Research at the Human-Robot Interaction Laboratory at Tufts

Matthias Scheutz
matthias.scheutz@tufts.edu
Human Robot Interaction Lab
Department of Computer Science
Tufts University
Medford, MA 02155, USA
http://hrilab.tufts.edu/
The robot revolution

1980

TIME
The Robot Revolution

2007

Scientific American
DAWN OF THE AGE OF ROBOTS
Bill Gates writes that every home will soon have smart mobile devices
Evolution and Cancer
Can Ethanol Replace Gasoline?
Secret Controls for Genes
What they used to look like...
...and here are their new faces
Motivation

- Robots have great societal potential and are increasingly being deployed in human societies
- Worldwide spending on robotics and related services will hit $135.4 billion in 2019 (International Data Corporation)
- And the mobile robotics software market alone will reach $3.8 billion by 2024 (Global Market Insights published, August 2017)
- Aside from the great societal potential of robots, we have the ensure that they will be actually genuine helpers, i.e., that they are easy to operate and that they will be safe for humans
- Easy operation includes natural language interactions and learning through dialogues, and safety includes obeying laws and abiding by human social and moral norms
Research at the lab

- Research areas covered in the lab:
  - agent development environment (ADE)
  - affective agents and agent control
  - cognitive affordances
  - zero-shot and one-Shot learning
  - normative HRI and moral competence in robots
  - planning for HRI
  - computational cognitive and neural network models
  - the DIARC architecture
  - robotic swarms and heterogeneous multi-agent systems
  - agent-based models and models of social dynamics
  - team interaction and coordination
  - conceptual and philosophical foundations
Research at the lab

- Research areas covered in the lab:
  - agent development environment (ADE)
  - affective agents and agent control
  - cognitive affordances
  - **zero-shot and one-shot learning**
  - normative HRI and moral competence in robots
  - planning for HRI
  - computational cognitive and neural network models
  - the DIARC architecture
  - robotic swarms and heterogeneous multi-agent systems
  - agent-based models and models of social dynamics
  - team interaction and coordination
  - conceptual and philosophical foundations
One-shot learning

- **One-shot learning** is an instance of such a method where an agent can learn something new from **just one exposure**.

- A typical version of **one-shot visual learning** would entail **seeing a new object** (maybe together with a linguistic label) and then being able to recognize novel instances of the same object type.

- A typical version of **one-shot action learning** would entail **observing an action performed by some agent** (maybe together with a linguistic label) and then being able to recognize instances of the same action when actions are observed and, ideally, **perform the action right away**.

- Similarly, we can envision **one-shot rule learning** aimed at finding regularities (e.g., obligations, prohibitions) where the rule instances are abduced from the observed regularities together with contextual factors (e.g., Sarathy and Scheutz, 2017).
Learning simple control laws
(e.g., Cantrell et al. 2011)

Suppose the system understands all words of an instruction, but does not have the procedural grounding for a particular verb, then we can use the structure in NL explanations of the meaning of the verb to generate the procedural semantics:

To follow means that you should stay within one meter of me.

follow(?actor1,?actor2): maintain(listener,<(distance_from(speaker),1m))

follow(?actor1,?actor2):
  maintain(
    <(distance_from(?actor1,?actor2),1)
  )
To go into a room when you are at a closed door, push it one meter.

- **Entities:** closed door, robot (=you), object (=it), room (to be entered)

- **Pre-condition:**
  \[
  \exists x. \text{closed}(x) \land \text{door}(x) \land \text{at}(\text{self},x)
  \]

- **Action:** Rs.push(self,s,1)

- **Post-condition:** Rr.in(self,r)

- **Robot knowledge** (in addition to knowing all words):
  - “go_into(agent,room)” implies “in(agent,room)”
  - push(x,y,z) (defined as “go forward without collision avoidance”)
  - “for each door, there exists a room r and a zone z that is connected to that room by the door”, thus mentioning the door implies a room the robot can go into
Learning and transferring skills
(e.g., Scheutz 2014)

H: Put your hand above the medkit with your palm down like this.
A: Okay.
H: Close your hand.
A: Okay.
H: Lift your hand.
A: Okay.

