4	5	Knowledgeable
Irresponsible	1	2	3	4	5	Responsible
Unintellegent	1	2	3	4	5	Intelligent
Foolish	1	2	3	4	5	Sensible

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived int
robots. International Journal of Social Robotics, 1(1), 71-81.

EXAMPLE OF QUESTIONNAIRE: ROSAS*

- 18-item scale (The Robotic Social Attribute Scale; RoSAS) to
- Measure people’s judgments of the social attributes of robots.
- 3 Factors:
  - Warmth
  - Competence
  - Discomfort

Variable	Factor 1	Factor 2	Factor
Happy	**.831**	-.013	-.009
Feeling	**.811**	-.154	.043
Social	**.793**	.125	-.176
Organic	**.784**	-.149	-.022
Compassionate	**.778**	-.039	.030
Emotional	**.776**	-.204	.050
Capable	-.210	**.706**	-.052
Responsive	-.141	**.680**	.040
Interactive	-.225	**.652**	-.006
Reliable	-.061	**.651**	-.028
Competent	-.111	**.646**	-.040
Knowledgable	-.007	**.620**	-.021
Scary	.052	-.012	**.693**
Strange	-.053	.150	**.601**
Awkward	.049	.037	**.601**
Dangerous	-.035	.024	**.597**
Awful	.360	-.250	**.555**
Aggressive	.265	.009	**.547**
Eigenvalue	22.031	9.336	5.047
Percent variance explained	26.54%	11.25%	6.09%

* Steven J. Stroessner, The Robotic Social Attributes Scale (RoSAS): Development and Validation, 2017

EXAMPLE OF QUESTIONNAIRE: ROSAS*

- 3 Factors:
  - Warmth
  - Competence
  - Discomfort

Variable	Factor 1	Factor 2	Factor 3
Happy	**.831**	-.013	-.009
Feeling	**.811**	-.154	.043
Social	**.793**	.125	-.176
Organic	**.784**	-.149	-.022
Compassionate	**.778**	-.039	.030
Emotional	**.776**	-.204	.050
Capable	-.210	**.706**	-.052
Responsive	-.141	**.680**	.040
Interactive	-.225	**.652**	-.006
Reliable	-.061	**.651**	-.028
Competent	-.111	**.646**	-.040
Knowledgable	-.007	**.620**	-.021
Scary	.052	-.012	**.693**
Strange	-.053	.150	**.601**
Awkward	.049	.037	**.601**
Dangerous	-.035	.024	**.597**
Awful	.360	-.250	**.555**
Aggressive	.265	.009	**.547**
Eigenvalue	22.031	9.336	5.047
Percent variance explained	26.54%	11.25%	6.09%

* Steven J. Stroessner, The Robotic Social Attributes Scale (RoSAS): Development and Validation, 2017

MEASURING TRUST

"Trust = a dyadic relation in which one person accepts vulnerability because they expect that the other person's future action will have certain characteristics; these characteristics include some mix of performance (ability, reliability) and/or morality (honesty, integrity, and benevolence)."

Measurement MDMT includes these two types of trust:

- **performance trust** in human-robot interaction
- **moral trust**, based on the notion that the system may or may not act morally.

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

Figure 2. Items within each subscale of the Multi-Dimensional Measure of Trust (MDMT), developed after sorting study and used in expectation change study.

