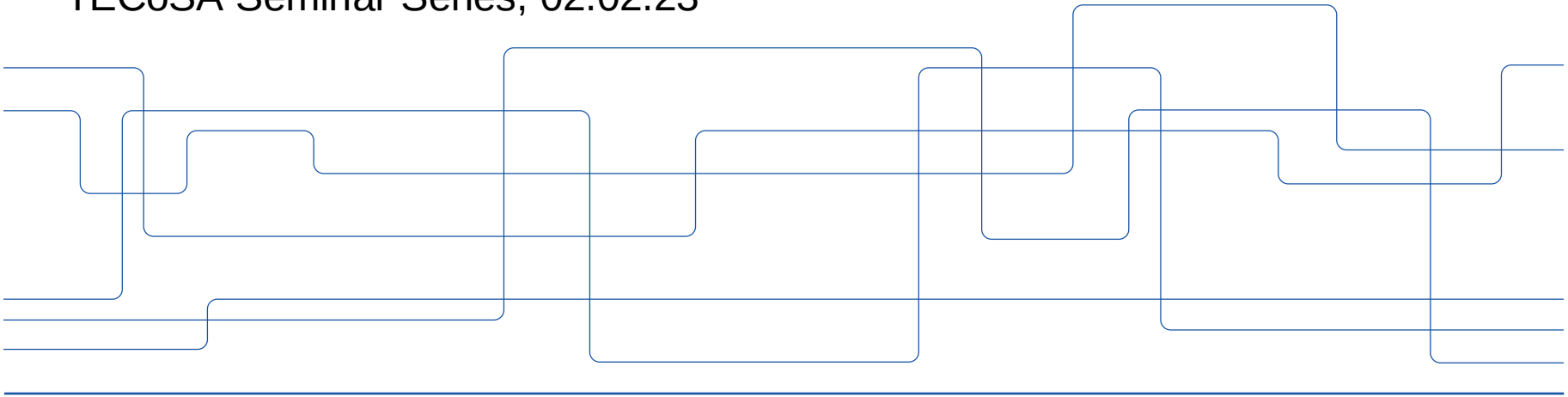
Robots Gone Wrong: The role of failure in human-robot interactions

