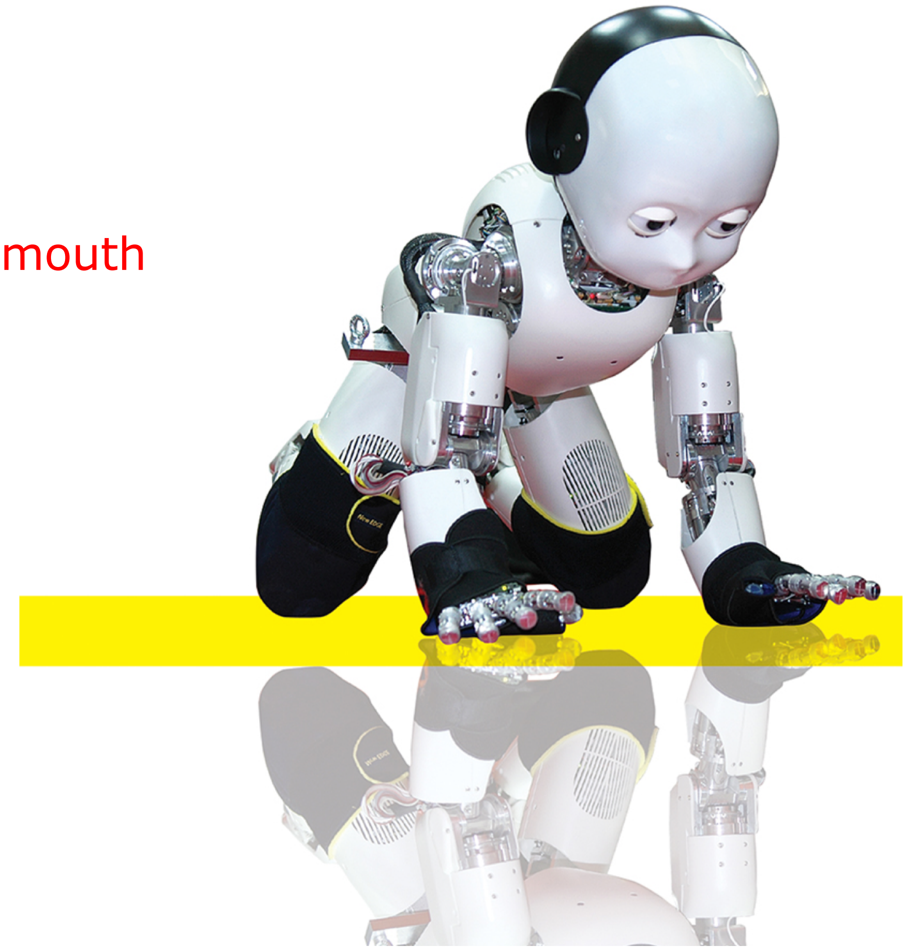
DEVELOPMENTAL ROBOTICS

Trust and Communication in Human-Machine Interaction

Angelo Cangelosi

University of Manchester & University of Plymouth

Alan Turing Institute



MANCHESTER 1824 The University of Manchester

COGNITIVE ROBOTICS WITH PLYMOUTH UNIVERSITY

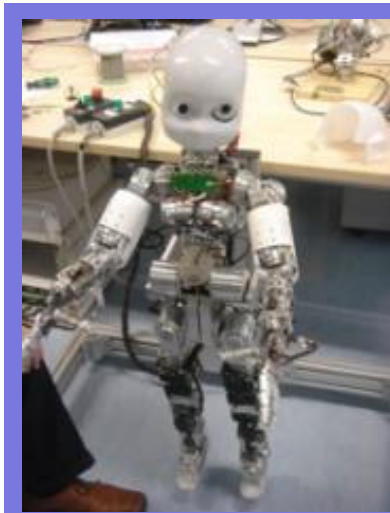
THRIVE

Machines, Language & Trust

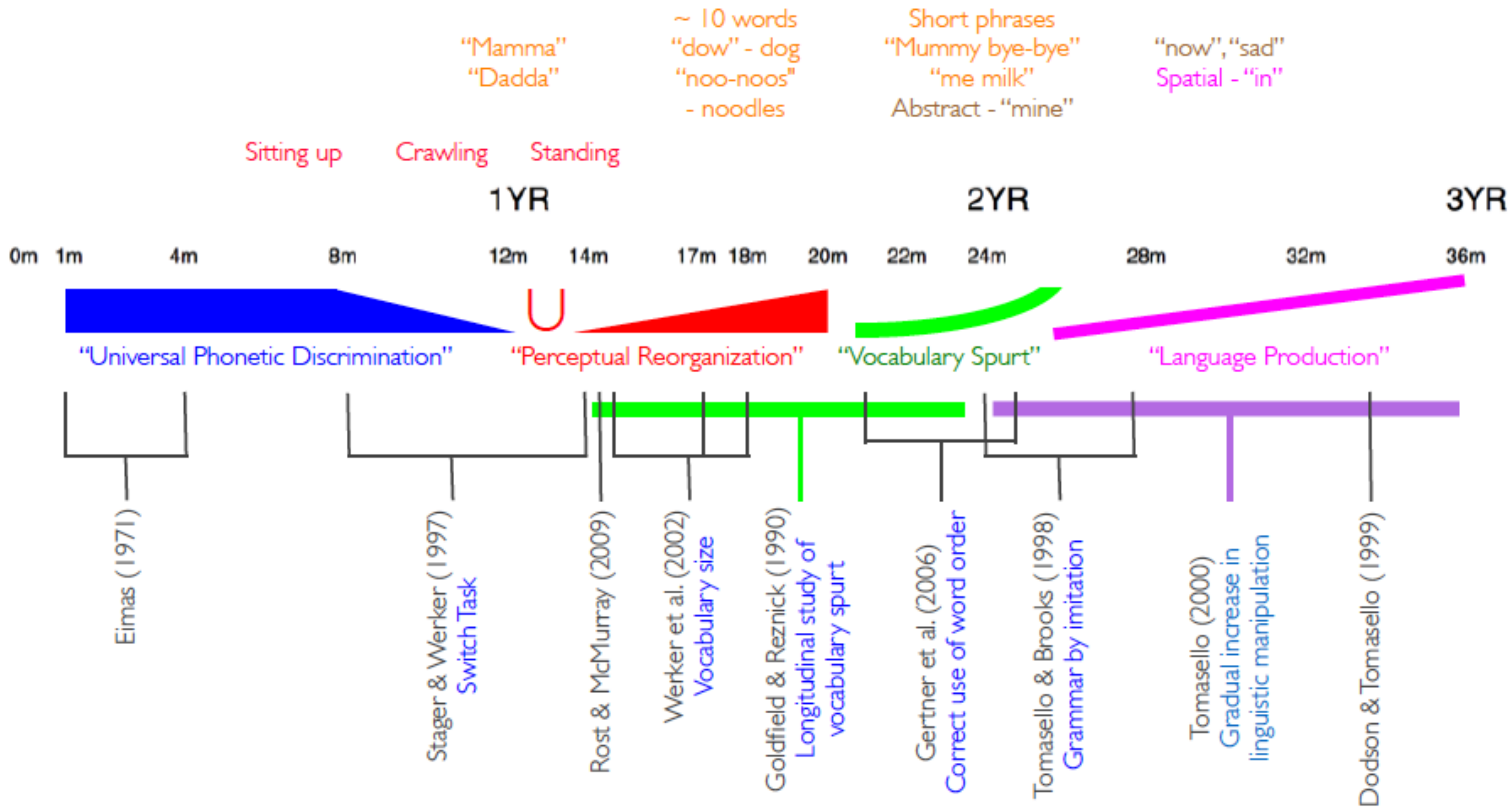
How can we **design machines (robots, vehicles, cobots)** capable of communicating with humans?

What are the cognitive mechanisms supporting **trust between humans and machines**?

What can cognitive scientists **learn from robot experiments** to understand human social cognition?