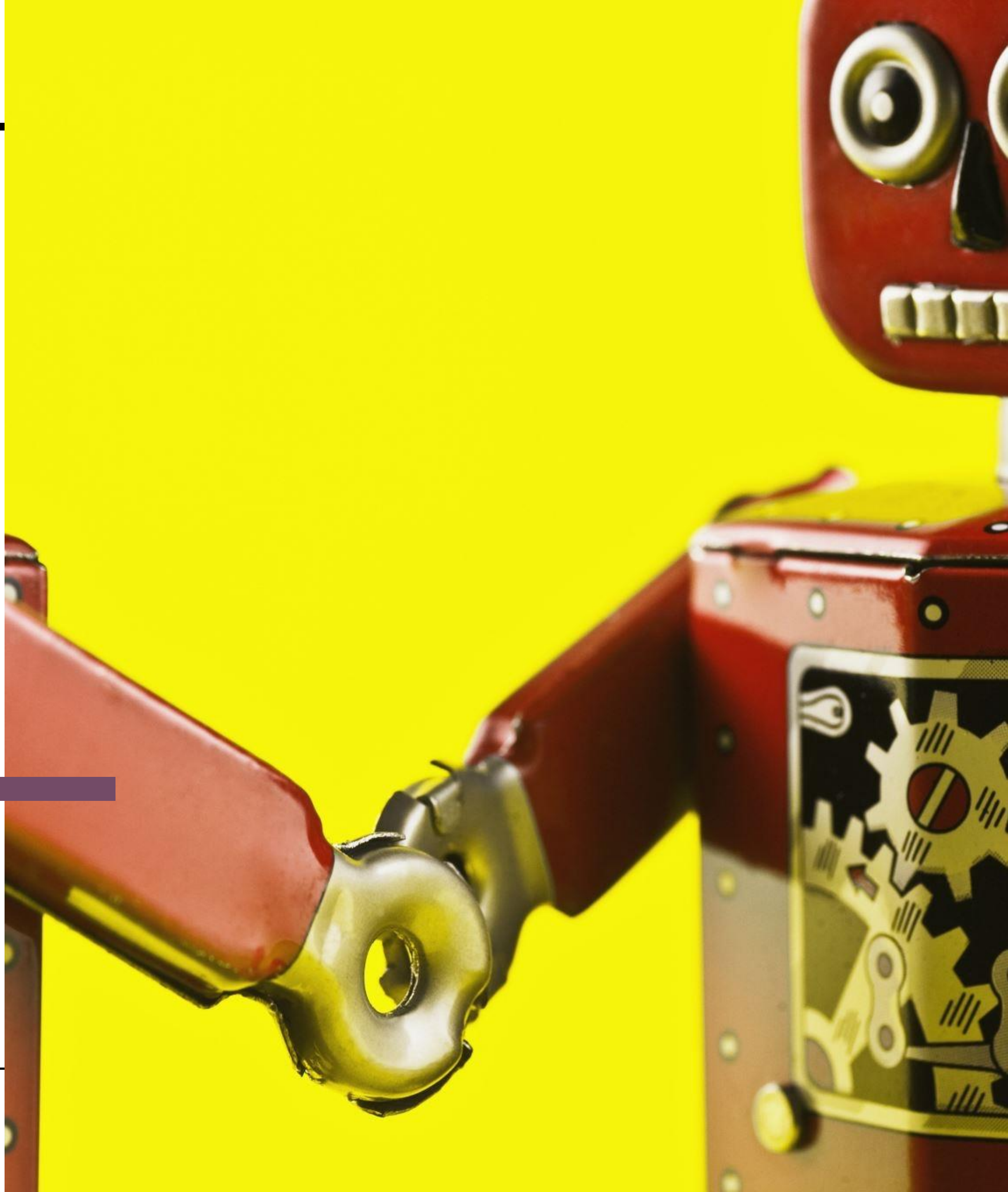
Malle, B. F., & Ullman, D. (2021). A multi-dimensional conception and measure of human-robot trust. In C. S. Nam and J. B. Lyons (eds.), Trust in human-robot interaction: research and applications (pp. 3-25). Elsevier.