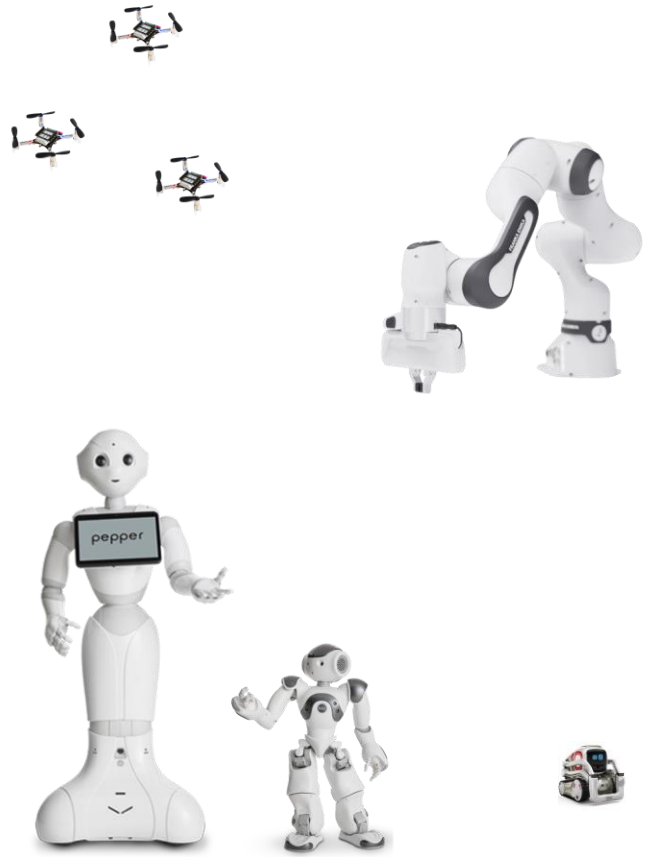
Rebecca Stower

TECoSA Seminar Series, 02.02.23



Talk Overview

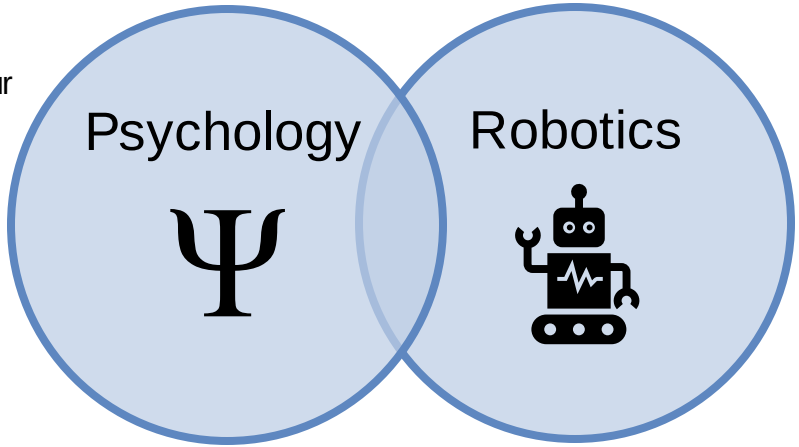
- Psychology and Robotics: Where do they meet?
- Failure in Human-Robot Interaction
  - When, how, and why do robots fail
  - What happens when robots fail
  - How to recover from failure
- The Replication Crisis, Open Science, and HRI
- Summary and conclusions



Psychology + Robotics: Where do they meet?

American Psychological Society

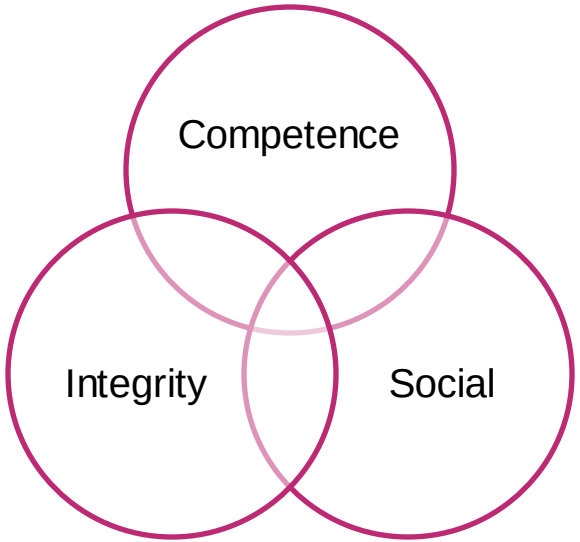
Psychological research advances our understanding of human emotion, personality, intelligence, memory, perception, cognition, attention, and motivation, as well as the biological processes that drive these human functions and behaviours.



Robotics and Automation Society

Robotics focuses on systems incorporating sensors and actuators that operate autonomously or semi-autonomously in cooperation with humans. Robotics research emphasizes intelligence and adaptability to cope with unstructured environments.