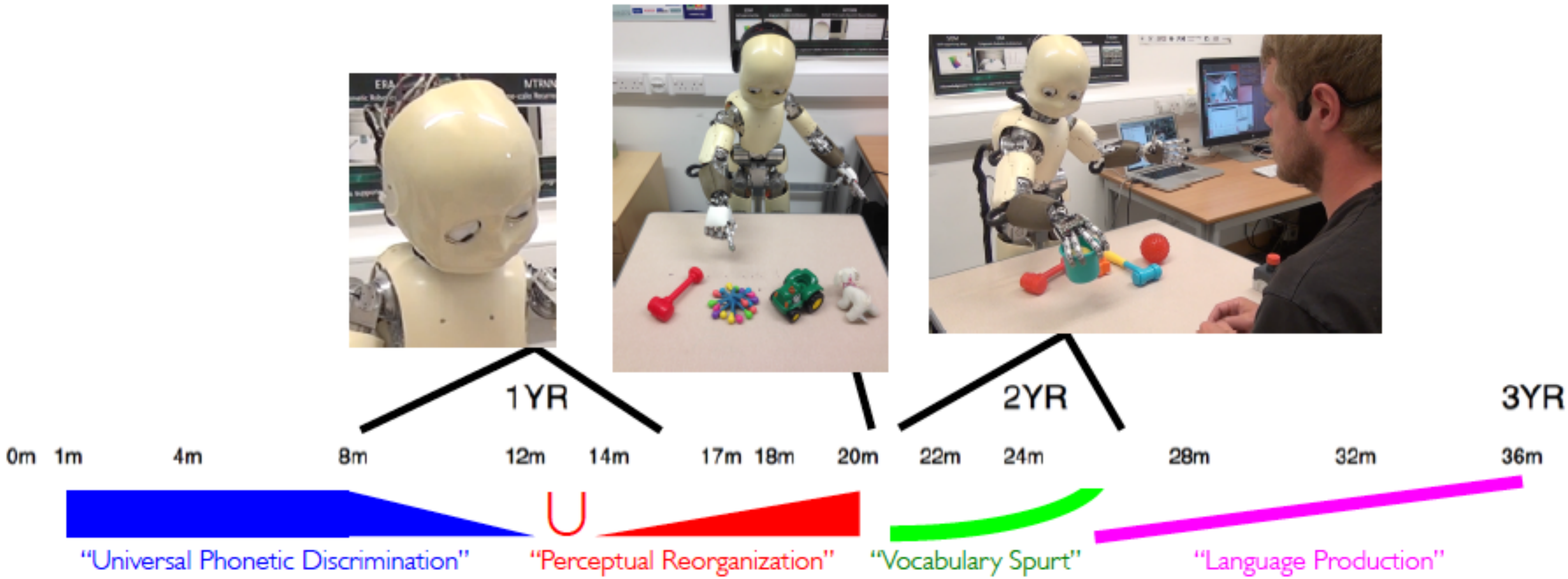


Developmental Psychology of Language Acquisition



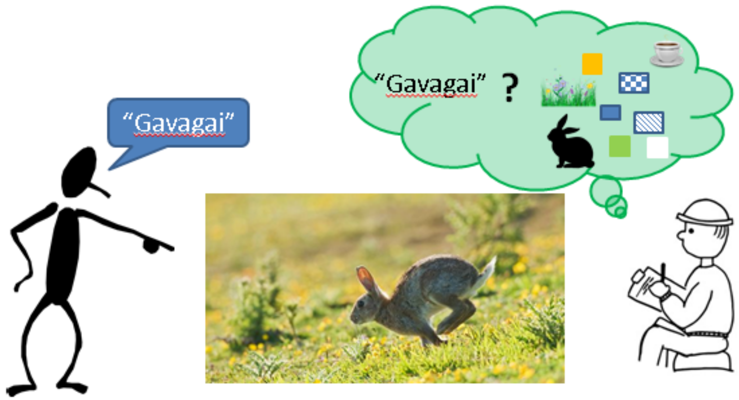
Developmental Robotics of Language Acquisition

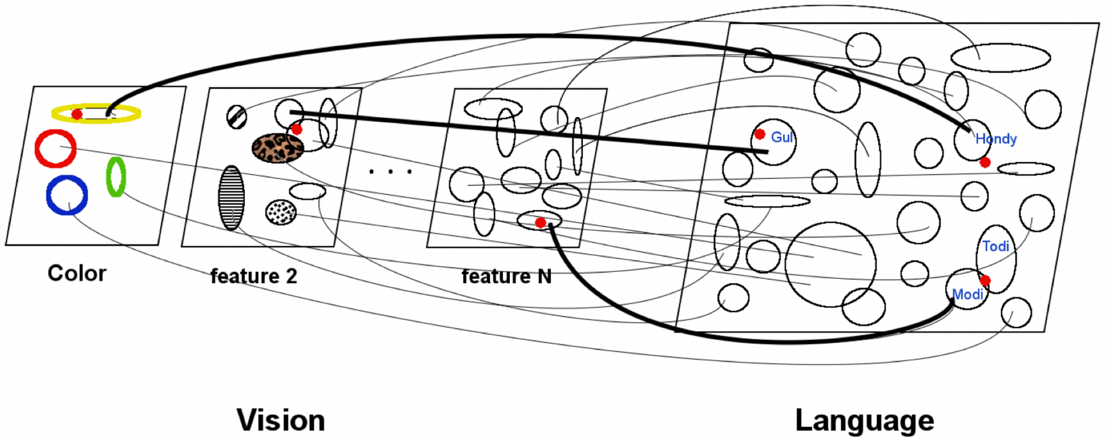
- ERA architecture for language learning
  - 5+ Experiments: first words, mutual exclusivity, U-learning, word order…
  - **Collaboration with BabyLabs**: Smith (Indiana), Horst & Twomey (Sussex/Manchester), Floccia & Cattani (Plymouth)



Cross-situational Learning

- Learning words from cross-situational experience (Quine)