YET

- Very often we cannot find differences in these subjective measures
- So, it is important to also ask directly the impressions that participants have
- Plus, these subjective measures depends on the scenario chosen

OVERVIEW OF LECTURE

- Experimental design
  - Research questions
  - Hypotheses
  - Variables
  - Type of experiments
- Typical measures in HRI
- Situating the experiment in the scientific world
- Running an experiment



Situating the experiment in the scientific world

"Reliability"

Yielding the same or compatible results in different experiments or statistical trials"

1. *How to achieve?* the dependent variable as precisely as possible
2. Depends on the definition
3. Measure of the variable (use an operational definition of what we want to measure)

"Validity"

Internal validity: the results obtained are due to the manipulation If the results obtained are not due to the manipulation, but rather to some other factors, then the study lacks internal validity. To get validity a good experimental design must be created.

External validity: If the results obtained are general for humanity, (and not rather valid for the specific situation).

Importance/ Generality

How do the results generalize

Is it important?

Situating the experiment in the scientific world

Threats to internal validity:

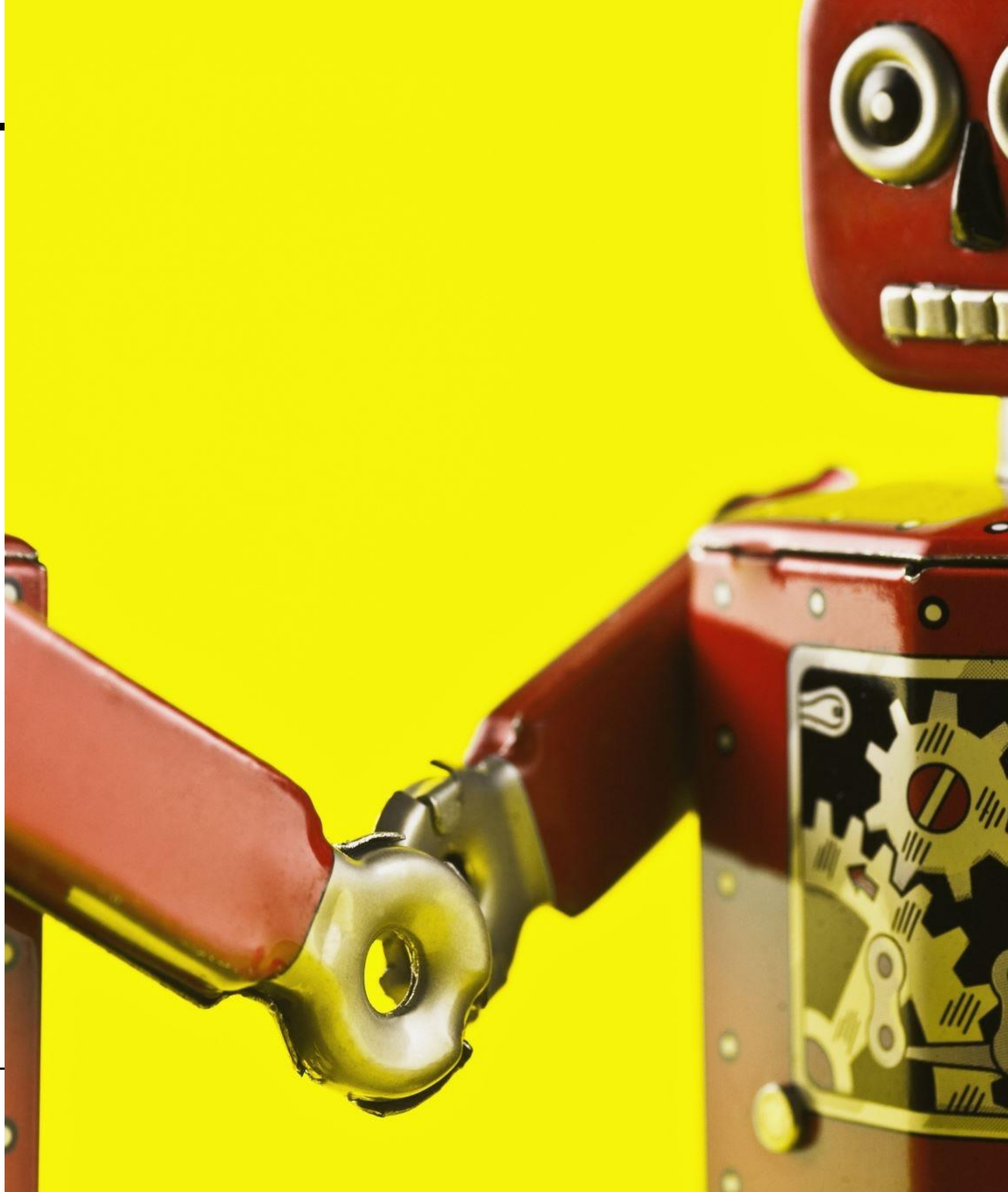
- Group threats: if the experimental and control groups are different to start with (then we may be measuring those differences and not the result of the manipulation)
- Regression to the mean: if the participants produce extreme scores in a pretest that by chance are likely to be the score of a post-test.
- Time threats: when the passage of time, with events happening produce changes in behaviour of participants
- History: events of the participants lives may lead to results that are unrelated to the manipulation.

