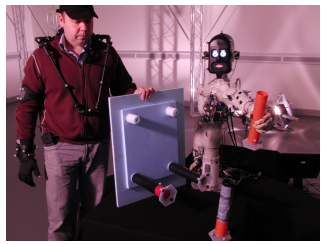
Trustworthy Robot Assistants

Clare Dixon
Department of Computer Science
University of Liverpool

Fraunhofer IPA

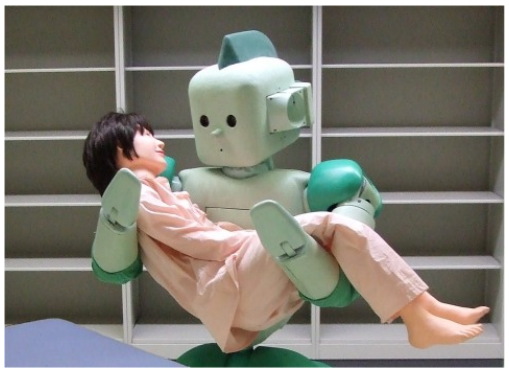


1 University of Liverpool (UoL)
2 University of Hertfordshire (UoH)
3 Bristol Robotics Lab (BRL)
www.robosafe.org

Farshid Amirabdollahian[2]
Kerstin Dautenhahn[2]
Kerstin Eder[3]
Michael Fisher[1]
Dejanira Araiza Illan[3]
Kheng Lee Koay[2]
Anthony Pipe[3]
Maha Salem[2]
Joe Saunders[2]
Matt Webster[1]
David Western[3]

Robotic Assistants

- Robot assistants are being developed to help, or work closely with humans in industrial, domestic and health care environments.
- The robots will need to be able to act autonomously and make decisions to choose between a range of activities.
- How do we make sure they are trustworthy and safe?



RI-MAN
rtc.nagoya.riken.jp/RI-MAN/

Pearl
www.cs.cmu.edu/ nursebot/

Wakamaru
www.mhi-global.com/products/detail/wakamaru.html

What is Trustworthiness and Safety?

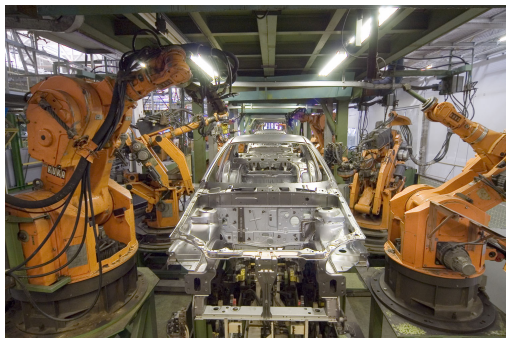
- Safety involves showing that the robot does nothing that (unnecessarily) endangers the person.
- The International Organization for Standardization (ISO) has developed ISO 13482, a standard relating to safety requirements for personal care robots.
- Trustworthiness involves social issues beyond pure safety.
- It is not just a question of whether the robots are safe but whether they are *perceived* to be safe, useful and reliable.
- A definition of trust: the firm belief in the reliability or truth or strength of a person or thing[1]
- There are also legal (and ethical) issues such as what happens when
  - the robot spills a hot drink on someone;
  - the robot doesn't remind the person to take their medicine;
  - the robot doesn't go to the kitchen when told?

[1] Concise Oxford Dictionary

Robots in the Workplace and at Home

Currently many robots in used in industry or domestic use operate in limited physical space or have limited functionality. This helps assure their safety.

- Robots industrial environments are limited so they can only move in a fixed area and have limited interactions with humans eg welding or paint spraying robots.
- Small or limited capability domestic robots, e.g., vacuum cleaning robots, robot lawn mowers, pool cleaning robots etc