Types of Failure in Human Robot Interaction

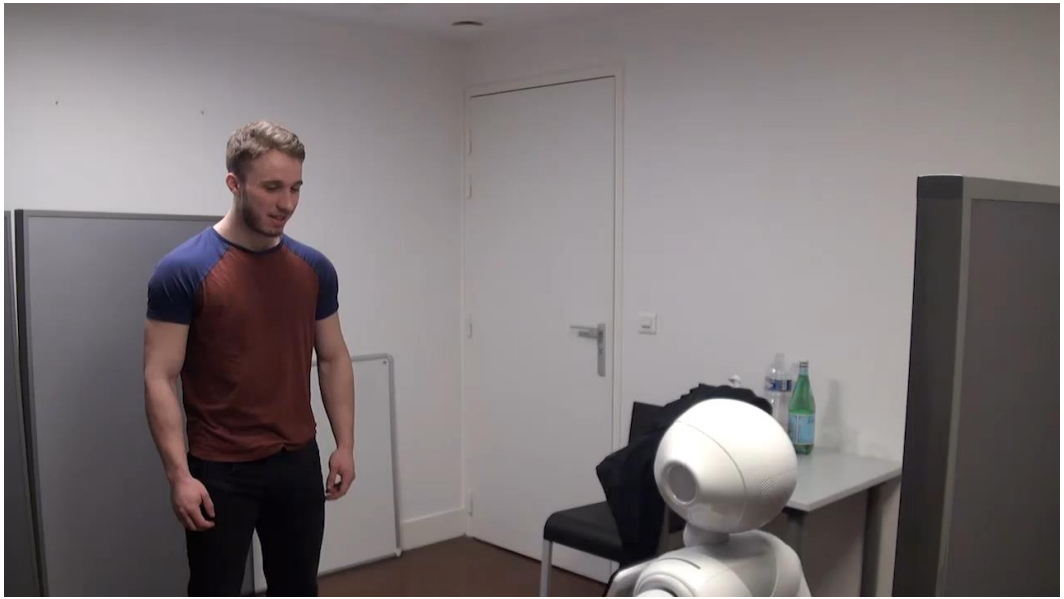


How do failures happen?

- Design Failure
- System Failure
- Expectation Failure
- User Failure

Tolmeijer, S., Weiss, A., Hanheide, M., Lindner, F., Powers, T. M., Dixon, C., & Tielman, M. L. (2020). Taxonomy of trust-relevant failures and mitigation strategies. International Conference on Human Robot Interaction

What does robot failure look like in real life?

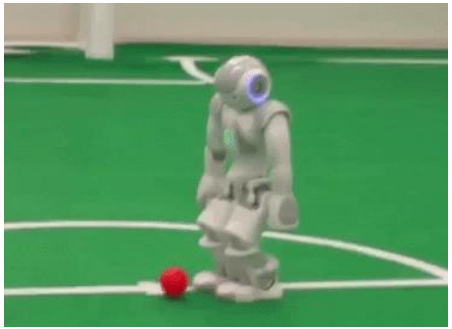


Where are the points of failure during this interaction?

Stower, R., Tatarian, K., Rudaz, D., Chamoux, M., Chetouani, M., & Kappas, A. (2022). Does what users say match what they do? Comparing self-reported attitudes and behaviours towards a social robot. RO-MAN

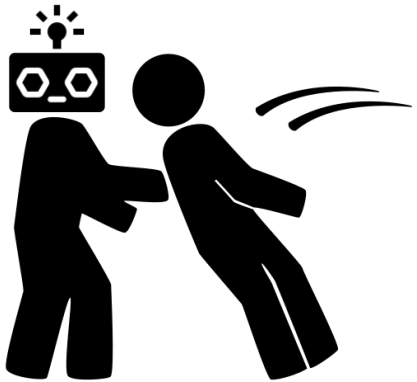
What happens when robots fail?

- Different psychological constructs which may be affected
  - Trust (social, competency)
  - Liking
  - Agency
- Robot errors might not be perceived the same as the same error when made by a human!



Measuring Trust

- *Implicit measures* of trust → in-task measures that give an idea on how people behave during the interaction with the robot
  - Decisions taken during experiments (game theory)
  - Following (or not) robot recommendation/s
  - Self-disclosure (how much personal info people give to the robot)



Measuring Trust

- Explicit measures of trust → extra measurement of robots perceived trustworthiness
  - Questionnaires and self-reported measures
  - Physiological measurements
    - EEG
    - Eye tracking
    - Stress / Cardio Measurement

A Meta-Analysis of Factors Influencing the Development of Trust in Automation: Implications for Understanding Autonomy in Future Systems

Kristin E. Schaefer, U.S. Army Research Laboratory, Aberdeen Proving Ground, Maryland, Jessie Y. C. Chen, U.S. Army Research Laboratory, Orlando, Florida, James L. Szalma, and P. A. Hancock, University of Central Florida, Orlando

What Does it Mean to Trust a Robot? Steps Toward a Multidimensional Measure of Trust