Threats to external validity:

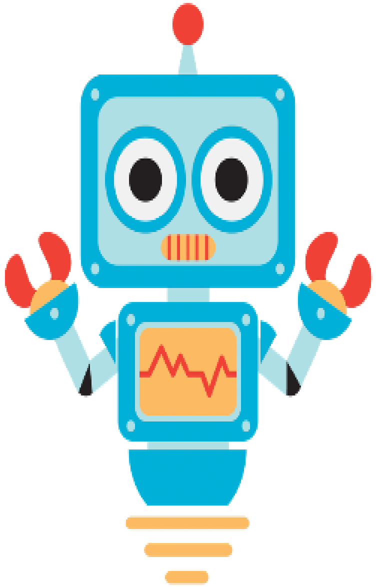
- Over-use of special participant groups (most HRI is tested with males, aged between 20-24, studying computer science and electrical engineering, most psychology studies are carried out with females, aged between 20-24 studying psychology).
- Too few participants…..

OVERVIEW OF LECTURE

- Experimental design
    - Research questions
    - Hypotheses
    - Variables
    - Type of experiments
- Typical measures in HRI
- Situating the experiment in the scientific world
- Running an experiment



TÉCNICO LISBOA



RUNNING AN EXPERIMENT:

PRACTICAL ISSUES AND METHODOLOGICAL ASPECTS OF CONDUCTING STUDIES IN HRI

TÉCNICO LISBOA

FIRST: PLANNING, LEGAL AND SAFETY

Usually before running any study involving humans and robots that will be interacting physically, certain legal, safety and ethical issues need to be considered.

- Ethical Approval (many organisations require an Ethics Committee to give approval for any research done with humans);
- Creation of a study protocol
- The Ethics committee must look into:
  - Privacy of the data (videos, photos, records must be kept private)
  - Protection of minors and vulnerable adults
  - Mental or emotional stress and humiliation
  - Physical harm

STUDY PROTOCOL CREATION: EXAMPLE

RESEARCH QUESTIONS

- RQ 1: Understand how children envision a robotic assistant to integrate futuristic classrooms.

PARTICIPANTS

- Participants will be: Children from 11-13 years old
- Participants will be grouped in N to perform the activity.

MATERIAL

- Drawing material (e.g., pen, pencil)
- One audio recorder
- One video camera
- Consent forms
- Robot NAO running the application

PROCEDURE

The activity will be conducted in a Portuguese school.

To participate in this study, users are required to have the consent form signed

Users will be grouped in groups of 2.

