Walk the Talk!

Exploring (Mis)Alignment of Words and Deeds by Robotic Teammates in a Public Goods Game









Filipa Correia, **Shruti Chandra**, Samuel Mascarenhas, Julien Charles-Nicolas, Justin Gally, Diana Lopes, Fernando P. Santos, Francisco C. Santos, Francisco S. Melo, Ana Paiva

gaips

inesc id lisboa

TÉCNICO LISBOA

UNIVERSITÉ DE LYON

INSA INSTITUT NATIONAL DES SCIENCES APPLIQUÉES LYON

PRINCETON UNIVERSITY

What is a Public Goods Game?

Public Goods Game

Let’s play a game <u>together</u>…

Public Goods Game

Let's play a game <u>together</u>…

- Each one of you has 1 coin 

Public Goods Game

Let’s play a game <u>together</u>…

- Each one of you has 1 coin 
- Each one of you has 2 possible actions:
  - **Cooperate** - give 1 coin to a common pot
  - **Defect** - keep your coin

Public Goods Game

- After everyone took an action, our common pot will hopefully grow…

- Iff the contributions in the pot exceed 15 coins, its total amount will **grow by 30%** and the coins will be **shared among everyone!**

Public Goods Game

- Let’s play one round…
  - **Cooperate** - Raise your left hand
  - **Defect** - Raise your right hand

Public Goods Game

- Assuming the collect the minimum number of contributions… the common pot will be shared among everyone!

Public Goods Game

- Assuming the collect the minimum number of contributions… the common pot will be shared among everyone!

- **People that cooperated**, now have only the amount they received from the pot.

- **People that defected**, now have the amount they received from the pot **plus their initial coin!**

Public Goods Game

- Defectors will end up with more coins than the others…

- But if we keep playing this iteratively, some people might employ **mechanisms to promote cooperation** (e.g., punishing, criticising)

    - …even those that defect and want to take advantage of the others

Public Goods Game

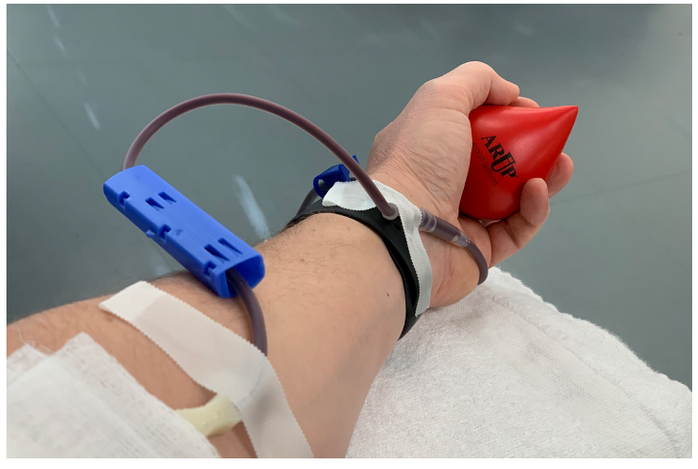
- What would you think of someone that criticises your action of defecting?

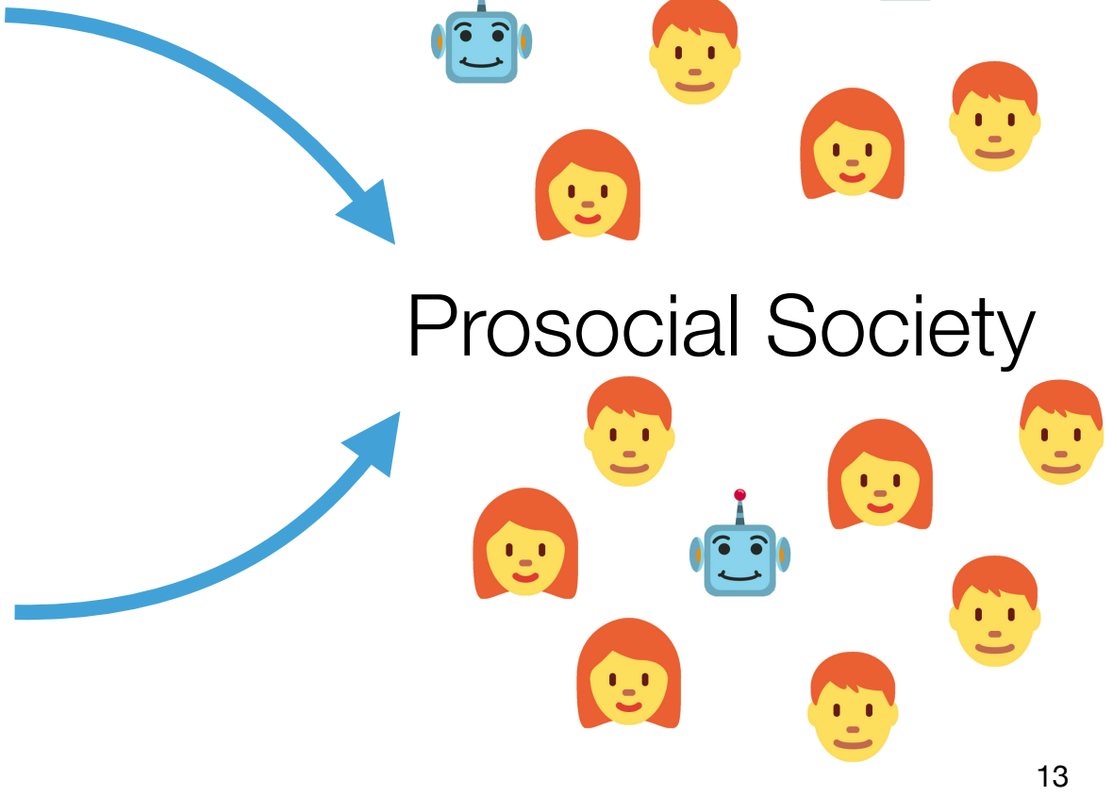
- Would it matter if that person has also defected or not?