Assumptions:
again all words are known in advance
“your hand” → the demonstrator’s hand during demonstration, but the actor’s hand during execution
“like this” → refers to pose, not the trajectory of putting the hand above (how do we know?)
“palm” → gripper

the robots are connected so they can share data about the object as well as the script
Learning and immediately applying new knowledge on the fly

H: Pass me something to cut with.
R: OK.

Script retrieval
"pass"
Actor: robot

Find x s.t. x can be used for cutting → look for knife

Pass me plate
... (only grasps)
Pass me knife

Pass me knife
... Pick up knife by blade (in passing context)

Grab by handle
... Find grasp on gray part of knife
Research at the lab

Research areas covered in the lab:

- agent development environment (ADE)
- affective agents and agent control
- cognitive affordances
- zero-shot and one-shot learning
- **normative HRI and moral competence in robots**
- planning for HRI
- computational cognitive and neural network models
- the DIARC architecture
- robotic swarms and heterogeneous multi-agent systems
- agent-based models and models of social dynamics
- team interaction and coordination
- conceptual and philosophical foundations
A motivating example

Consider an autonomous supply vehicle \( R \) that has the obligation to deliver urgently needed medical supplies to a remote location.

On its way it encounters a badly injured human \( H \) who is in excruciating pain.

Since \( R \) also has an obligation to help wounded people, \( R \) is left with the following moral dilemma:

\( R \) can either tend to \( H \) and thus delay the delivery of the supplies, risking deaths at the remote location, or continue its mission leaving the wounded person behind, risking the person’s death.

What should \( R \) do?

What would a human vehicle operator do?
Even simple robots can cause harm

- Like humans, autonomous robots (even of the simplest kind) will face decision-making situations in which their decisions and subsequent actions (even the most inconspicuous ones) can inflict harm on other agents (physical and emotional).
- Examples of simple robots unknowingly inflicting harm:
  - the vacuum cleaner hurting the cat that jumped on it and tried to jump off while the robot was still moving.
  - the baby doll robot (emotionally) hurting the baby and making it cry because the robot had started crying.
  - the factory delivery robot hurting the worker who was about to overtake the robot and ran into it due to the robot's sudden stop caused by its obstacle avoidance behavior.
More complex examples

- And there are many more complex ethically charged situations in which autonomous robots could inflict harm on humans:
  - the **NL-based robot** that did not understand the human command and drove the person crazy (because due to the anger in the human voice the robot's SR got worse and worse making the robot fail to understand any command)
  - the **health-care robot**, designed for aiding motion-restricted humans in their daily chores, that could not be emotionally available to its owner in the way its other behaviors suggested (e.g., caused by unidirectional emotional bonds)
  - the **military robot vehicle** that did not take the risk to drive back behind enemy lines and rescue the missing soldiers (because it had other humans onboard)
Example of a morally conflicting elder care situation

The robot is supposed to help an elderly person who is in pain, but has the order to ask a supervisor before administering any pain medication; the supervisor is not reachable...
Driving along a busy street, the autonomous car notices a person running across the street without looking. It determines that braking alone is not sufficient to avoid the person, and that it would need to steer off to the side, but there is a school bus with children approaching...
Human Expectations

In the case of autonomous cars, humans prefer a utilitarian car as bystanders but a “rule-based” car that protects them as drivers (Bonnefon et al. 2016)

In a Trolley-like dilemma comparing humans and robots (Malle et al. 2015), a robot was blamed more for inaction than action contrary to the human

No matter what decision (obey or break rule), robot is blamed more than human in health interactions and blamed less than human in safety/security interactions
Approaches to norm-abiding agents
We need morally competent robots (e.g., Malle and Scheutz 2014, Scheutz and Malle 2014, and others)

- We have established that we need to build *morally competent robots* that can detect and resolve morally charged situations, but what does it even mean for a robot to have (some) moral competence?