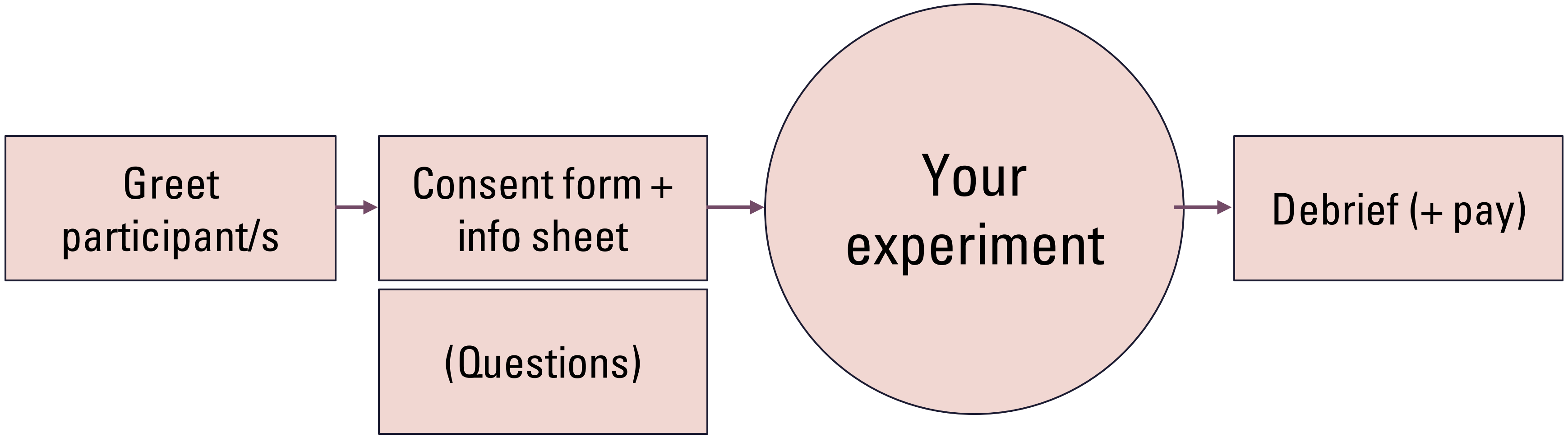
The activity is divided into three different phases:

PHASE 1 – QUANTITATIVE DATA COLLECTION

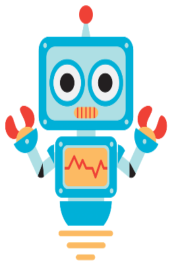
PHASE 2 – THE INTERACTION PHASE

PHASE 3 – THE REFLECTION PHASE

RUNNING AN EXPERIMENT: TYPICAL EXPERIMENTAL SESSION



Participants



- Respect your participants: you are representing your university and they are real people, they have rights!

- Protect your study variables do not influence participants on how you would like them to behave;

- Protect your sample! If you bring friends do not tell them in advance the aim of the study or you will be manipulating your results!

In online surveys: they will also be working (even in online studies)

Often attention checks are needed to guarantee the participants are engaged with the experiment

INTERACTION WITH PARTICIPANTS

Participants are real people and they are doing you a favour!

They CAN leave an experiment halfway through and you should still pay them (it's called a show-up fee for a reason)

(but don't worry, this rarely happens)

If you realise that a participant is doing everything wrong for some reason, don't call them out: it's too late to make them start again. Wait for them to finish and then consider discarding their data

Never give participants individual feedback on their performance in the experiment

Don't tell them how they did compared to other participants

If anything, give them the average result of all participants

PARTICIPANTS' INDIVIDUAL DIFFERENCES

- People are different, and it's hard to find something that is 100% generalisable
- Some differences you can control (to some extent)
  - Gender
  - Age group
  - Language fluency

- You should collect demographics data
  - We need to know if / how your results generalise
- Note on gender: recently recommended categories for HCI experiments are:
  - Male
  - Female
  - Non-binary
  - Prefer to self describe
  - Prefer not to say

Running the experiment: Minimise Bias

Participants are eager to please you

- If they think you want them to answer in a certain way, they will
- Check the questions you are asking! Are they biasing?

"I enjoyed interacting with the robot" vs.

"I did not enjoy interacting with the robot" vs.