- Would you change your action next time?

Why does this game matter?

Motivation





Research Questions

Research Questions

In a Public Goods Game…

Research Questions

In a Public Goods Game…

- How do people perceive a robotic teammate that criticises them when they defect?

Research Questions

In a Public Goods Game…

- How do people perceive a robotic teammate that criticises them when they defect?
    - And what if that robot does not “practice what it preaches”?

Research Questions

In a Public Goods Game…

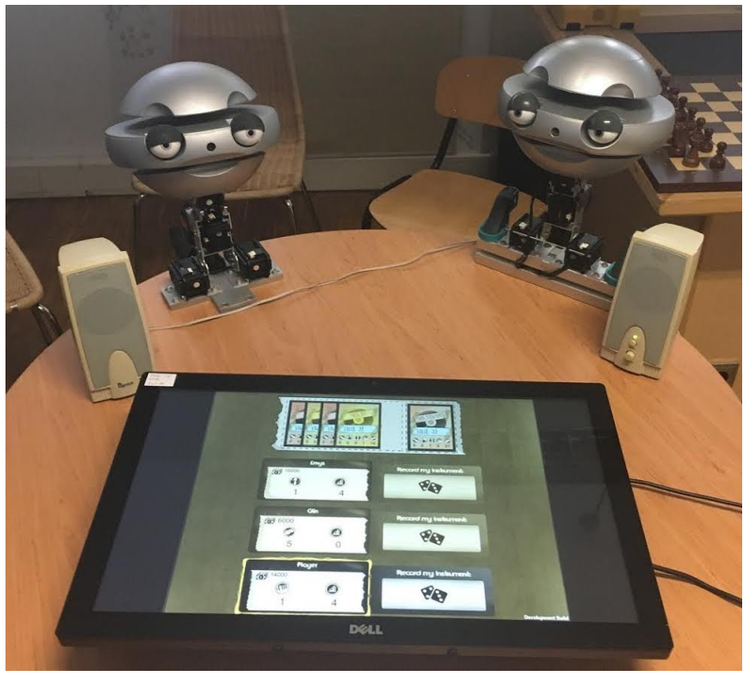
- How do people perceive a robotic teammate that criticises them when they defect?
    - And what if that robot does not “practice what it preaches”?
- Do people cooperate more with the team when being criticised by a robot that acts according to its criticism?

Scenario

Scenario

- “For The Record”
- N-player collaborative game
- Variant of a Public Goods
- Musical metaphor:

“The band needs to collect the maximum number of successful albums without collapsing”



Scenario

- Each round, there is a social dilemma:

Scenario

Threshold game with **uncertain returns**

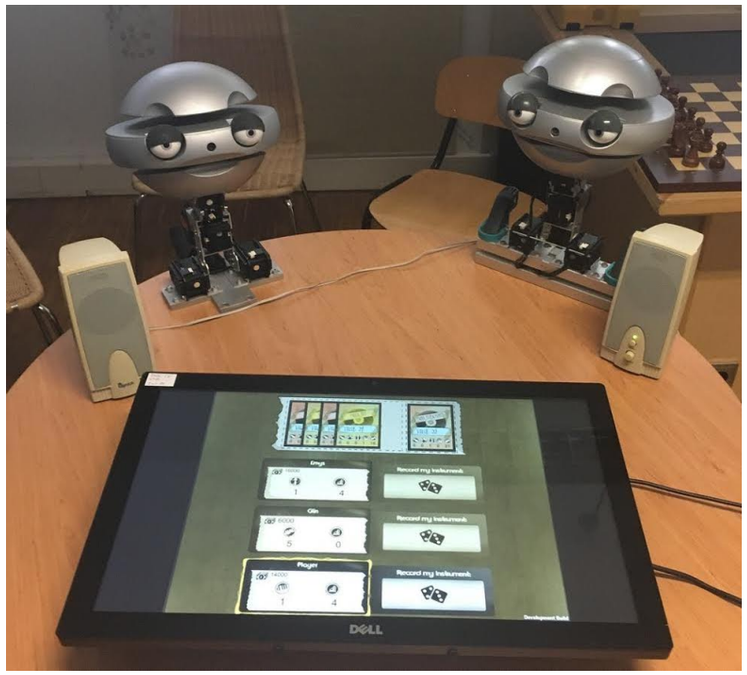
digital dice

control the outcome (win/lose)

Scenario

3 players

- 1 person
- 2 autonomous robots



Hypotheses

Hypotheses

H1 - People will **perceive more negatively a selfish robot** when it criticises, compared to a pro-social robot when it makes the same criticism.

