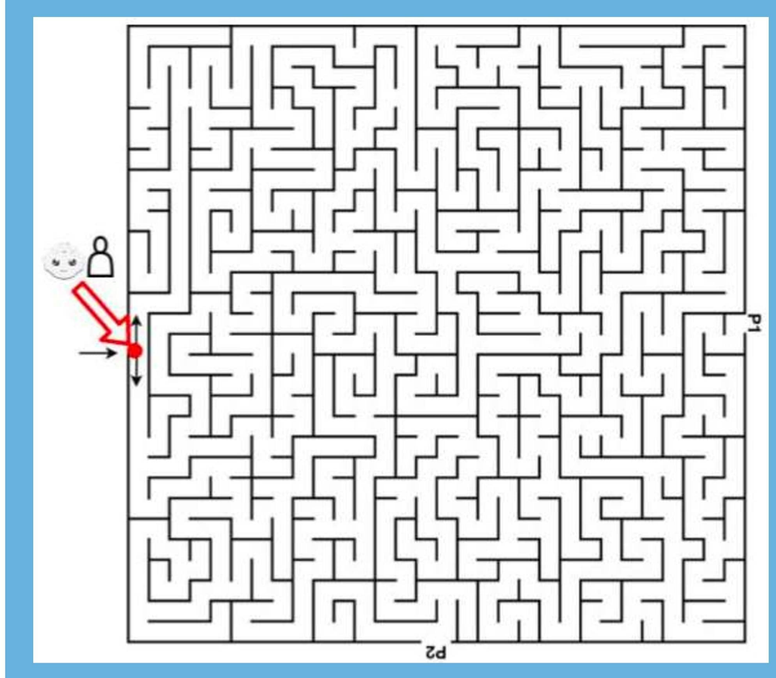
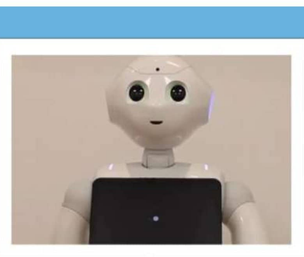
Exploring Theory of Mind for Human-Robot Collaboration

M. Romeo[1], P. E. McKenna[2], D. A. Robb[2], G. Rajendran[2], B. Nesset[2], A. Cangelosi[1], H. Hastie[2]
UKRI Node on Trust
[1]University of Manchester, [2]Heriot-Watt University

Overview

The ability to impute mental states to oneself or others, or Theory of Mind (ToM), has been linked to trust between humans. Less is known about how a robot mimicking ToM affects users' trust and behaviour. We explore this through a study where we compare 3 robot personas in a cooperative maze navigation task: (i) neutral (*n*=235), (ii) explaining its reasoning in technical terms (*n*=236), (iii) mimicking ToM (*n*=235). Results show that 'robot ToM' led to more cautious navigation decisions.





Condition	Utterance Example
Baseline	Go towards P2.
Sensor	Using my robot vision I can see P2 is the only reachable exit.
ToM	I suspect you think P2 is correct. I agree, we should follow your instinct and go right to reach it.

Figure 1. Maze example and Pepper's navigation suggestion utterances. ToM levels confirmed by manipulation check.