"Please rate your experience with the robot"

- Don't tell participants what your manipulation is before the experiment
  - (you should tell them afterwards, in a debriefing)
- Ideally, if participants are randomly allocated to conditions, you should also not know what condition one participant is in
- Wait outside the room where the experiment is running
- Consider having 50% participants do the experiment with a male experimenter, 50% with a female experimenter

Running the experiment: Minimise Noise

Minimise Noise

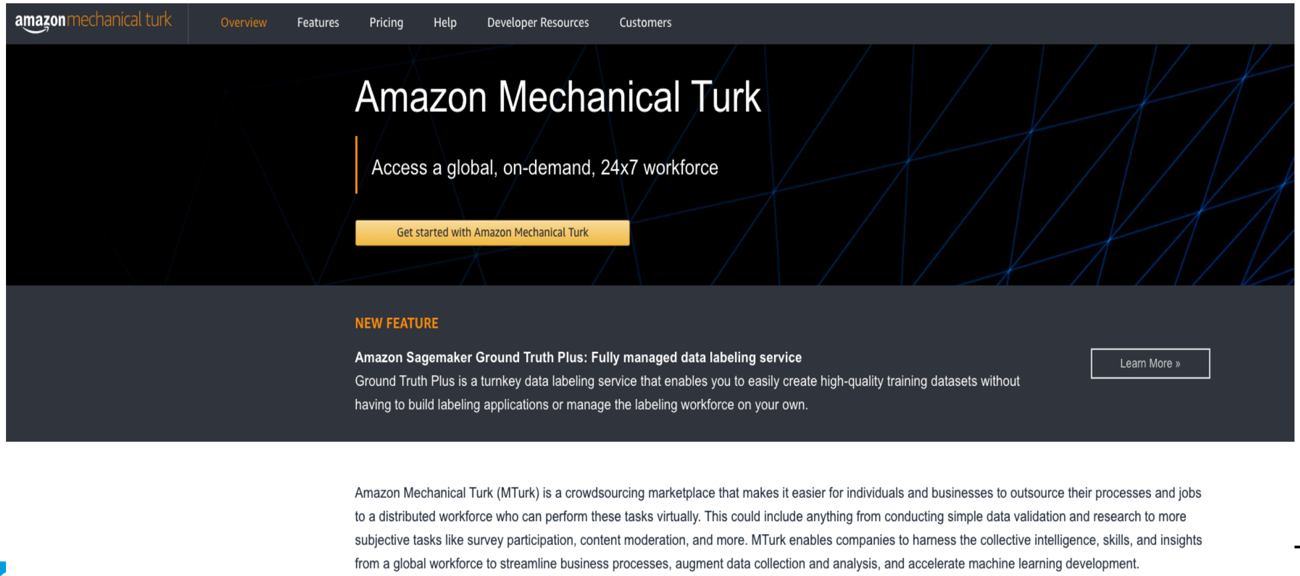
- Noise is everything you can't control:
  - how participants are feeling on that particular day
  - how hot it is
  - participants' personality…
- Your goal is to keep as much of this variability as you can under control

- Keep the environment constant
  - Lighting
  - Computer orientation
  - Audio levels
  - Device calibration
- Keep "yourself" constant
  - Make a little "script" and always use it when greeting participants
  - Think about what questions participants might ask you before the experiment, and be ready to always answer them in the same way
  - Don't read / recite the instructions, put them on a screen / printout and let participants read them
  - Automate as much as possible