Hypotheses

H1 - People will **perceive more negatively a selfish robot** when it criticises, compared to a pro-social robot when it makes the same criticism.

H2 - People will **cooperate more when criticised by a pro-social robot** than when criticised by a selfish robot.

User Study

Experimental Design

- Mixed experimental design
    - Within-subjects variable - strategy of robotic partner

Pro-social

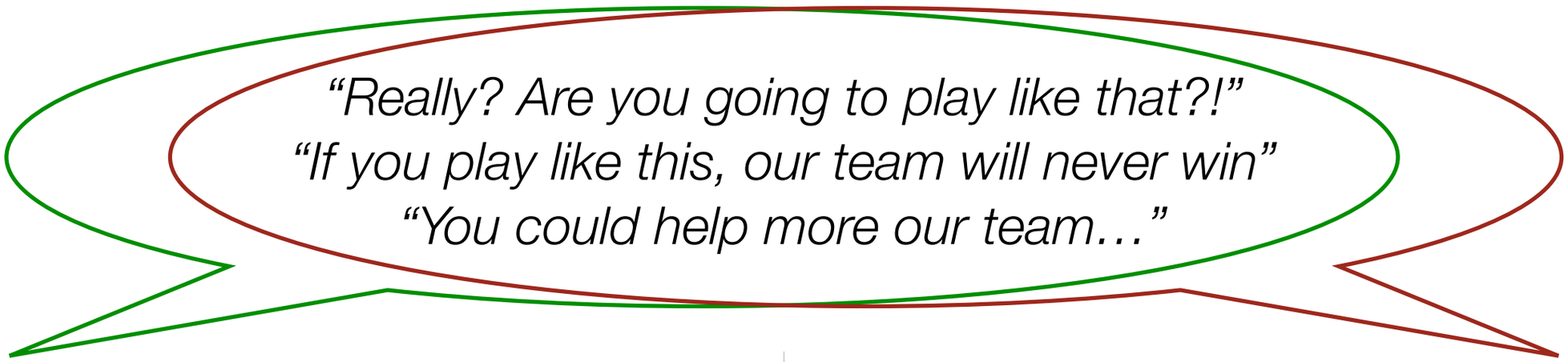
Unconditionally chooses ***to cooperate***