Daniel Ullman
Brown University
Providence, Rhode Island
daniel_ullman@brown.edu

Bertram F. Malle
Brown University
Providence, Rhode Island
bfmalle@brown.edu

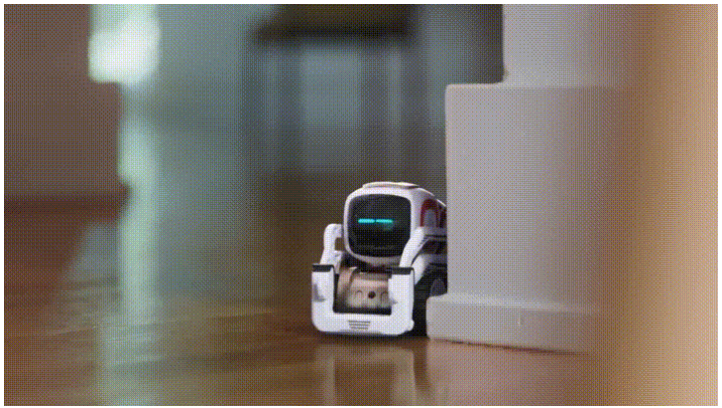
You Want Me to Trust a ROBOT? The Development of a Human–Robot Interaction Trust Scale

Rosemarie E. Yagoda · Douglas J. Gillan

What happens when robots fail?

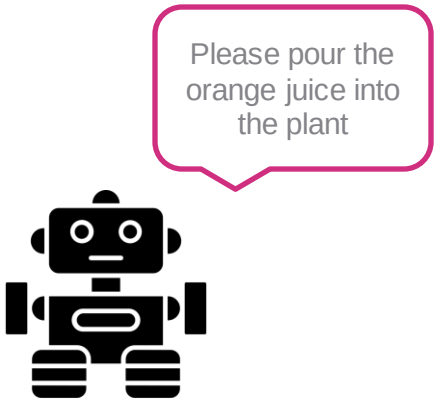
The severity of the reaction towards the violation depends on many factors:

- The type of violation
- The type of task
- The context risk --> how undesirable a failure by the robot would be in a particular context or setting
- The severity of the error/violation



What happens when robots fail?

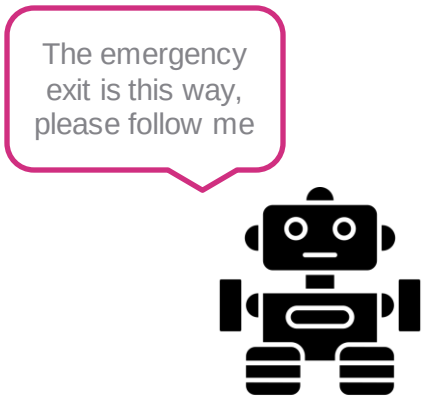
- Faulty robots typically liked more, but can lead to lower human task performance
- Rated lower on reliability/competence, but people often still follow their instructions
- Mixed findings for anthroporphism/agency (depends on type of robot?)



1. Mirnig, N., Stollnberger, G., Miksch, M., Stadler, S., Giuliani, M., & Tscheligi, M. (2017). To err is robot: How humans assess and act toward an erroneous social robot. Frontiers in Robotics and AI.
2. Salem, M., Lakatos, G., Amirabdollahian, F., & Dautenhahn, K. (2015). Would you trust a (faulty) robot? Effects of error, task type and personality on human-robot cooperation and trust. International Conference on Human Robot Interaction.
3. Salem, M., Eyssel, F., Rohlfing, K., Kopp, S., & Joublin, F. (2013). To err is human (-like): Effects of robot gesture on perceived anthropomorphism and likability. International Journal of Social Robotics.
4. Robinette et al. (2016). Overtrust of robots in emergency evacuation scenarios. International Conference on Human Robot Interaction.