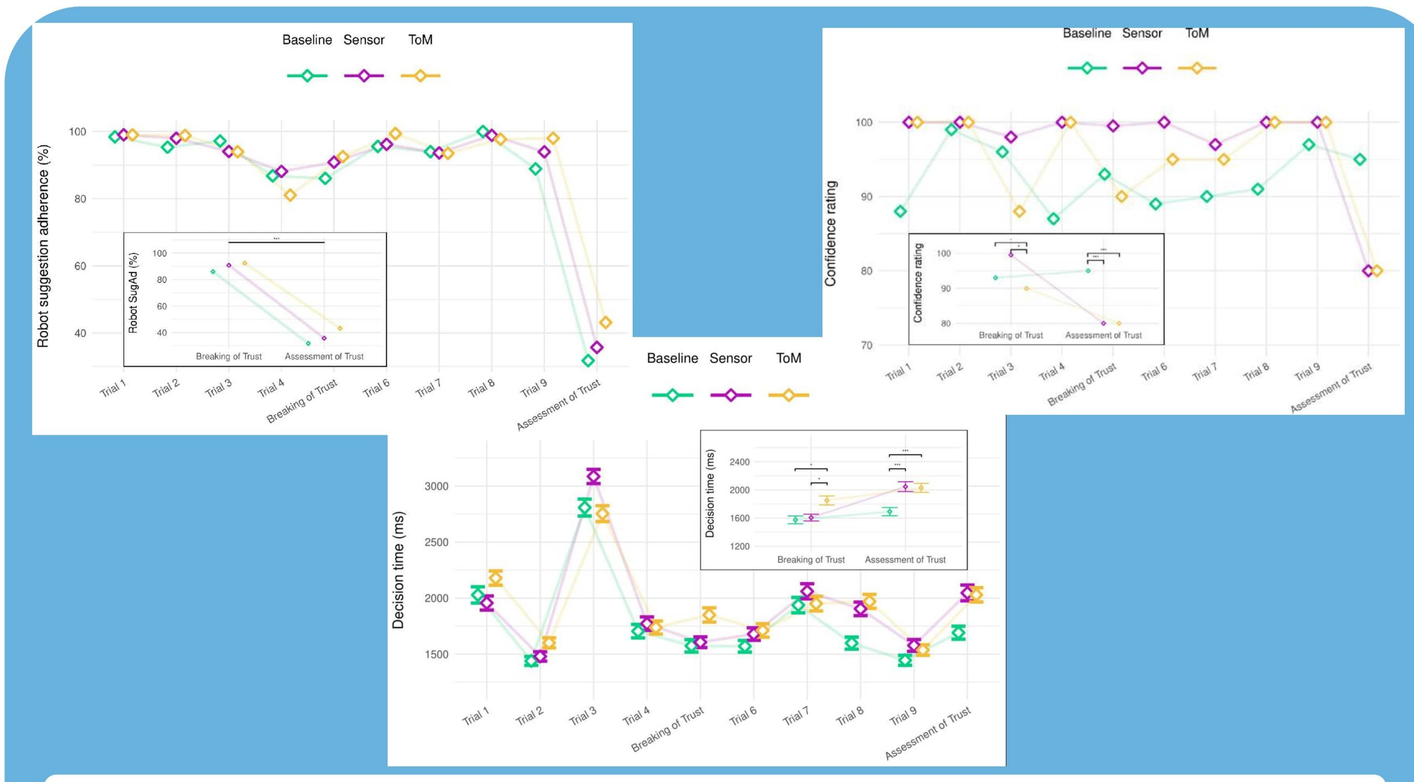


Figure 2. Results: Suggestion Adherence (left); Decision time (centre); Self-reported Confidence (right). Main plots: trend across trials. Inset plots: group comparisons & p-values.

Key Features

- Design: Online, between-subjects
- IV: ToM priming & maze utterance
- DV: Subjective & behavioural measures of trust (in-task & MDMT q'nnaire)
- Task: solve 10 mazes
- Trial: follow/not follow robot's advice. T5 → Trust broken
- Pepper advice vary by persona

Results

- MDMT inconclusive
- follow/not follow did not vary per group
- Mimicked ToM: longer decision times
- Sensor: higher confidence

Future Work

- Adding transparency of the interaction as a variable
- Analysis of different trust repair strategies
- Applications in: recommender systems, robot navigation, collaborative manufacturing

Engineering and Physical Sciences Research Council
HERIOT WATT UNIVERSITY
MANCHESTER 1824 The University of Manchester
Imperial College London
UKRI Trustworthy Autonomous Systems Hub