Selfish

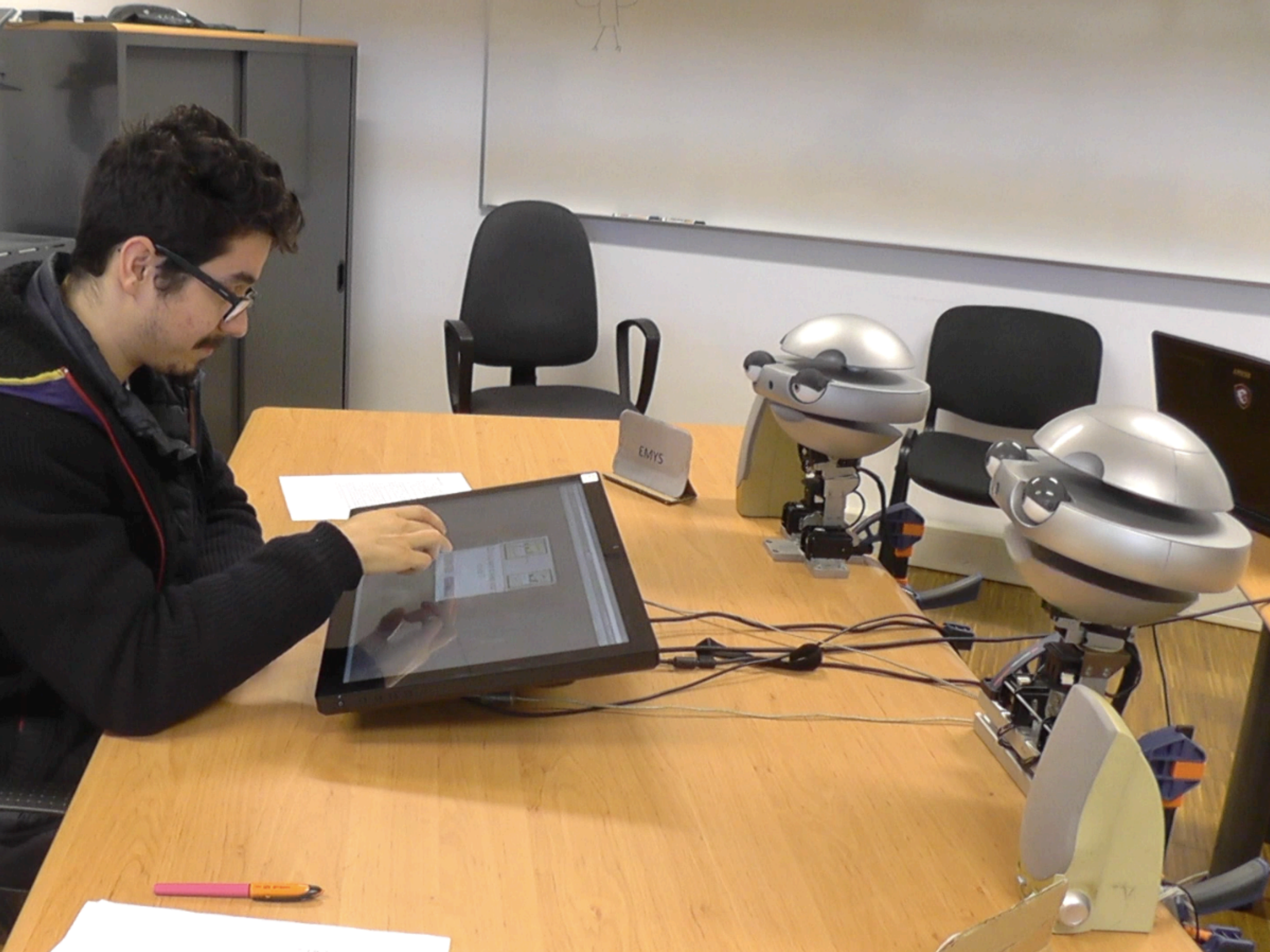
Unconditionally chooses ***to defect***

Experimental Design

- Mixed experimental design
    - Within-subjects variable - strategy of robotic partner
    - Between-subjects variable - who is the criticiser



Pro-social Critic	Selfish Critic
Alignment W&D	**Misalignment W&D**


EMYS

Experimental procedure

1. Training (10-15 minutes)
    - Game rules
    - Training game with the researcher scripted for the team to win
2. Interaction with the robots (5 minutes)
    - Scripted for the team to lose, based on previous findings [1]
3. Questionnaire (10 minutes)

Total Time ~ 30 minutes

[1] Correia, F., Mascarenhas, S.F., Gomes, S., Arriaga, P., Leite, I., Prada, R., Melo, F.S. and Paiva, A., 2019, March. **Exploring prosociality in human-robot teams**. In 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI) (pp. 143-151). IEEE.

Measures - Social Attributes (H1)

RoSAS Questionnaire [1]

- Warmth
- Discomfort
- Competence

[1] C. M. Carpinella, A. B. Wyman, M. A. Perez, and S. J. Stroessner, "**The robotic social attributes scale (rosas): development and validation**" in *ACM/IEEE Int. Conf. on Human-Robot Interaction*, 2017.

Measures - Cooperation Rate (H2)

- Number of times participants cooperated (out of 4)