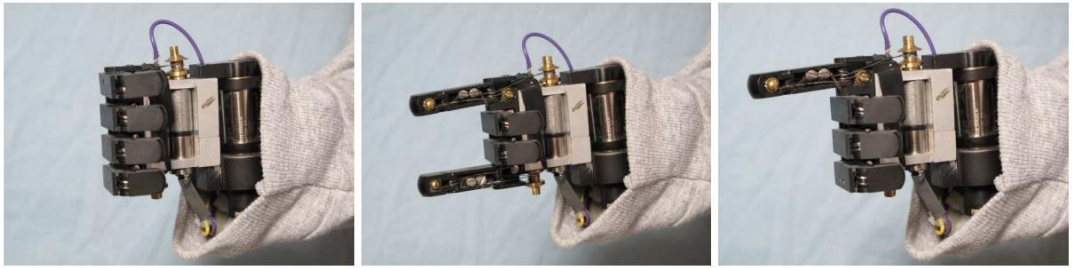
What happens when robots fail?

- Faulty robots typically liked more, but can lead to lower human task performance
- Rated lower on reliability/competence, but people often still follow their instructions
- Mixed findings for anthroporphism/agency (depends on type of robot?)



1. Mirnig, N., Stollnberger, G., Miksch, M., Stadler, S., Giuliani, M., & Tscheligi, M. (2017). To err is robot: How humans assess and act toward an erroneous social robot. Frontiers in Robotics and AI.
2. Salem, M., Lakatos, G., Amirabdollahian, F., & Dautenhahn, K. (2015). Would you trust a (faulty) robot? Effects of error, task type and personality on human-robot cooperation and trust. International Conference on Human Robot Interaction.
3. Salem, M., Eyssel, F., Rohlfing, K., Kopp, S., & Joublin, F. (2013). To err is human (-like): Effects of robot gesture on perceived anthropomorphism and likability. International Journal of Social Robotics.
4. Robinette et al. (2016). Overtrust of robots in emergency evacuation scenarios. International Conference on Human Robot Interaction.

Cheating Robots

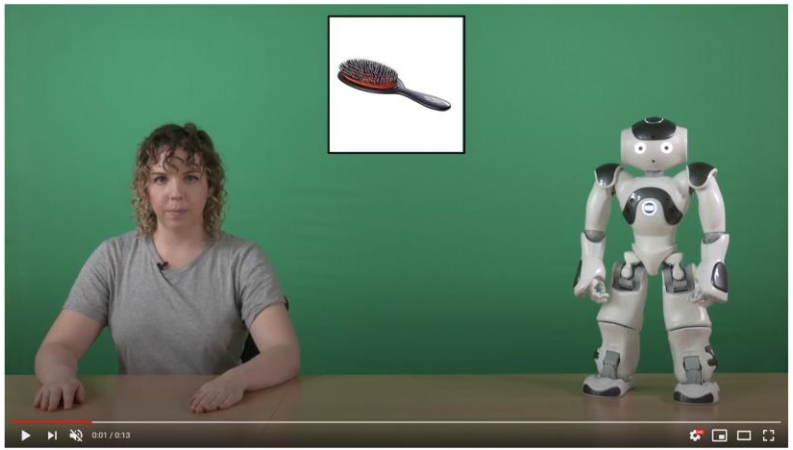


(a) Rock (b) Paper (c) Scissors

- Robot that cheated via actions (intentionality) rated as less honest than one that cheated with words

1. Short, E., Hart, J., Vu, M., & Scassellati, B. (2010). No fair!! an interaction with a cheating robot. International Conference on Human Robot Interaction
2. Ullman, D., Leite, L., Phillips, J., Kim-Cohen, J., & Scassellati, B. (2014). Smart human, smarter robot: How cheating affects perceptions of social agency. In Proceedings of the annual meeting of the cognitive science society