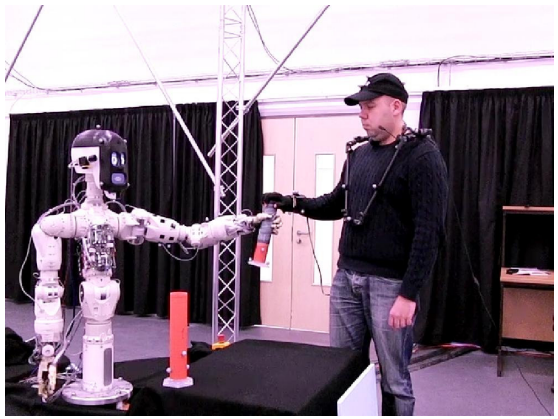


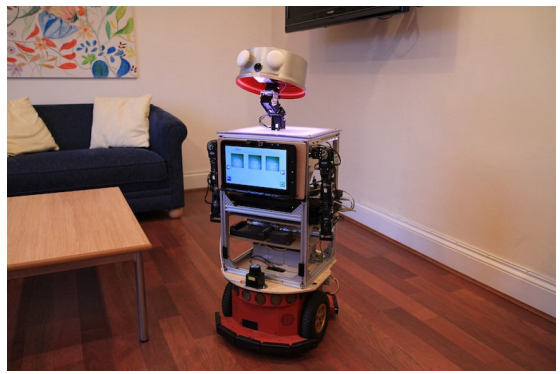
Trustworthy Robotic Assistants Project

The EPSRC funded Trustworthy Robotic Assistants Project develops three different approaches to verification and validation of robotic assistants.

Each approach is aimed at increasing trust in robotic assistants.

- Formal Verification (Liverpool)
- Simulation-based Testing (Bristol Robotics Laboratory)
- End-user Validation (Hertfordshire)

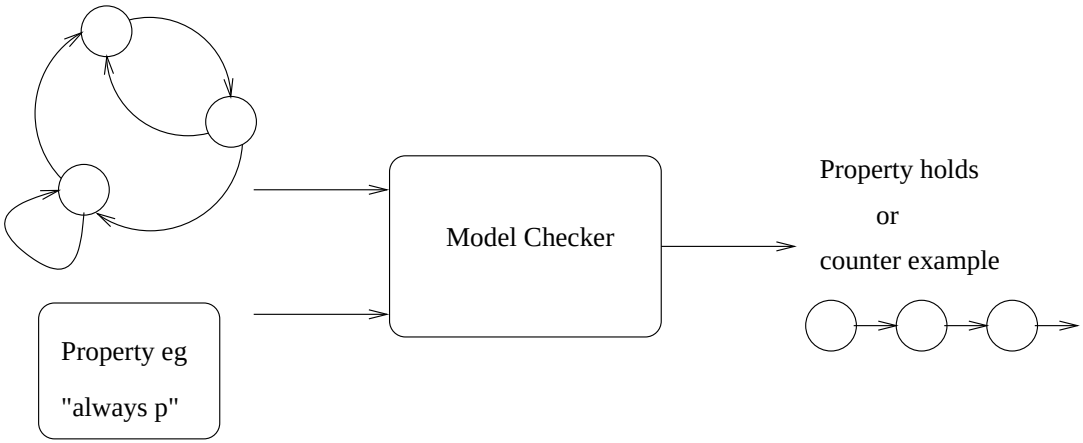




We consider two use cases personal robot assistant located in a domestic type house and a co-operative manufacturing task.

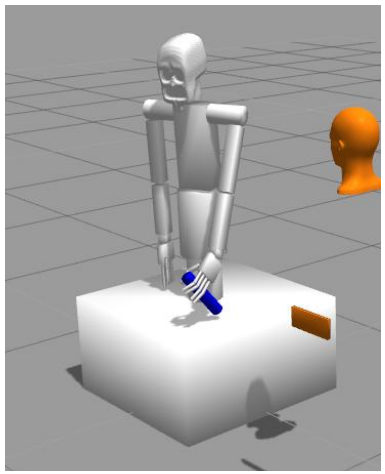
Formal Verification

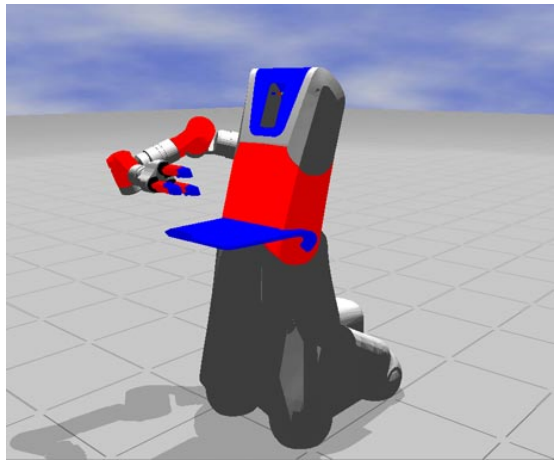
- A mathematical analysis of all behaviours using logics, and tools such as theorem provers or model checkers.
- We focus on temporal verification using automatic tools and techniques that do not require user interaction.
- Model checking is a fully automatic, algorithmic technique for verifying the temporal properties of systems.
- Input to the model checker is a model of the system and a property to be checked on that model.
- Output is that the property is satisfied or a counter model.