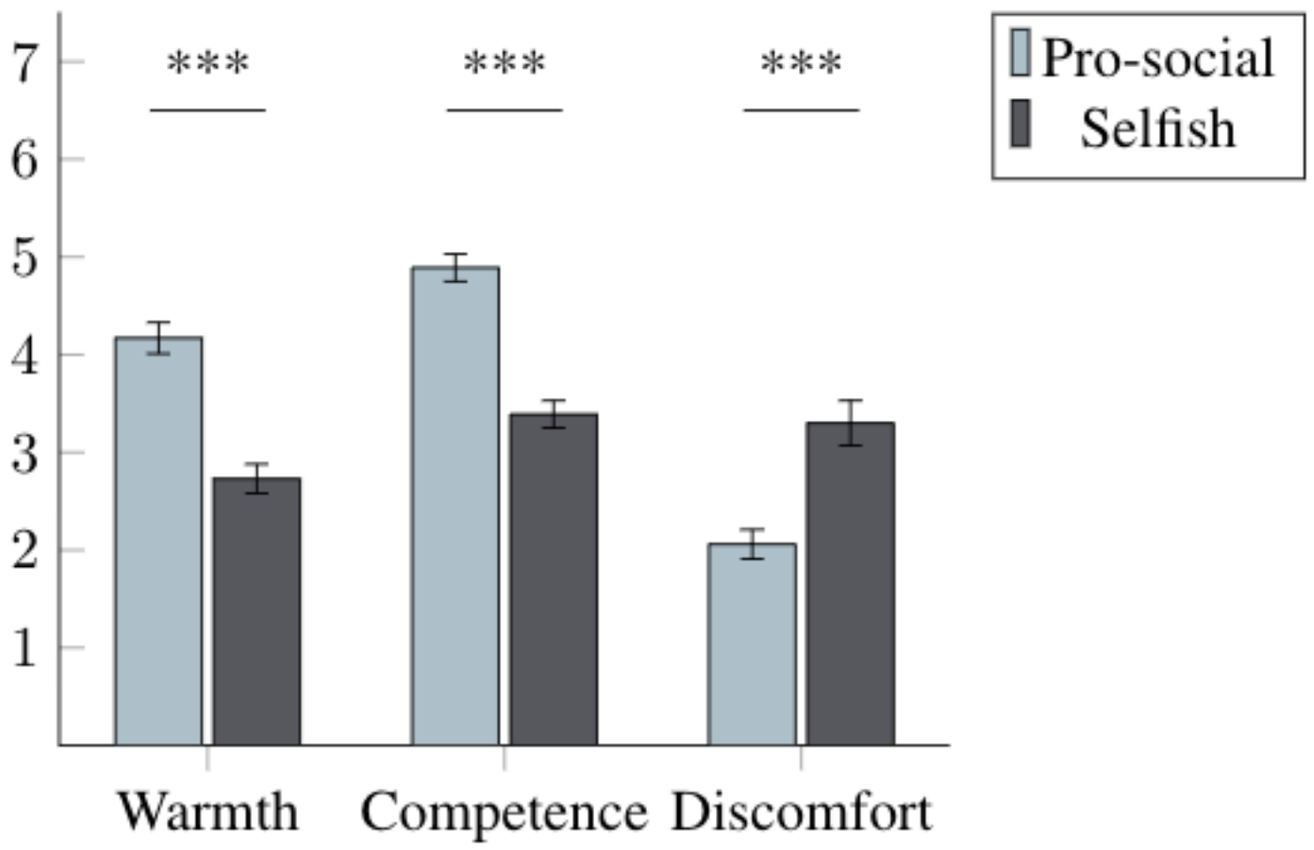
Participation

- 46 participants (out of 50)
  - 22 in the PC condition (Alignment W&D)
  - 24 in the SC condition (Misalignment W&D)
- students from university campus
- 26 males, 20 females
- [18-49] years old (M = 24.04, SD = 5.62)

Results

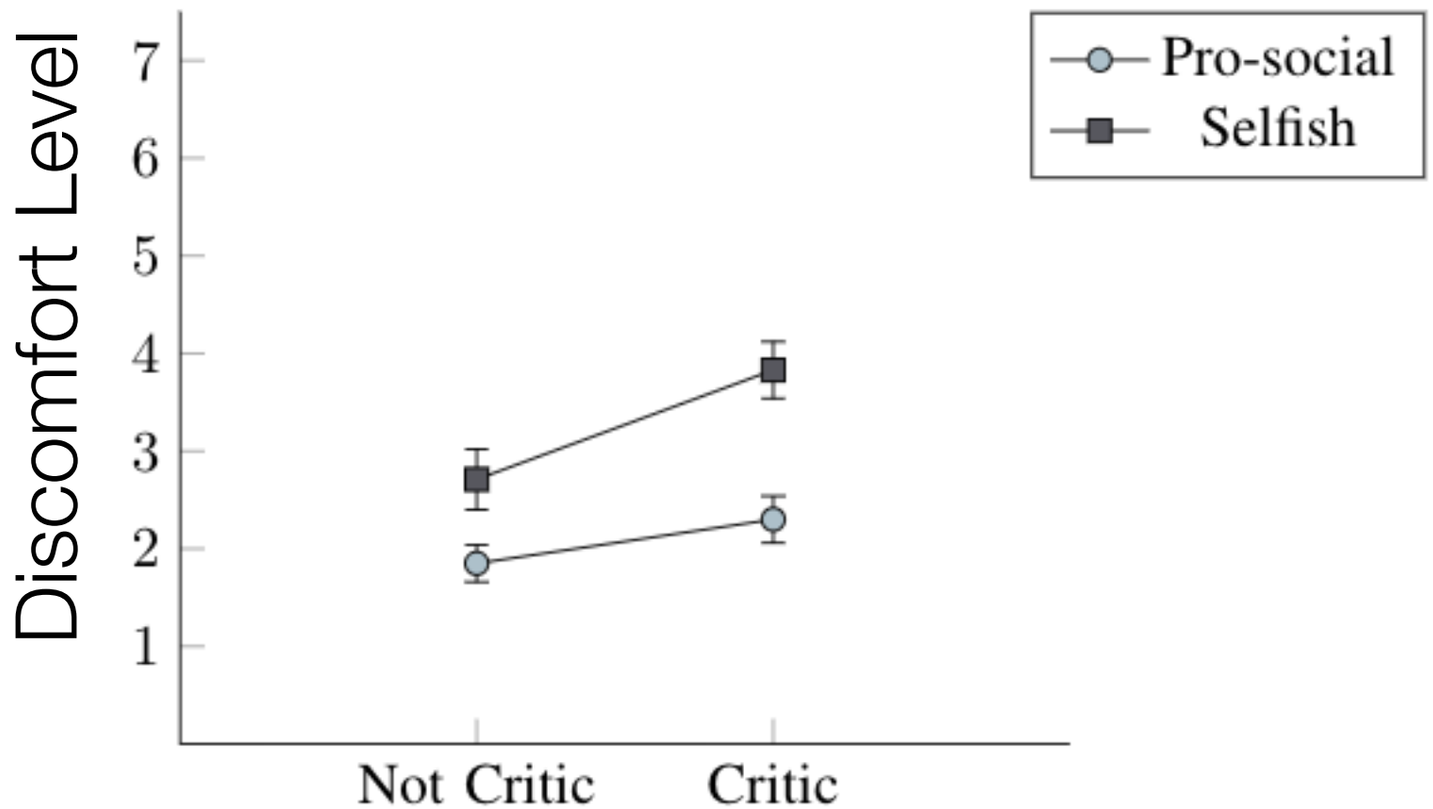
Results - Social Attributes

- We found a statistically **significant main effect** of the **strategy** on the perceptions of **warmth** ($p < 0.001$), **competence** ($p<0.001$) and **discomfort** ($p < 0.001$).



