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

What does robot failure look like in real life?

Where are the points of failure during this interaction?

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
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?)

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?)

The emergency exit is this way, please follow me

Cheating Robots

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

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

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

<table>
<thead>
<tr>
<th></th>
<th>For integrity-based violations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Apology</td>
<td>Denial</td>
</tr>
<tr>
<td>Promising to do better next time</td>
<td>Trustworthy Action</td>
</tr>
<tr>
<td>Providing additional reasons to trust the robot again</td>
<td></td>
</tr>
<tr>
<td>Justify the failure</td>
<td></td>
</tr>
</tbody>
</table>

Correcting for Robot Failures

User approves/modifies planned trajectory.

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

Robot fails to detect an object in the real-world environment.

User uses VR interface to add missing object/s.

Virtual robot plans trajectory.

User approves/modifies planned trajectory.

Real world robot executes movement.
Compare virtual reality and screen user interfaces to correct robot errors

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)

Causes of the Replication Crisis

Some common causes:

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


https://medianwatch.netlify.app/post/z-values/
Replication Crisis – just a problem of Psych?

IS THERE A REPRODUCIBILITY CRISIS?

- 52% Yes, a significant crisis
- 38% Yes, a slight crisis
- 3% No, there is no crisis
- 7% Don't know

1,576 researchers surveyed

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!


2023-02-20
What can be done?

PREREGISTRATION

Plan study

Upload study and analysis plan to an open repository

Collect & analyze data

Write report

Publish report

• 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

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

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

Writing paper/s
• Know your audience/target venue
• Include all relevant info
Questions?

✉️ stower@kth.se
🐦 becbot
🌐 becbot.github.io