Simulation Based Testing

- This is an exhaustive testing methodology widely used in the design of micro-electronic and avionics systems.
- These appeal to Monte-Carlo techniques and dynamic test refinement in order to cover a wide range or practical situations.
- Tools are used to automate the testing and analyse the coverage of the tests.

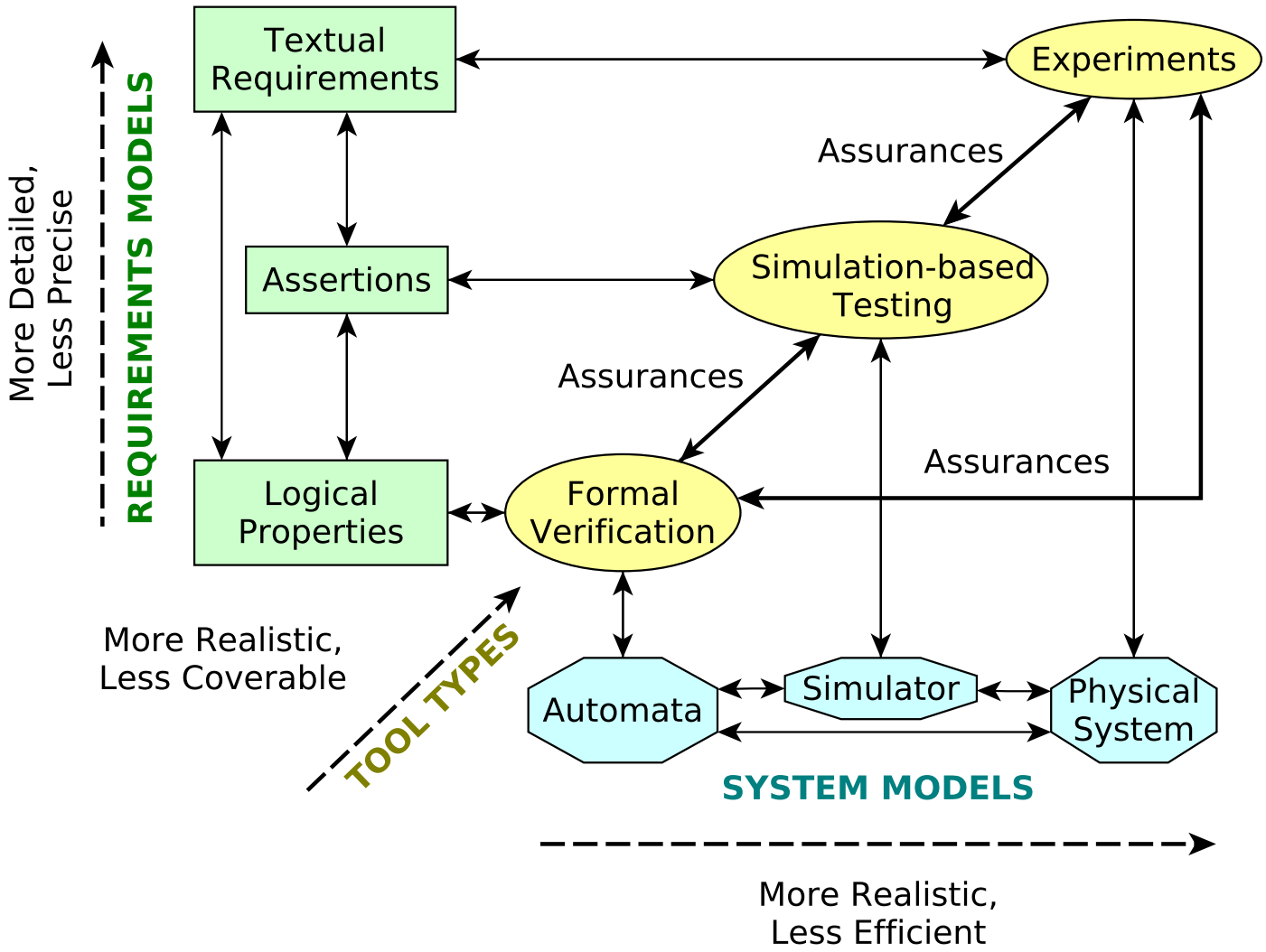




End User Validation

- This approach involves experiments and user evaluations in practical robotic scenarios.
- Experiments take place in the UoH *Robot House* a domestic house equipped with equipment (sensors, video etc) to monitor the house and the interactions.
- Scenarios relating to robot human interaction are developed to test some hypothesis and experiments with users carried out.
- This helps establish whether the human participants indeed view the robotic assistants as safe and trustworthy.

Overall Approach



The Role of Trust in Human-Robot Interaction

- Robotic assistants need to be trusted (to some extent) or it is unlikely that they will be bought and used.
- We need to be able to demonstrate the safety and reliability of robotic assistants (at least to help with certification).
- Trust is about the perception of, not just the actual safety and reliability.
- However, very safe and reliable robots may not be trusted because of human factors.
- With the verification aspects we aim to verify the decision making of the robot not lower level robot control.
- An explanation of the decisions made by the robot might help particularly when these don't match what the user expected.

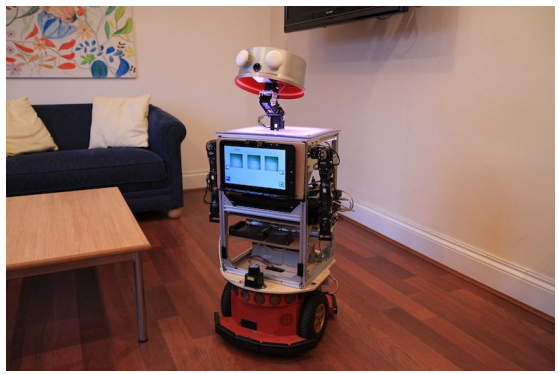
Experiments with Trust and Reliability

In the robot house UoH experimented using two scenarios where the robot appeared faulty or not.

In both scenarios the person was asked to carry out a task with the robot.

Results suggested that although errors in a robot's behavior are likely to affect participant's perception of its reliability and trustworthiness, this doesn't seem to influence their decisions to comply with instructions (or not).