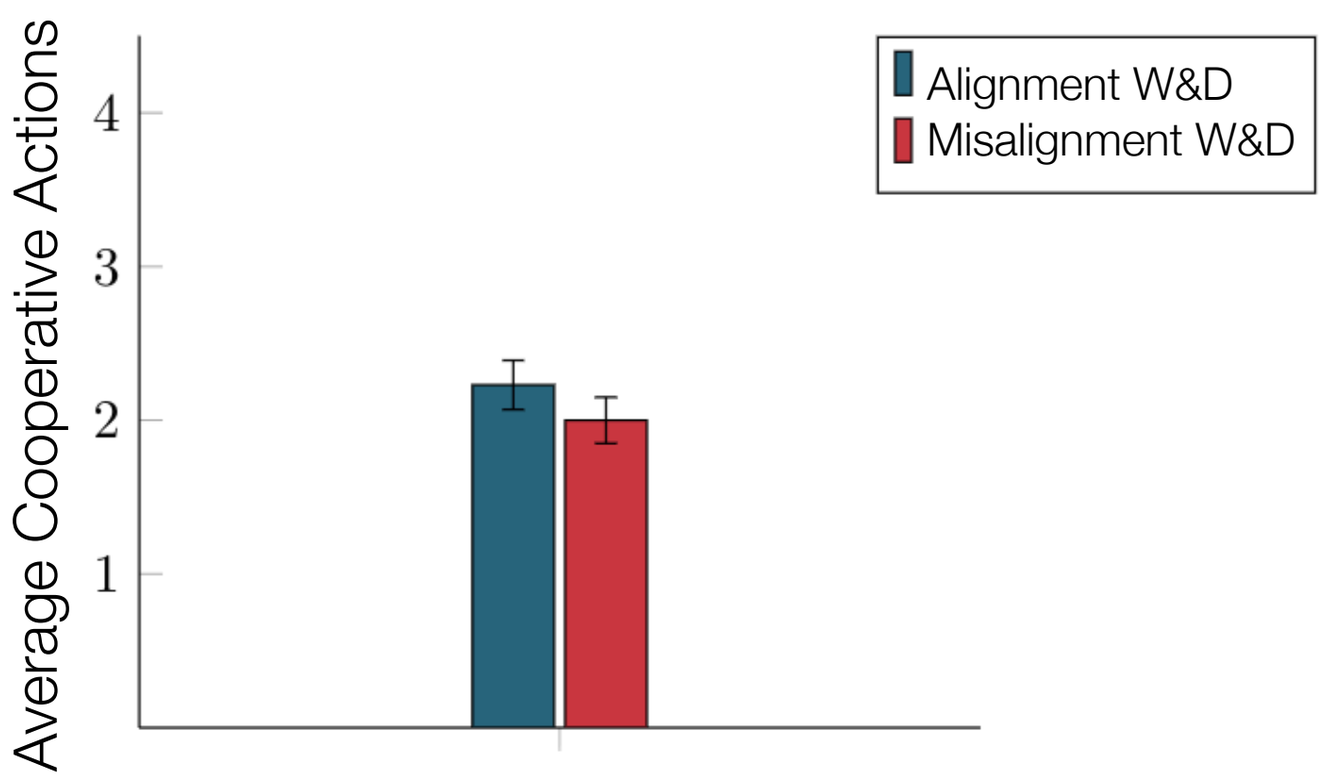
Results - Social Attributes

- We found a statistically **significant interaction effect** between the **strategy** and the **critic behaviour** on the perception of **discomfort** (p = 0.001).



Results - Cooperation Rate

- We found no significant differences on the cooperation rate between the two conditions ($p = 0.318$).



Discussion

Discussion - Social Attributes

H1 - People will **perceive more negatively a selfish robot** when it criticises, compared to a pro-social robot when it makes the same criticism.

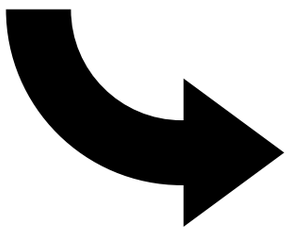
Discussion - Social Attributes

H1 - People will **perceive more negatively a selfish robot** when it criticises, compared to a pro-social robot when it makes the same criticism.

…only in terms of discomfort!

Discussion - Social Attributes

H1 - People will **perceive more negatively a selfish robot** when it criticises, compared to a pro-social robot when it makes the same criticism.



People **will not perceive more negatively a prosocial** robot when it criticises…

Discussion - Cooperation Rate

H2 - People will **cooperate more when criticised by a pro-social robot** than when criticised by a selfish robot.

Discussion - Cooperation Rate

H2 - People will **cooperate more when criticised by a pro-social robot** than when criticised by a selfish robot.