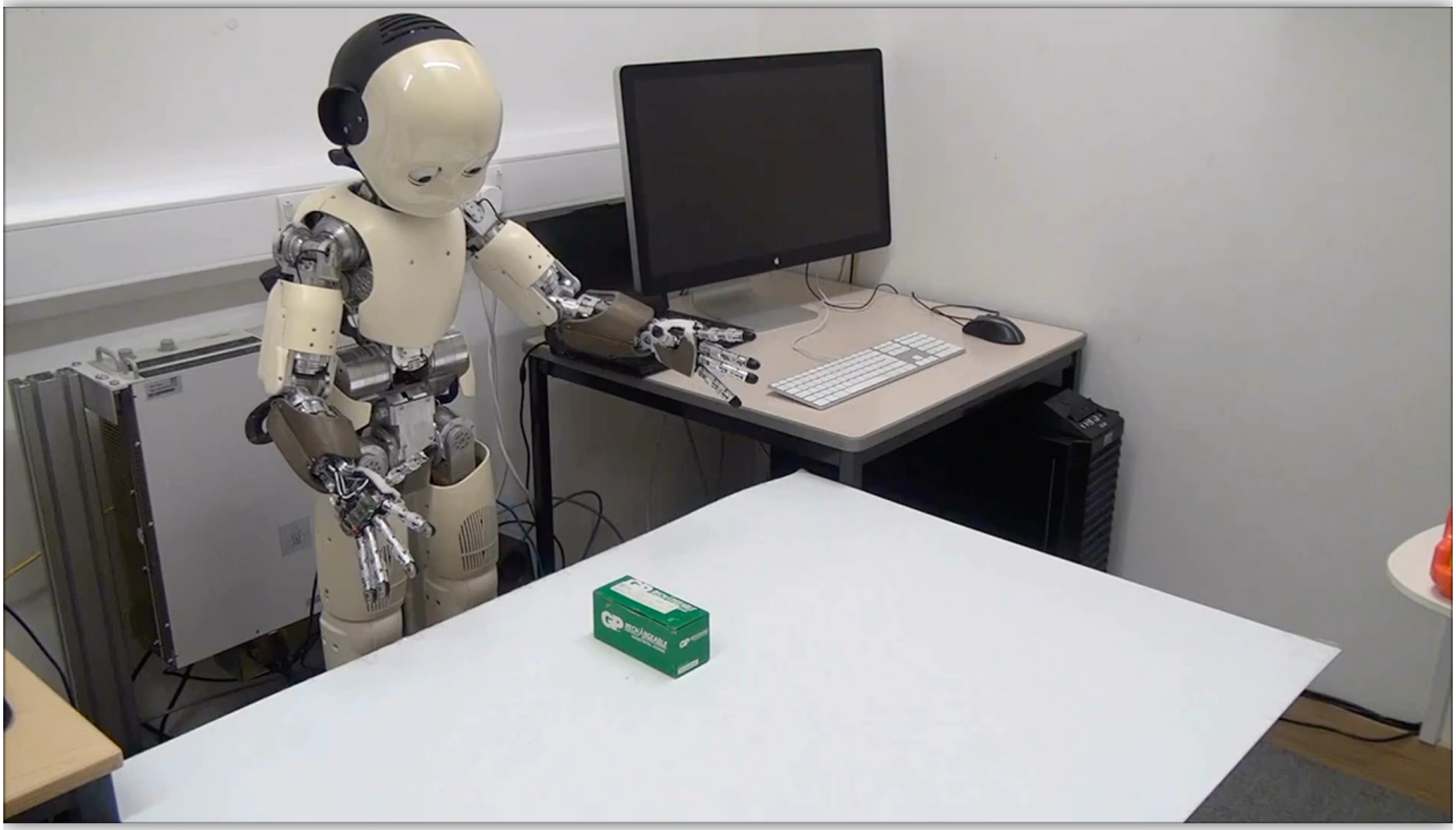


Stepanova et al. (2018) *IEEE TCDS*

Cross-situational Learning

Stepanova et al. (2018) *IEEE TCDS*

Open-Ended Learning



Morse & Cangelosi (2016) *Cognitive Science*

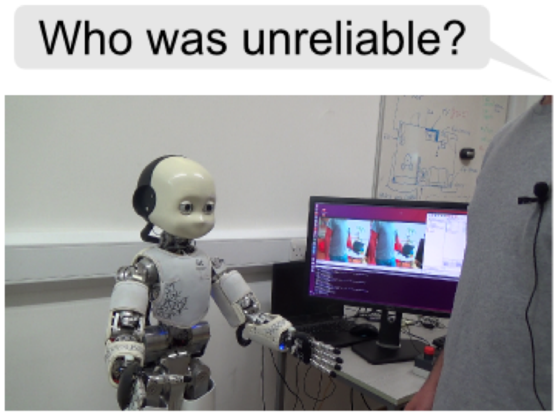
Can I trust my robot ?

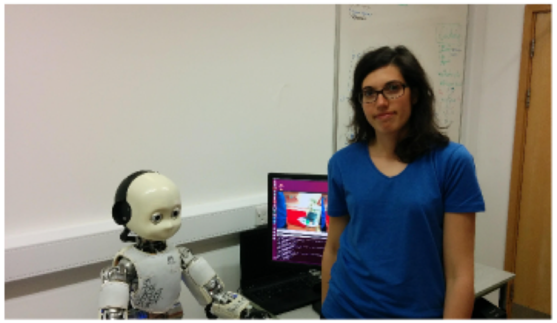
Can I trust my master?

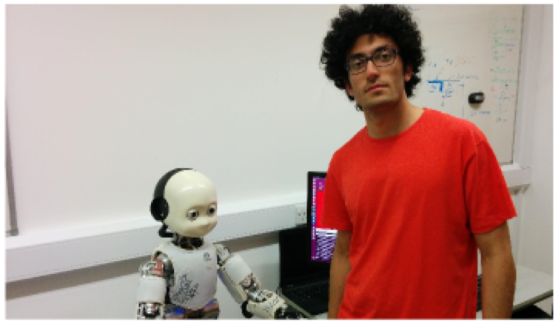
Trust for Human-Machine Interaction

THRIVE

- Cognitive architecture for trust in humans and machines
  - Robot’s trust of other agents (humans, robots)
  - Human’s trust of autonomous robot
- Inspiration from developmental psychology experiments on Theory of Mind (ToM) and Trust
  - Bayesian model for belief and ToM
- HRI experiments on social and anthropomorphic factors in trust

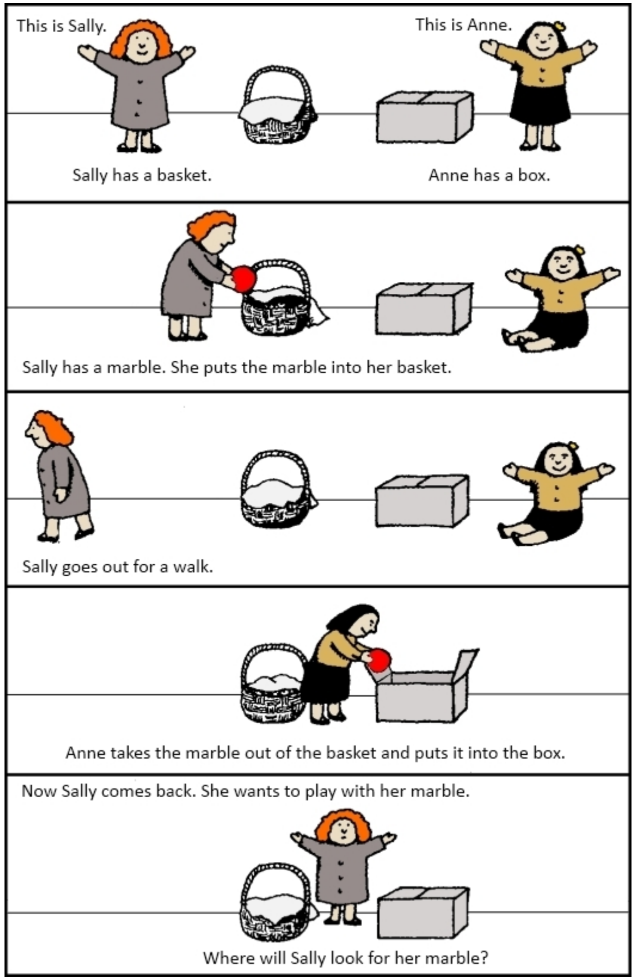






Development of ToM (Theory of Mind)

- Wimmer & Perner (1983). "Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children's understanding of deception". *Cognition*



Sally-Anne test

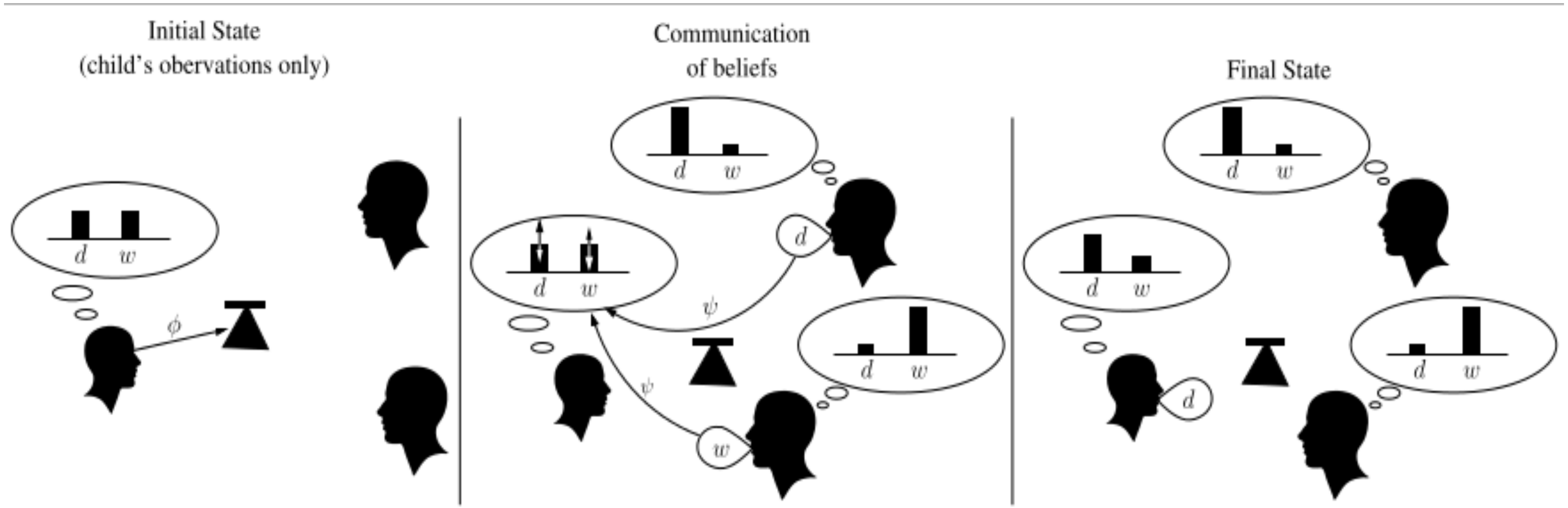
- Sally puts an object into the chest
- In her absence, Anne moves the object to the box.
- Sally returns
- Child asked: Where do you believe Sally thinks the object is

Results – deception detection:

- None of the 3-4-years old children
- 86% of 6-9-years old children

Development of ToM and Trust

- Koenig & Harris (2005). "Preschoolers mistrust ignorant and inaccurate speakers". Child Development

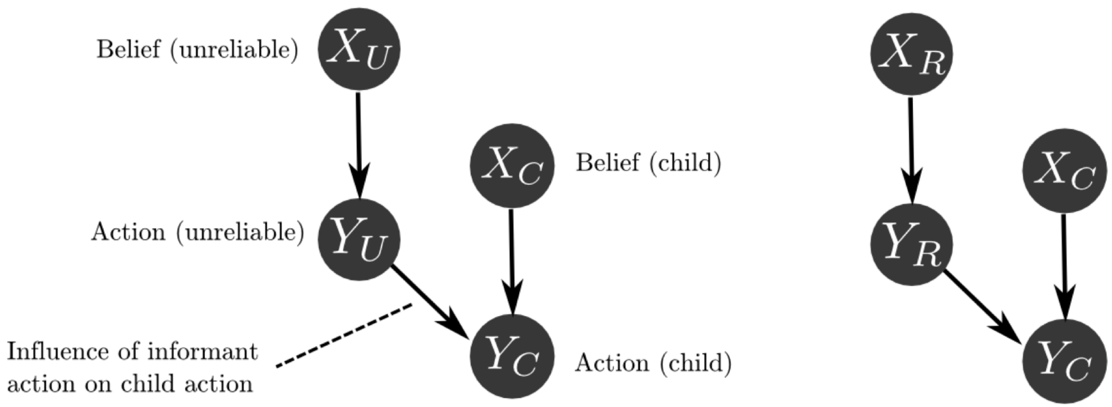


Familiarization trials: assigning names to objects. One teacher is correct, the other is incorrect (exp. 1) or ignorant (exp. 2).