Their willingness to comply with the robot's instructions seem to depend on the nature if the task, in particular, whether its effects are irrevocable.



Social Trust in Autonomous Systems

Social trust is a belief in the honesty, integrity and reliability of others–"a faith in people."[2]

- Media stories about "killer robots" don't help–we need more positive stories.
- If there are high benefits in using robot assistants (or other autonomous systems) it might encourage potential users to try them and build up trust.
- There is a need to engage with regulators to ensure there is a suitable regulatory framework to protect the users but also allow autonomous systems to come to market.
- With respect to safety they need to be at least as good as humans but how can this be demonstrated?
- Other issues apart from trust are also important: safety, reliability, legal, ethical, privacy, security, ...

[2] Americans and Social Trust: Who Where and Why, Pew Research Center, Social & Demographic Trends, 2007

Concluding Remarks and Papers

We gave an overview to the project Trustworthy Robot Assistants and discussed approaches to trust and safety for robotic assistants.

Webster, M., Dixon C., Fisher M., Salem M., Saunders J., Koay K.L., Dautenhahn, K and Saez-Pons, J.
Towards Reliable Autonomous Robotic Assistants Through Formal Verification: A Case Study
IEEE Transactions on Human-Machine Systems, 46(2): 186-196, 2016.

Dixon C., Webster, M., Saunders J., Fisher M. and Dautenhahn, K.
The Fridge Door is Open"-Temporal Verification of a Robotic Assistant's Behaviours, Advances Autonomous Robotic Systems (TAROS), Springer, LNAI vol 8717, pages 97-108, 2014. Paper won the Springer Best Paper Award.

Amirabdollahian, F., Dautenhahn, K., Dixon, C., Eder, K., Fisher, M., Koay, K.L., Magid, E., Pipe, A., Salem, S., Saunders J., and Webster, M.
Can You Trust Your Robotic Assistant?
In 5th International Conference in Social Robotics (ICSR 2013), LNAI vol 8239, pages 571-573. Springer, 2013.