Limitations & Future Work

Limitations & Future Work

1. In the PC condition (Alignment W&D), the robot did not criticise all the defectors, only the participant…

Limitations & Future Work

1. In the PC condition (Alignment W&D), the robot did not criticise all the defectors, only the participant…

    - It would be interesting to compare with a prosocial robot that **criticises all defectors including other robotic teammates.**

Limitations & Future Work

1. In the PC condition (Alignment W&D), the robot did not criticise all the defectors, only the participant…

    - It would be interesting to compare with a prosocial robot that **criticises all defectors including other robotic teammates.**

2. The presence of a free-rider on the team may not be worth for participants to change their strategy.

Limitations & Future Work

1. In the PC condition (Alignment W&D), the robot did not criticise all the defectors, only the participant…

    - It would be interesting to compare with a prosocial robot that **criticises all defectors including other robotic teammates.**

2. The presence of a free-rider on the team may not be worth for participants to change their strategy.

    - Comparing with robotic teammates that **both employ a prosocial strategy.**

Limitations & Future Work

1. In the PC condition (Alignment W&D), the robot did not criticise all the defectors, only the participant…

    - It would be interesting to compare with a prosocial robot that **criticises all defectors including other robotic teammates.**

2. The presence of a free-rider on the team may not be worth for participants to change their strategy.

    - Comparing with robotic teammates that **both employ a prosocial strategy.**

3. Number of rounds

Conclusions

Conclusions

- We explored the **expression of criticisms** by **social robots**
- verbal **mechanism** to promote cooperation in **collaborative interactions**

Conclusions

- We found **no evidence** that (mis)alignment of words and deeds influenced **cooperation**
- **misalignment of words and deeds** by a robotic teammate may **affect negatively the discomfort** attributed to it

Conclusions

The perception a human has of a robotic teammate that criticises him is not damaged as long as…

Conclusions

The perception a human has of a robotic teammate that criticises him is not damaged as long as…

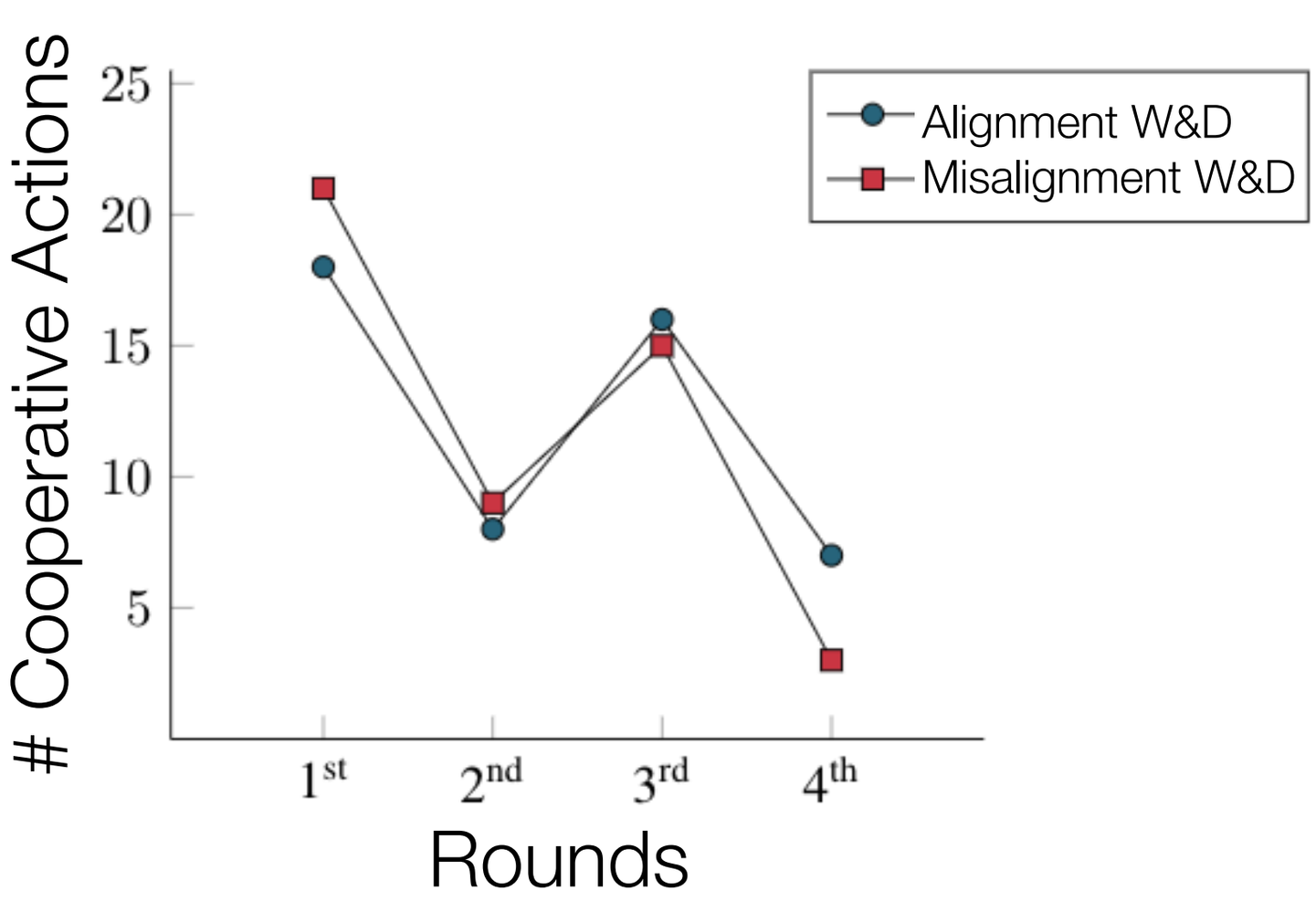
…the robot displays an alignment of words and deeds.