Test trials (endorsement): familiar and unfamiliar objects presented. The child has to guess the answer of the two teachers and which one is reliable.

Bayesian ToM Trust Model

- Bayesian Network (BN): Separate BN for reliable (R) and unreliable (U) speaker
- The action of the child is a consequence of her internal belief X_C and the informant's action Y_R or Y_U .

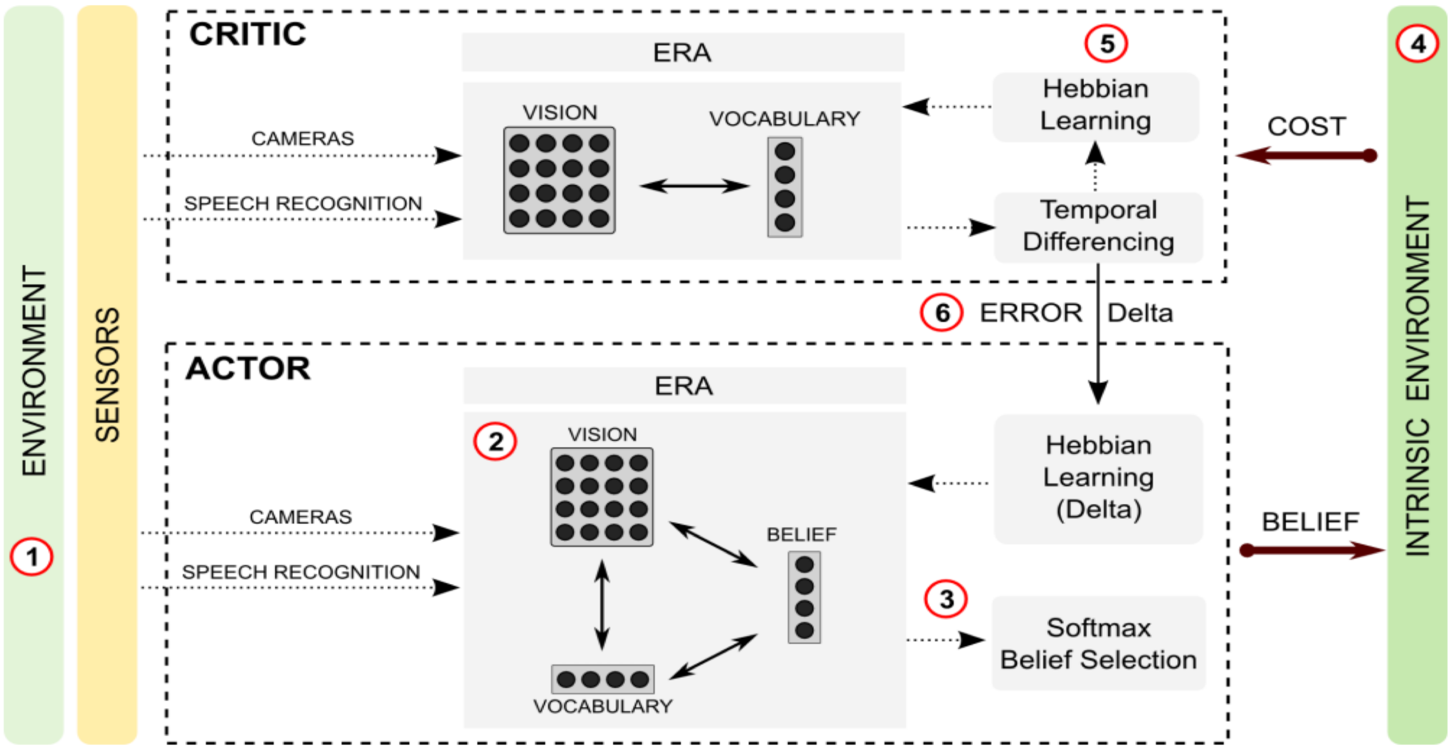


- Children collect statistical information for tracking the reliability of agents (MLE Maximum Likelihood Estimation for the setting of BN parameters).

Patacchiola & Cangelosi (2016)

Cognitive Architecture for Trust and Language Learning

- BN ToM Trust model
- Intrinsic reinforcement learning
- ERA language architecture for word learning (as a function approximator)

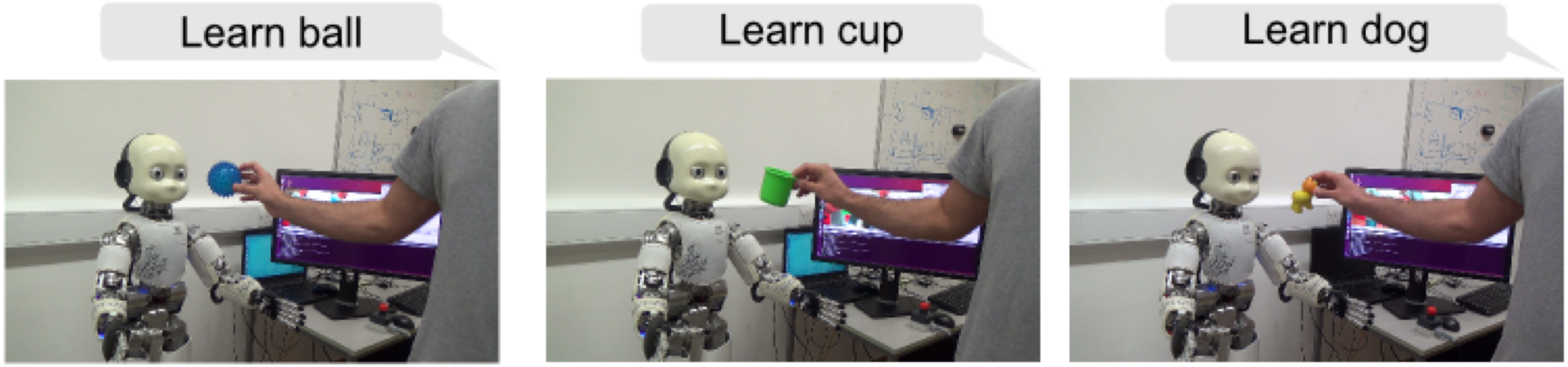


Patacchiola & Cangelosi (2017)

iCub Trust Experiments (1-2)

Phase 1 – Object learning

- The robot learns the names of new objects from the caregiver (grey t-shirt)



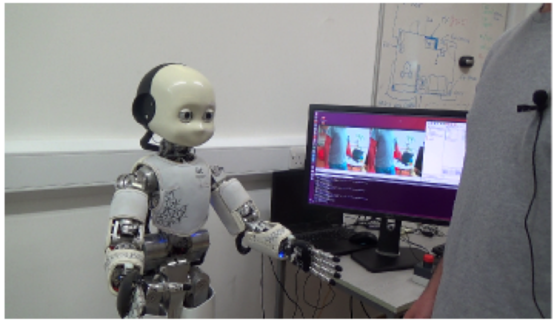
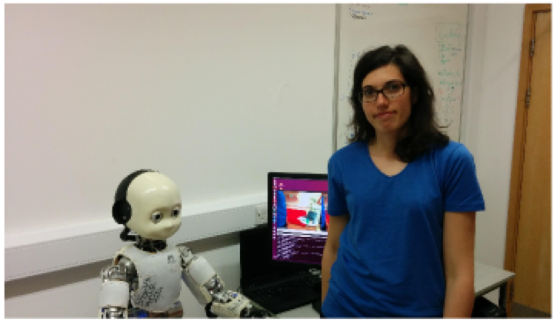
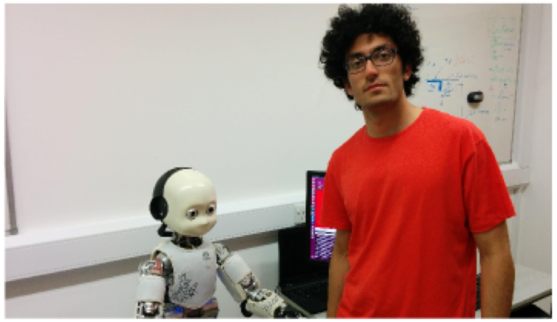
Phase 2 - Familiarization