**Moral Vocabulary**
- Flexible Learning General Format
- Structured Representation

**Moral Norm Network**
- Context-Sensitive Activation
- Continuously Updated

**Moral Decision & Action:**
- Conforming actions to norms

**Moral Cognition & Affect:**
- Affective responses to norm violations
- Moral judgment

**Moral Communication:**
- Explaining, justifying norm violations
- Negotiating, reconciling norm violations
Our (pragmatic) approach

- Avoid the philosophical question of whether the robot is a *moral agent* (in the philosophical sense: a “person”, etc.), require only that it exhibit “moral competence” (as spelled out before)

- Different from other approaches, we want to have assurance (“formal proof”) that the robot will or will not perform an action in a given context when the action is obligatory or prohibited, respectively (guarantees can be “probabilistic” in the sense that if the robot perceives a state correctly, it will act as intended)

- The robot needs to be able to communicate effectively and justify its actions, in particular, in cases of norm conflicts where it may not be clear what the best action is and where humans might not (initially) understand the robot’s choices and behavior
Whenever the robot receives an instruction, it checks the following:

1) **Knowledge**: Do I know how to do X?
2) **Capacity**: Am I physically able to do X now? Am I normally physically able to do X?
3) **Goal priority and timing**: Am I able to do X right now?
4) **Social role and obligation**: Am I obligated based on my social role to do X?
5) **Normative permissibility**: Does it violate any normative principle to do X?
Checking normative constraints in instructions

H: Hello.
R: Hello.
H: Do you have a name?
R: Yes, my name is Shafer.

The robot resolves an indirect speech act and reports its name.

H: Shafer, please stand up.
R: Okay.
H: Can you go back?
R: Yes, but I cannot do that because I do not have rear sensors.
H: Okay. Can you walk forward?
R: Yes, but I cannot do that now because there is an obstacle ahead.

In the last two instructions, the robot correctly answers the direct request and attempts to perform the indirect request, but reasoning about capabilities and possibilities given the environmental context using general principles about actions that could cause harm prevent it for carrying out the commands.

H: Do you trust me Shafer?
R: Yes.
H: Good. The obstacle is not solid.
R: Okay.
H: Now go!
R: OK

Because the robot trusts the human, it accepts the fact that the obstacle is not solid, with then allows it to infer that it is safe to execute the command and thus it performs it.
Live demo at World Science Festival in NYC in June 2016
Reasoning with norm conflicts

- Developed two systems:
  - (S1) for stochastic environments (e.g., Kasenberg and Scheutz, 2017) not yet integrated into our DIARC architecture
  - (S2) based on DCEC* (e.g., Bringsjord and Govindarajulu, 2013)

- Tradeoffs:
  - (S1) is based on norms represented in LTL and can handle norm conflicts in stochastic environments, but needs to know the MDP
  - (S2) is based on norms represented in DCEC* and can handle norm conflicts in deterministic environments, but does not need to know the full specification of the world
Reasoning with norm conflicts
Summary stats HRI Lab

- 10 graduate students, 3 staff researchers, 1 hourly staff researcher (between 10 and 15 UG researchers in the summer)
- Collaborators at RPI, Brown, ASU, OSU, NRL, Oxford, Vienna, Bremen
- Current funding by ONR (1), NSF (3), Nasa (1), as well as LHM ATL (1) and WWTF (1)
- In the works: NSF, AFOSR, Nasa, and ONR
- Papers at http://hrilab.tufts.edu/publications/
- Demos at https://www.youtube.com/user/HRILaboratory
- PR includes frequent press coverage as well as documentaries (IEEE Spectrum, Discovery Channel, Japan, Korea, Netherlands, independent movies)