Thank you!

filipacorreia@tecnico.ulisboa.pt

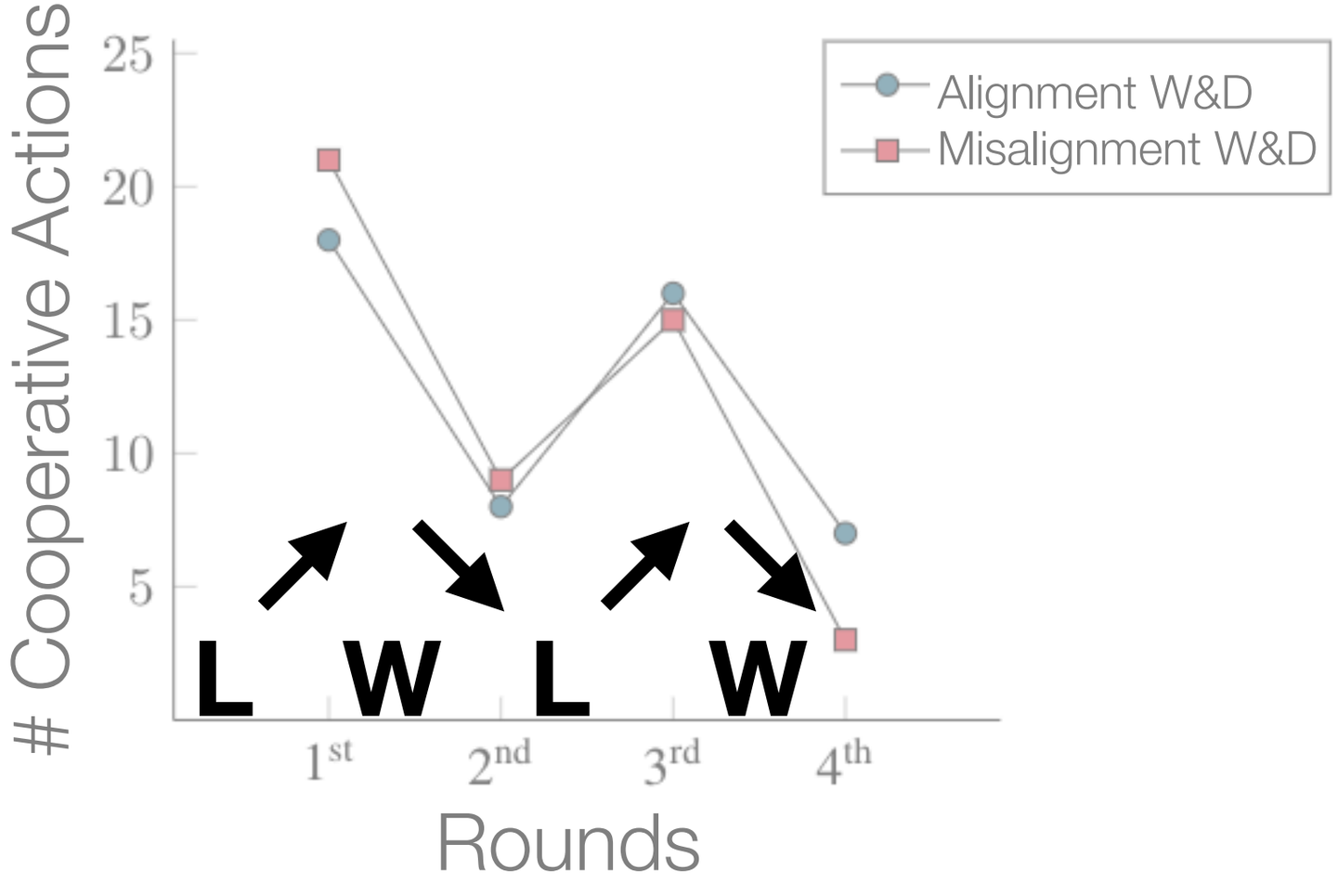
Results - Cooperation Rate per Round

- We looked at the cooperation rate in each decision point between the two conditions…



Results - Cooperation Rate

- It may be related to scripted sequence of the outcomes per round.



Results - Cooperation Rate

- Cooperation right after the first criticism (p=0.09)
    - 72.2% (13 out of 18) in the Alignment W&D
    - 45.0% (9 out 20) in the Misalignment W&D

Results - Cooperation Rate

- Cooperation right after the first criticism (p=0.09)
    - 72.2% (13 out of 18) in the Alignment W&D
    - 45.0% (9 out 20) in the Misalignment W&D

… small sample size…