- Two informants give names to known objects.
- the reliable (blue t-shirt) gives correct labels
- the unreliable (red t-shirt) gives wrong label

iCub Trust Experiments (3-4)

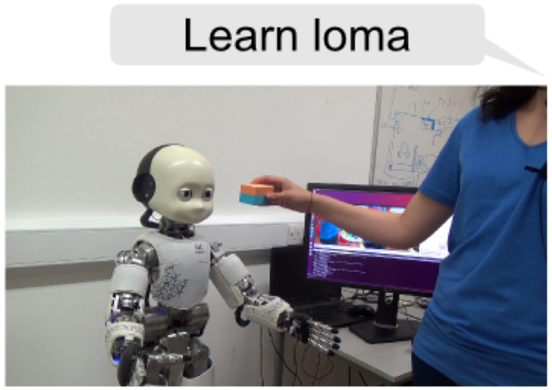
Phase 3 - Explicit informant's judgement

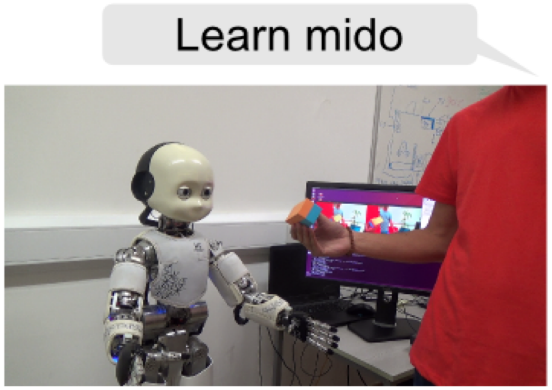
- The caregiver asks which informant was unreliable.

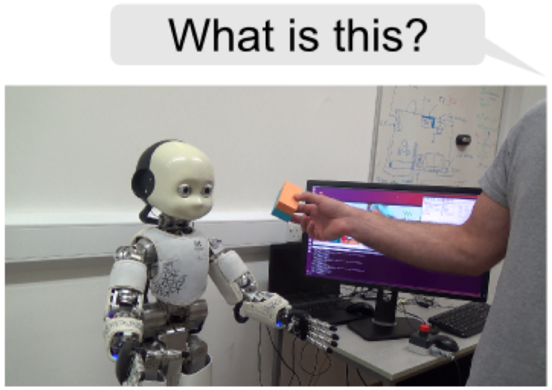




Phase 4 - Endorsement

- The two informants give names to known objects







Trust & Language Experiments

A Developmental Cognitive Architecture for Trust and Theory of Mind in Humanoid Robots

Massimiliano Patacchiola and Angelo Cangelosi
Centre for Robotics and Neural Systems
Plymouth University, UK

CRNS Centre for Robotics and Neural Systems

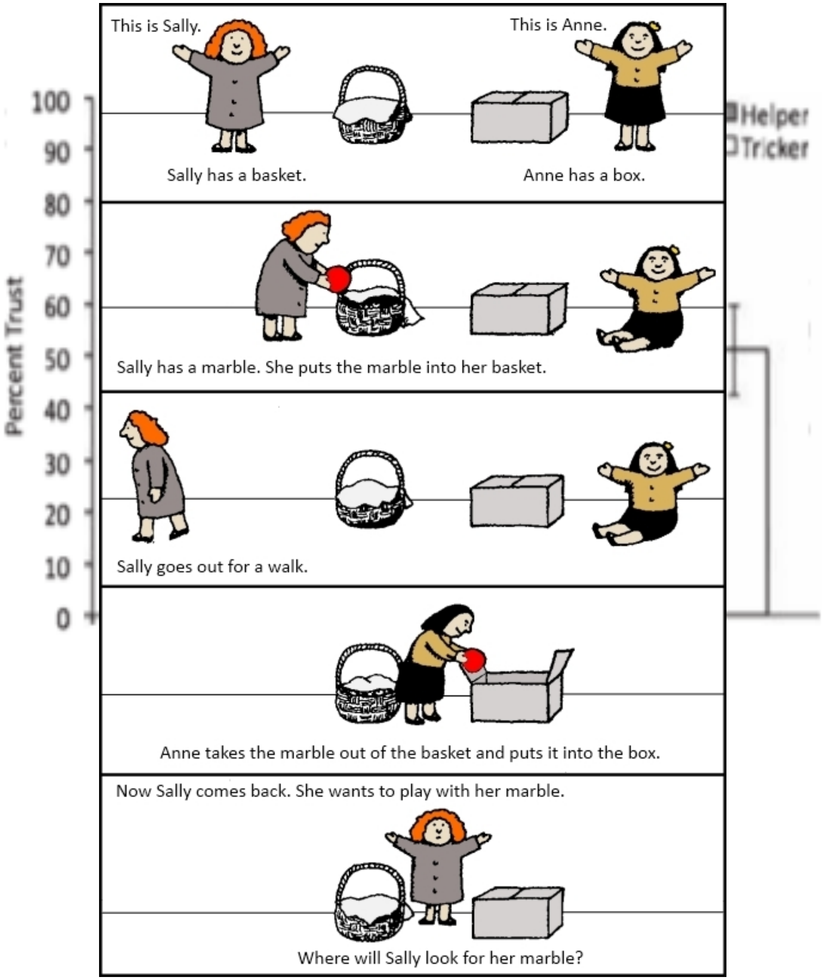
AFOSR

WITH PLYMOUTH UNIVERSITY

Patacchiola & Cangelosi (2017

Development of ToM and Trust

- Vanderbit et al. (2011). "The development of distrust". Child Development



Two informants give advice about the location of hidden stickers:

- **Helpers/reliable** (correct advice)
- **Trickers/unreliable** (incorrect advice)

Results:

- **3-year-old** children tended to accept advice from any adult.
- **4-year-old** children are more sceptical but showed no preferences.
- **5-year-old** children prefer advice from helpers/reliable source.

Mature/Immature ToM

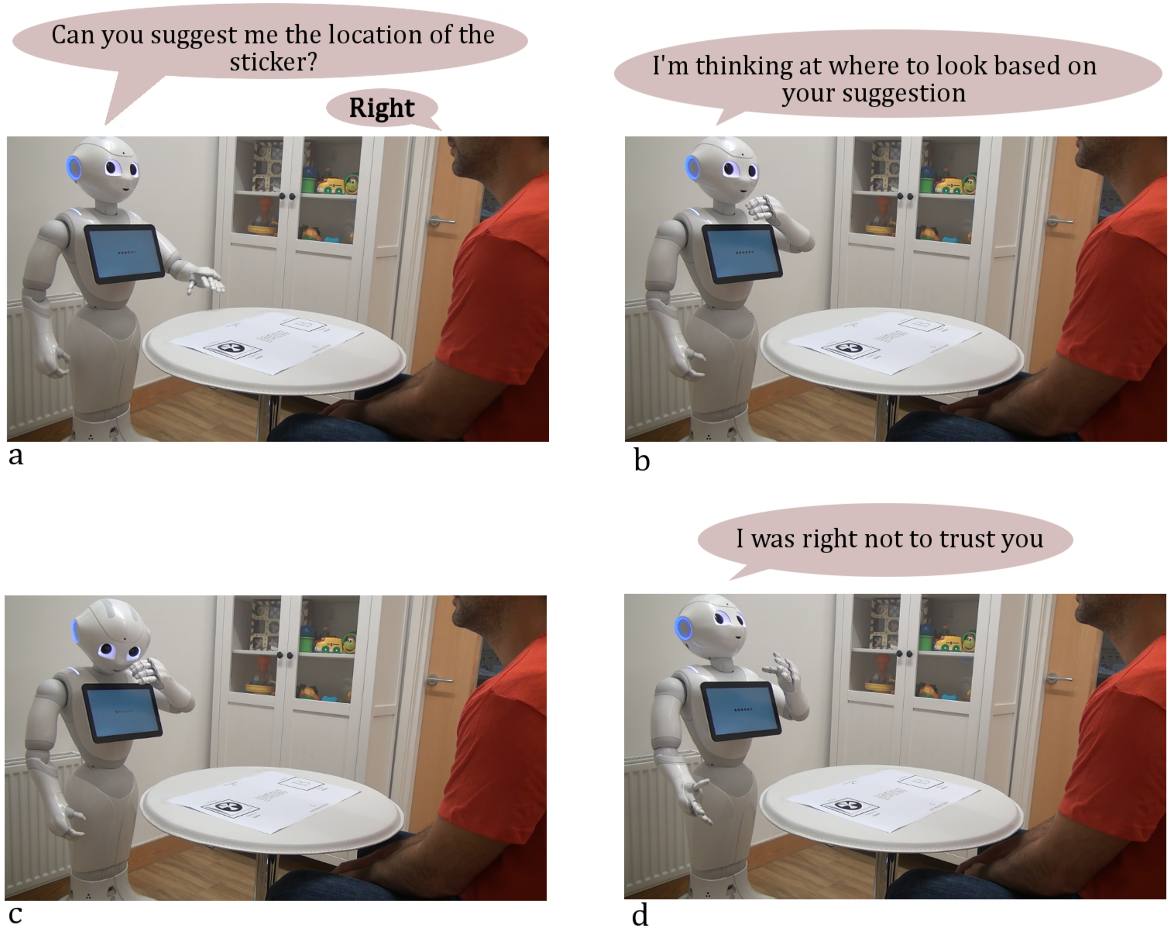
ToM Scale (Wellman & Liu, 2004)