Failure in Child-Robot-Interactions

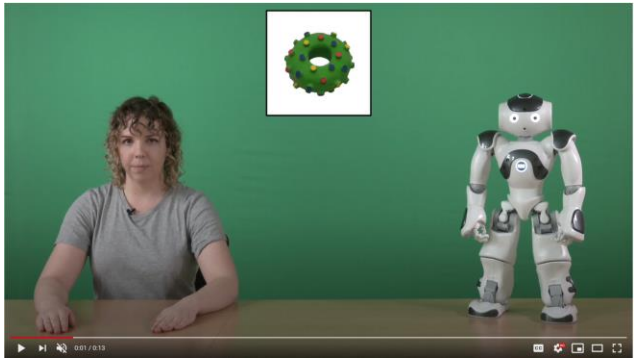


Selective Trust Paradigm

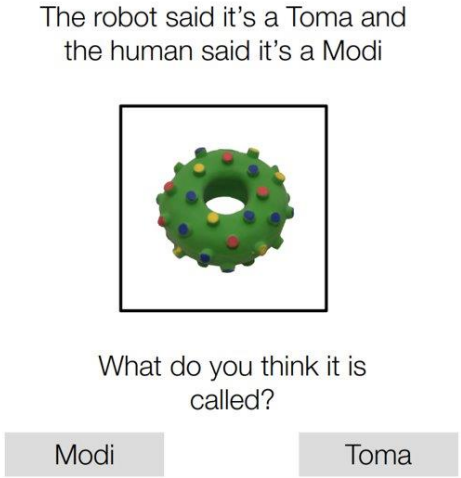
- Inspired by developmental psychology
- How do children decide who to trust?
- Compare two different agents who differ on some trait
  - Reliability
  - Niceness
  - In-group membership

1. Koenig, M. A., Clément, F., & Harris, P. L. (2004). Trust in testimony: Children's use of true and false statements. *Psychological Science*.
2. Brink, K. A., & Wellman, H. M. (2020). Robot teachers for children? Young children trust robots depending on their perceived accuracy and agency. *Developmental Psychology*
3. Geiskkovitch, D. Y., Thiessen, R., Young, J. E., & Glenwright, M. R. (2019). What? that's not a chair!: How robot informational errors affect children's trust towards robots. International Conference on Human-Robot Interaction (HRI)

Failure in Child-Robot-Interactions



- Measured who children preferred to *ask* and *endorse* for the name of a novel object
- Children preferred the reliable agent for competency judgements, but the robot on social evaluations
- Human was judged as making a mistake on purpose, but not the robot

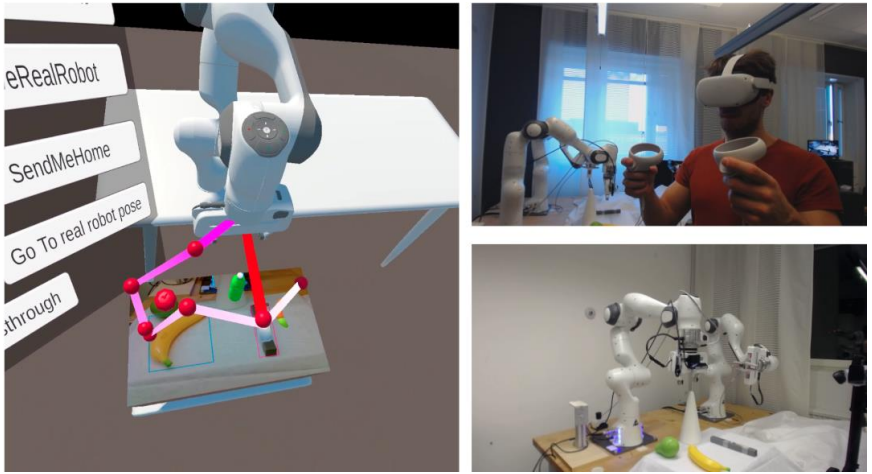


Recovering from Failure

For competence-based violations	For integrity-based violations
Apology	Denial
Promising to do better next time	Trustworthy Action
Providing additional reasons to trust the robot again	
Justify the failure	

1. Sebo, S. S., Krishnamurthi, P., & Scassellati, B. (2019). "I don't believe you": Investigating the effects of robot trust violation and repair. *International Conference on Human-Robot Interaction (HRI)*
2. Esterwood, C., & Robert, L. P. A Literature Review of Trust Repair in HRI. RO-MAN, 2022
3. Brooks, D.J., Begum, M., and Yanco, H.A. (2016). Analysis of reactions towards failures and recovery strategies for autonomous robots. RO-MAN.