ONLINE STUDIES: TOOLS

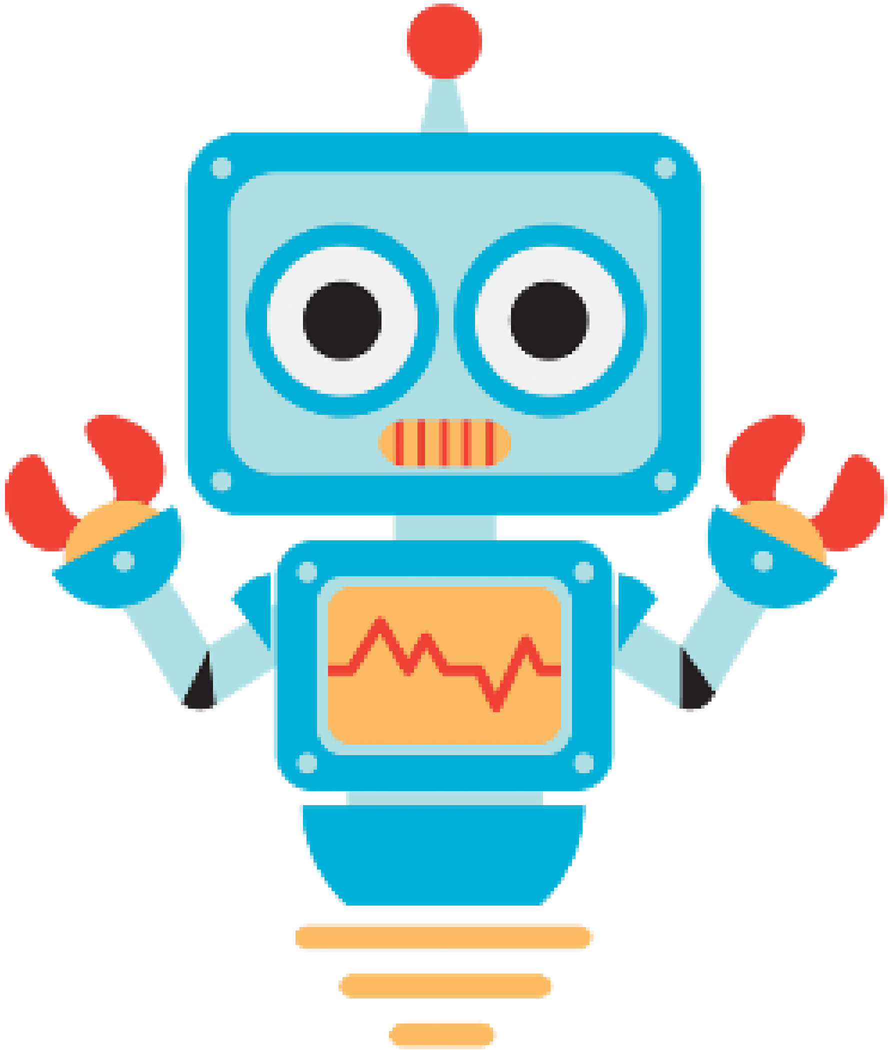
HRI research can also be conducted online!

- Mturk
  - Pros: widely used, you can have hundreds of participants overnight
  - Cons: run by big corp, criticised for unfair treatment of participants



- Prolific
  - Pros: higher quality data , made specifically for research
  - Cons: less widely used

https://www.umass.edu/research/guidance/mturk-guidance
https://researcher-help.prolific.co/hc/en-gb/articles/360034677314-Quick-guide-to-using-Prolific



IMPORTANT: DON'T FORGET

- Consent Forms signed by all the participants
- Participants anonymity (number tags and protection of data – DO NOT USE NAMES)
- Do not share data outside of your team the data you gathered (e.g.videos)
- Do not influence questionnaires responses!
- Don't influence the task: You cannot be in the same room that the participants are doing the task or somewhere else where a participant sees you.
- Don't blame the participants: If things do not go according to plan it is never the participant fault!
- Never interrupt a session until it is the participant that comes and ask for help.

QUESTIONS

TO READ

- *How to Design and Report Experiments, Andy Field &Graham Hole, SAGE Publications, 2003.*
- *Malle, B. F., & Ullman, D. (2021). A multi-dimensional conception and measure of human-robot trust. In C. S. Nam and J. B. Lyons (eds.), Trust in human-robot interaction: research and applications (pp. 3-25). Elsevier.*