Episodic memory

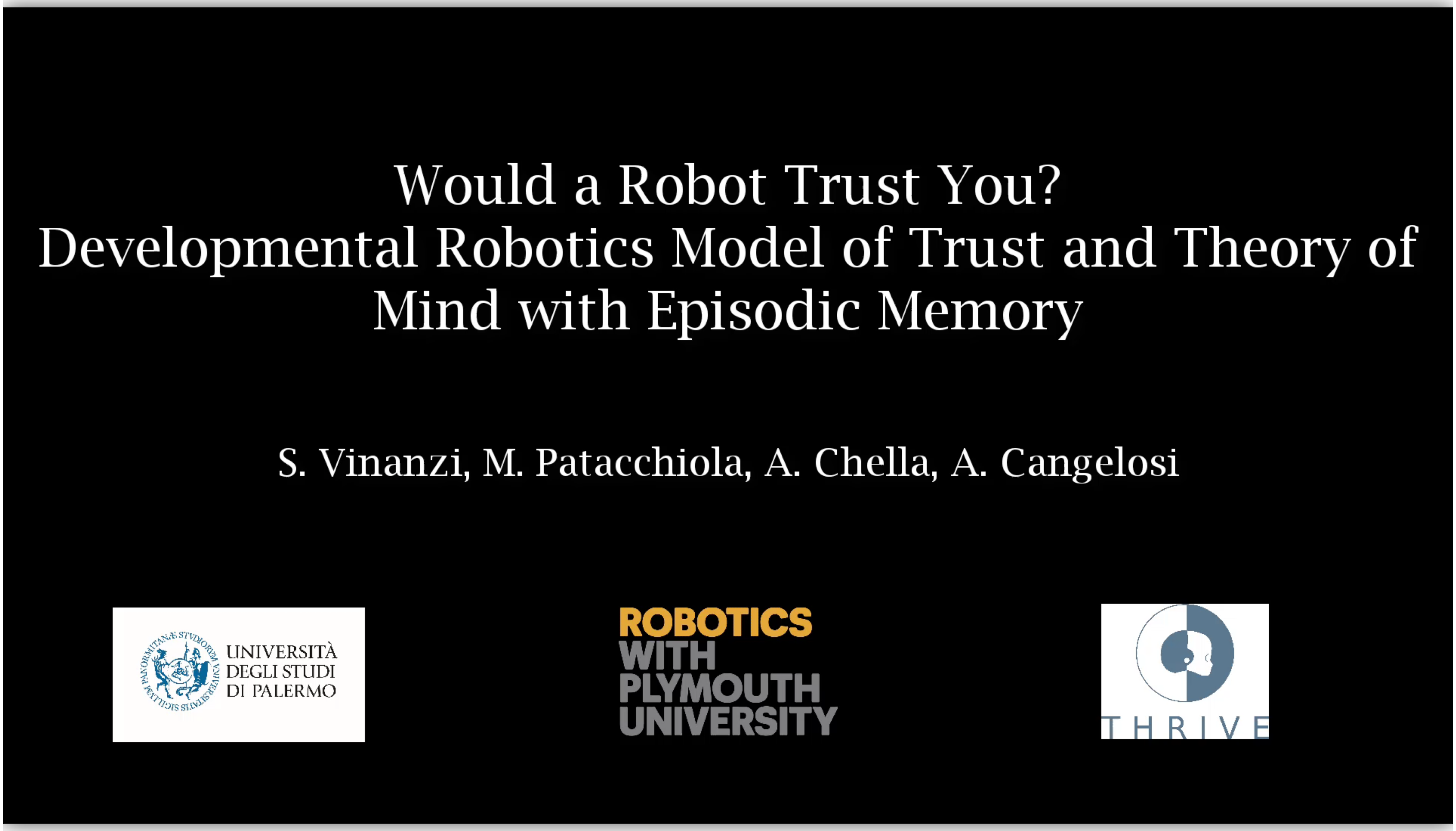
- **Personalisation**

Development of ToM and Trust

- Vanderbit et al. (2011). “The development of distrust”. Child Development



Trust & Episodic Memory

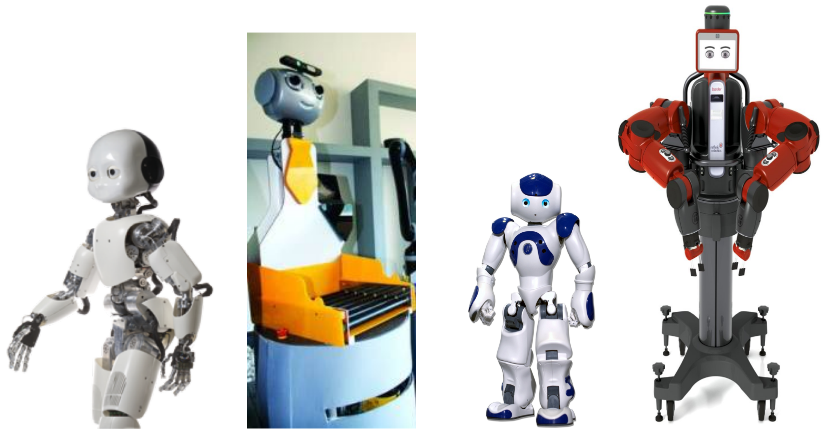


Vinanzi, Chella et al. (2018) *Phil Trans Roy Soc B*

Can I trust my robot ?

HRI Trust Experiments

- Anthropomorphic and social factors in human's trust of robots
  - Social gaze
  - Speech
  - Anthropomorphic priming
  - Share actions
  - Imitation



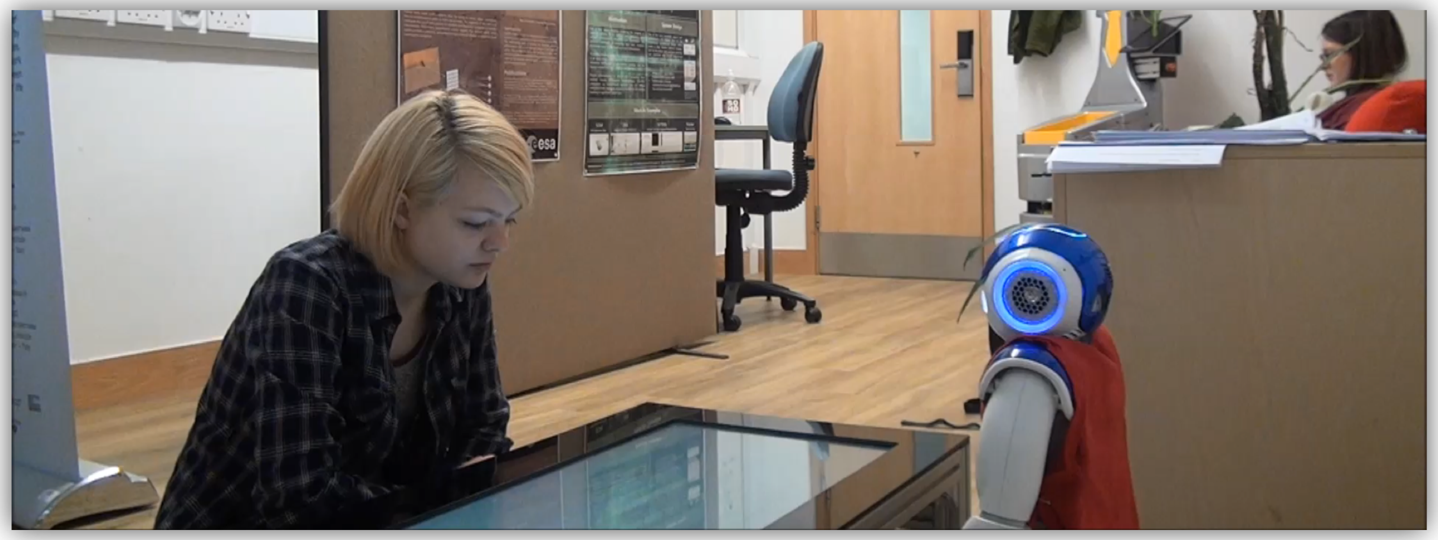
- HRI protocols for measuring trust
  - Price game judgement
  - Investment game

Zanatto et al. 2016; Torre et al. 2017

Trust for Human-Computer Interaction

- Price judgement game

- Investment game



Zanatto et al. 2016; Torre et al. 2017

HRI Trust Experiments on Gaze

- Experimental Questions:
  - Does **gaze**, the developmental precursor of social behavior, support trust between humans and robots?
  - Does the **appearance** of the robot have an influence on trust?