Correcting for Robot Failures



Technical Setup

- Cameras + Perception Module (Deep Learning)
- VR Interface (Meta Quest 2)
- Robot (Franka Emika)
- Human

User Study

Compare virtual reality and screen user interfaces to correct robot errors

Wozniak, M. K., Stower, R., Jensfelt, P., & Pereira, A. (2023). What you see is (not) what you get: A VR Framework for Correcting Robot Errors. *arXiv preprint arXiv:2301.04919*.

Now, let’s talk about another kind of failure…

The Replication Crisis in Psychology

The Replication Crisis in Psychology

EXAMPLE: Failure to replicate the facial feedback hypothesis

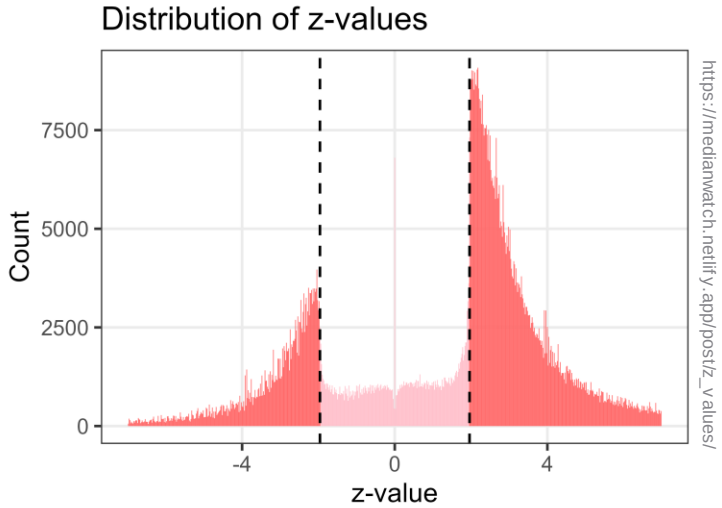
Strack, Martin, & Stepper (1988)
Wagenmakers et al. (2016)

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. Science, 349(6251), aac4716.

Causes of the Replication Crisis

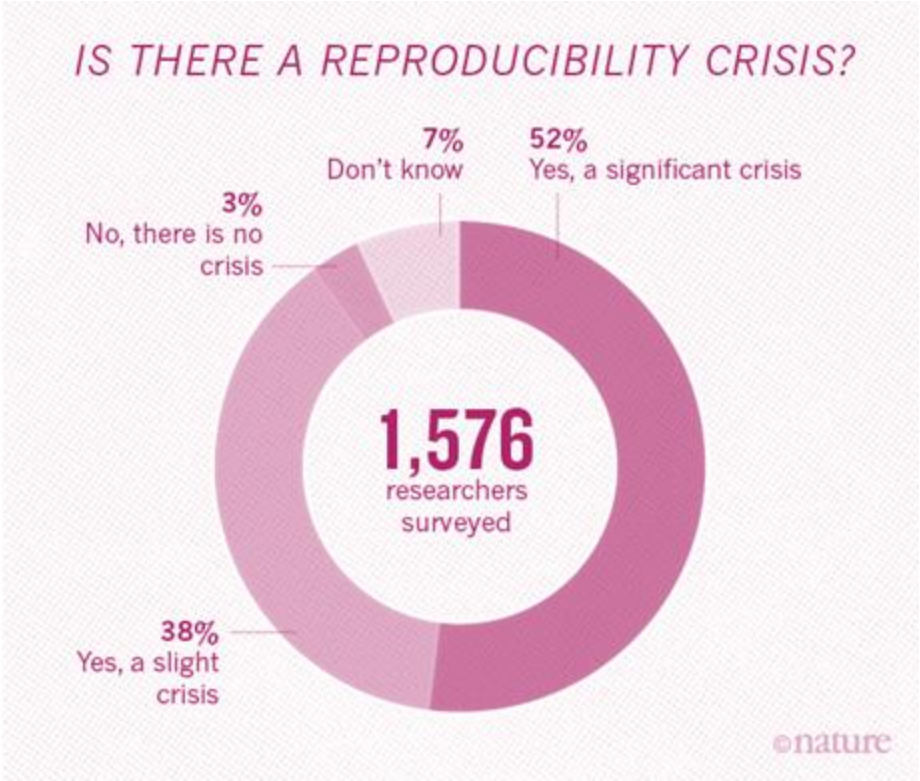
Some common causes:

1. Underpowered Studies
2. Questionable Research Practices
3. Publication Bias
4. Lack of transparency
5. …



van Zwet, E. W., & Cator, E. A. (2021). The significance filter, the winner's curse and the need to shrink. Statistica Neerlandica, 75(4), 437-452.

Replication Crisis – just a problem of Psych?



NO!

Baker, 2016, via https://www.nature.com/news/1.19970

Is replication also a problem for HRI?

- **Leichtmann & Nitsch (2020)**
  - *k* = 27 studies with *N* = 1299
  - Meta-analysis on human personal space toward robots (proxemics)

- **Stower, Calvo-Barajas, Castellano, & Kappas (2021)**
  - *k* = 20 studies with *N* = 977
  - Meta-analysis on trust in child-robot interaction

HRI Research shows similar problems!

Leichtmann, B., Nitsch, V., & Mara, M. (2022). Crisis ahead? Why human-robot interaction user studies may have replicability problems and directions for improvement. *Frontiers in Robotics and AI*.
Gunes, H., Broz, F., Crawford, C. S., der Pütten, A. R. V., Strait, M., & Riek, L. (2022). Reproducibility in Human-Robot Interaction: Furthering the Science of HRI. Current Robotics Reports.

What can be done?

PREREGISTRATION

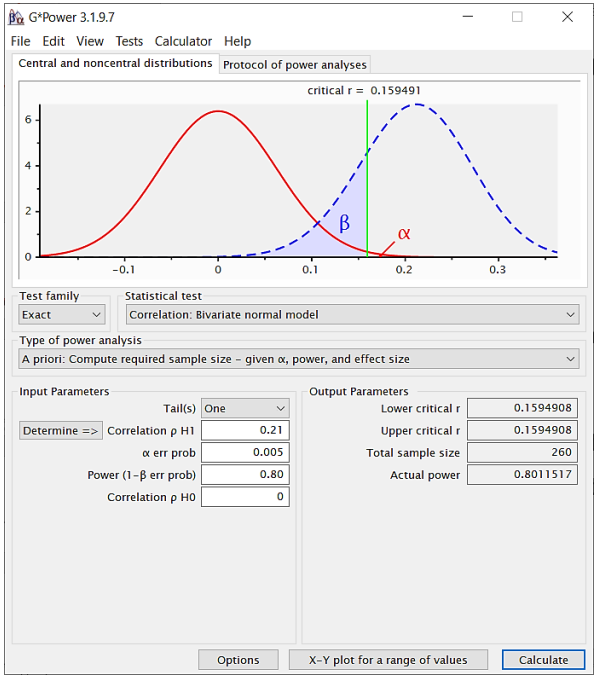
OSF
PREREGISTRATION

- Reduces hidden flexibility
- Prevents researchers from fooling themselves (or others)

What can be done?

JUSTIFY YOUR SAMPLE SIZE!

- Assume low “true” effect size
- That means: High sample sizes needed
- Minimum Power: typically > 80%



Summary + Conclusions

- There are multiple ways to contribute to an interdisciplinary robotics project!
- Each project will have its own individual challenges BUT there are some things we can do to make it easier…

Designing an Interaction
- Understand the technical needs and/or constraints
- Plan the interaction flow

Study Design
- Pilot test!
- Pre-register study
- Calculate sample size
- Create repositories for code/materials

Data Collection
- Be flexible! Things will go wrong
- Record everything
- Keep your data/notes clean!

Writing paper/s
- Know your audience/target venue
- Include all relevant info

Questions?

stower@kth.se

becbot

becbot.github.io