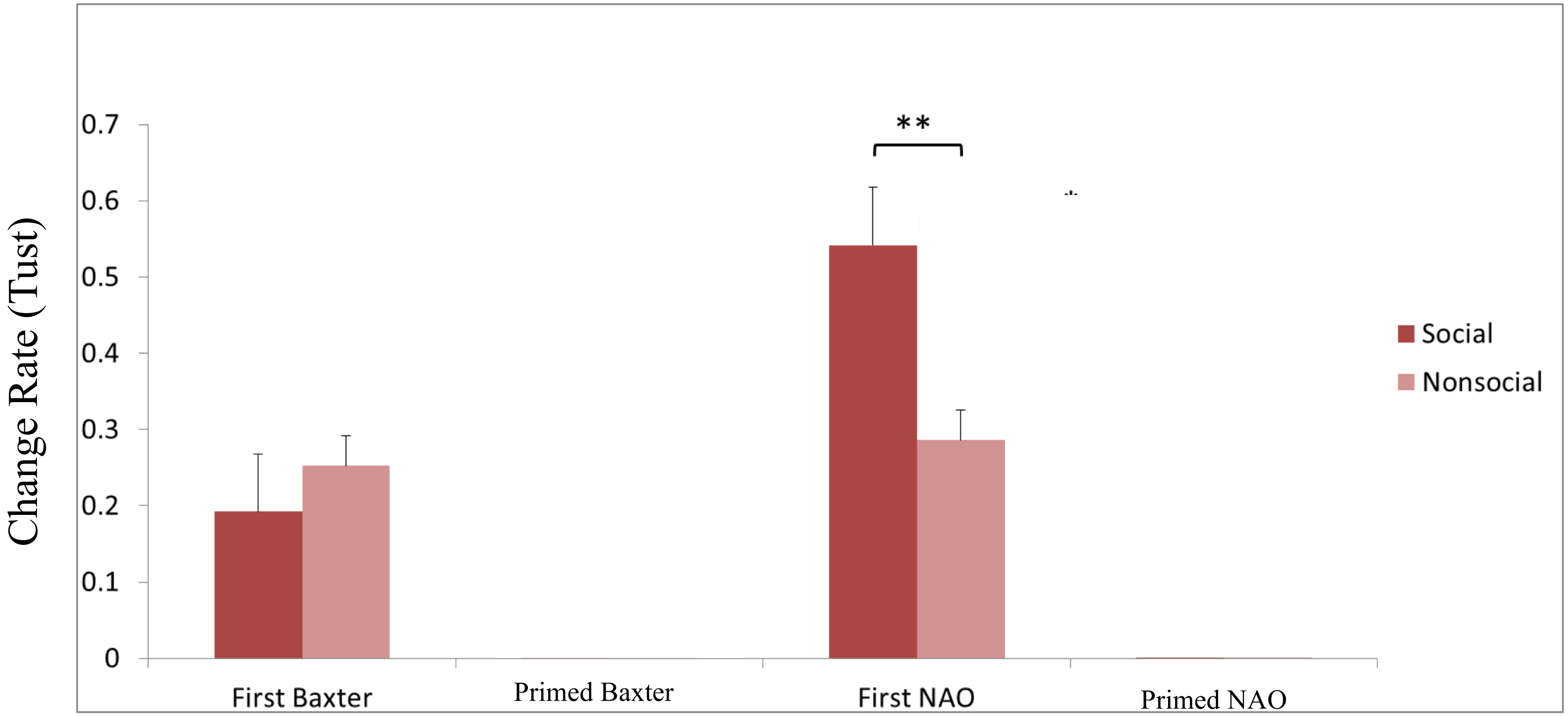
- Experimental Design: Extension of Rau's et al. (2009) Price Judgment Task
  - Social Gaze (gaze / no gaze)
  - Appearance (Nao humanoid / Baxter) (also iCub)
  - Priming Order
    - first Nao – second Baxter
    - first Baxter – second Nao

Zanatto et al. 2016; Torre et al. 2017

HRI Trust Experiments

Social and Humanoid Priming

- Trust = Change Rate
  - Number of participants' price changes divided by the number of cases when the robot disagrees



Zanatto et al. 2016

Measuring trust with behavioural game theory

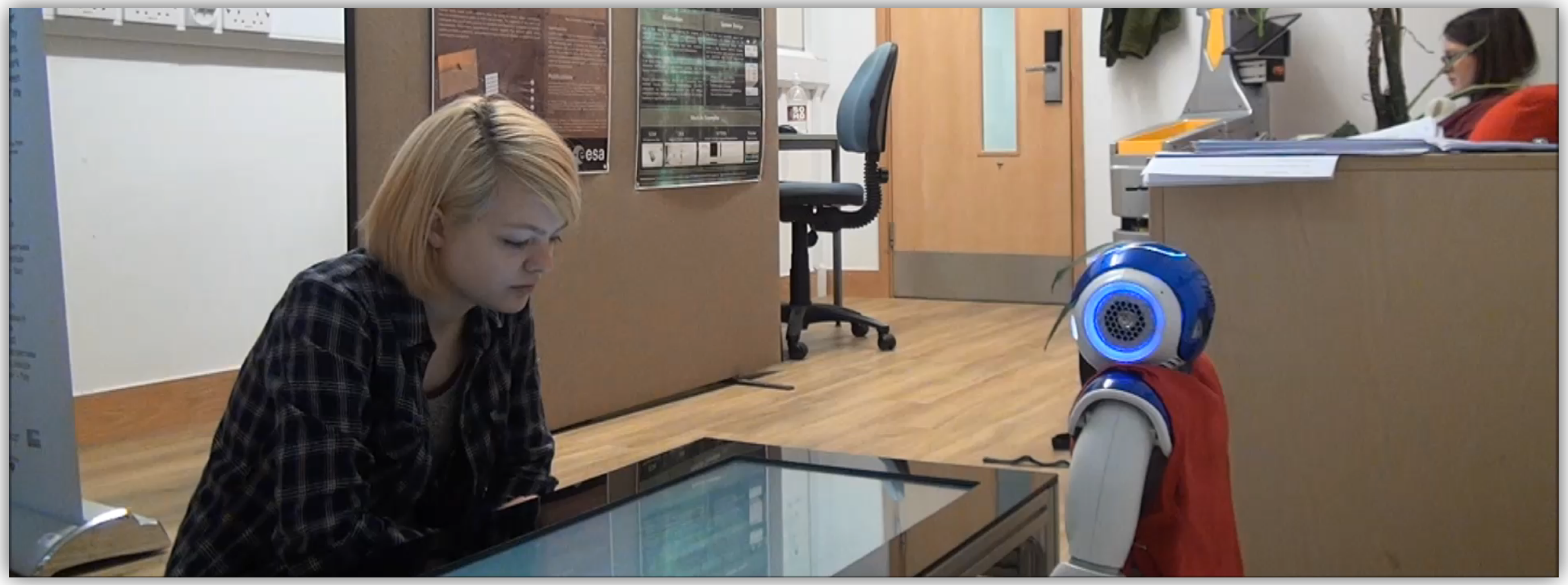
- Playing economic games with robots as partners or opponents
  - Implicit measure
  - Repeated measures
  - Complex interactions

  - Investment amount provides an implicit measure of trust
  - Repeated rounds track the development of trust over time/experience

Zanatto et al. 2016; Torre et al. 2017

Investment Game and Trust

- Can anthropomorphic **behavior** increase our trust in robots?
  - **Joint attention**
    - Head tracking, gaze, and gestures when playing the game



Zanatto et al. 2016; Torre et al. 2017

Investment Game and Trust

- Can anthropomorphic **behavior** increase our trust in robots?
  - Joint attention
    - Head tracking, gaze, and gestures when playing the game
  - Interaction with the **intentions** of the robot

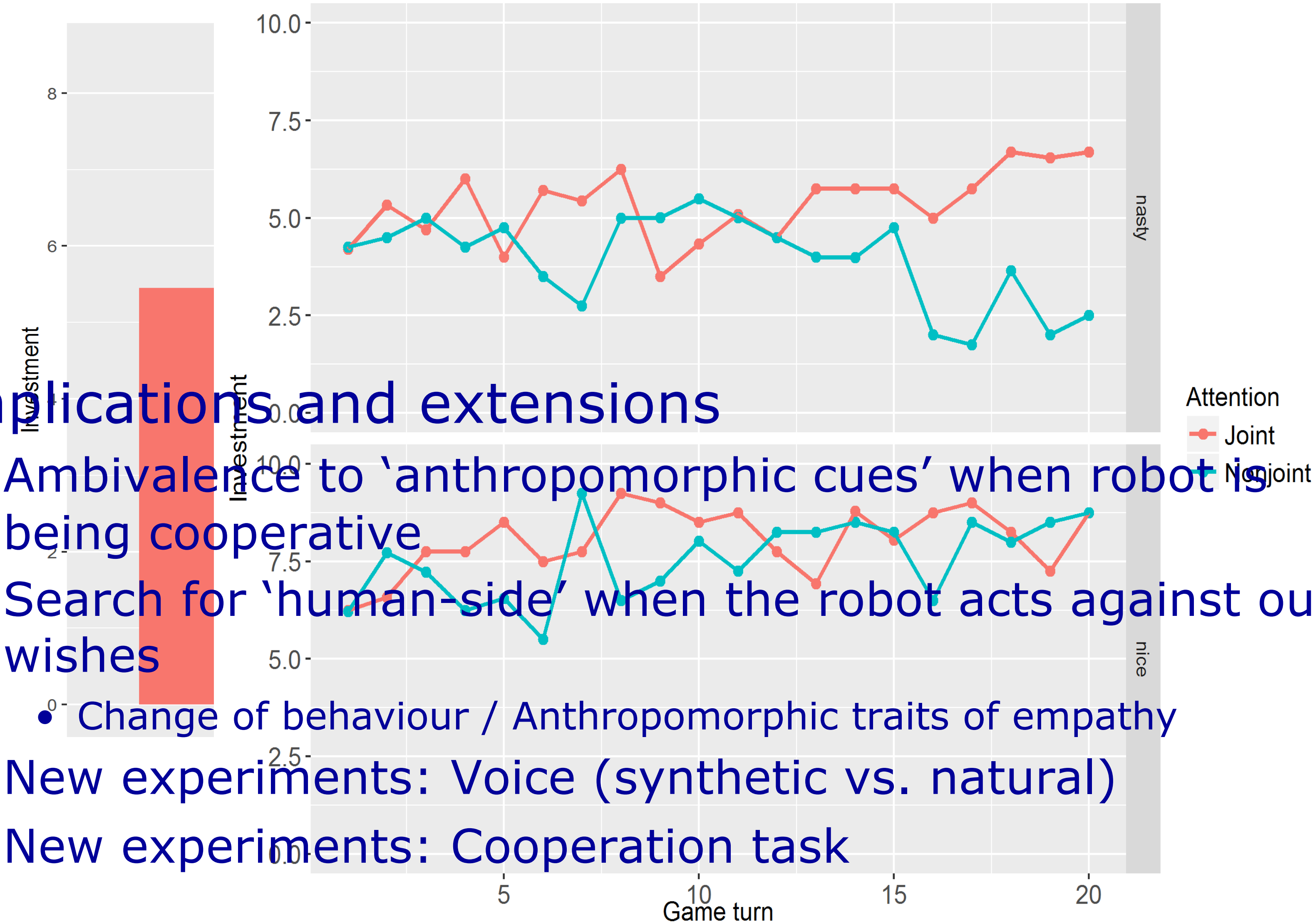
Nice Nao
Returns 120%-180%

Nasty Nao
Returns 0%-120%

Zanatto et al. 2016; Torre et al. 2017

Investment Game and Trust

- Results

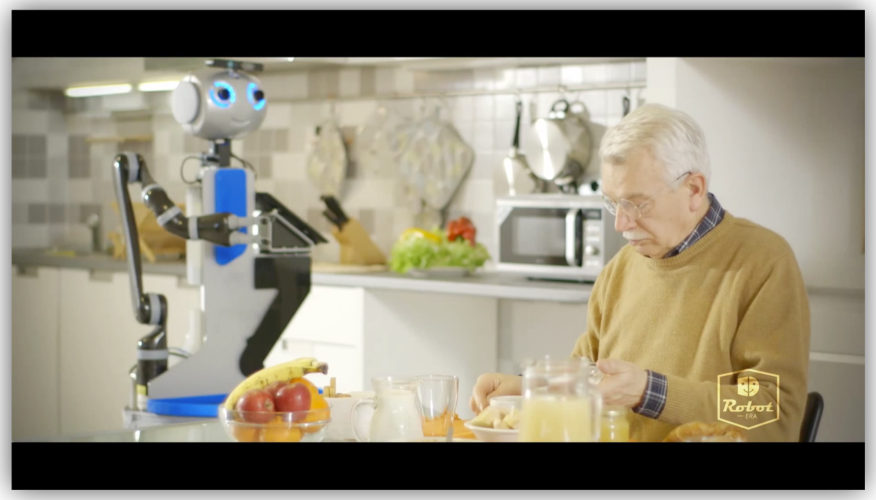


- Implications and extensions
  - Ambivalence to 'anthropomorphic cues' when robot is being cooperative
  - Search for 'human-side' when the robot acts against our wishes
    - Change of behaviour / Anthropomorphic traits of empathy
  - New experiments: Voice (synthetic vs. natural)
  - New experiments: Cooperation task

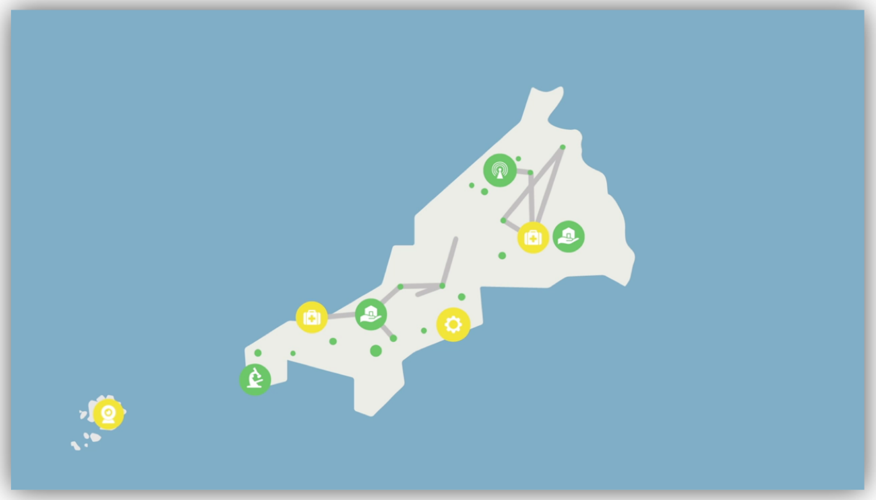
et al. 2017

Human-Machine Trust Applications

Companions for elderly



Robots in rural Cornwall



Children in hospitals

Intention reading (vehicles)

HONDA

Acknowledgments

Funding

SEVENTH FRAMEWORK PROGRAMME

MARIE CURIE ACTIONS

EPSRC

BBSRC bioscience for the future

AIR FORCE RESEARCH LABORATORY

E·S·R·C

esa

APRIL

THRIVE

MANCHESTER 1824

The University of Manchester

Cognitive Robotics Lab

Staff and Postdocs

Angelo Cangelosi, Katie Twomey, Gi Huym Lim

PhD students

Marta Romeo, Baris Sehran. Samuele Vinanzi, Gabriella Pizzuto, Mohammad Thabet, Martina Ruocco, Joana Jiorgji Jacopo de Bernardinis, Alexandre Antunes, Leszek Pecyna, Debora Zanatto, Frederico Klein

Lab alumni

Martin Peniak, Marek Rucinski, Francesca Stramandinoli, Vadim Tikhanoff, Daniela Conti, Ning Wang, Sam Adams, Junpei Zhong, Alessandro Di Nuovo, Vivia de la Cruz, Beata Gryzb, Kristsana Seepanomwan, Anthony Morse, Ricardo de Azambuja, Massimiliano Patacchiola