Cooperating AI

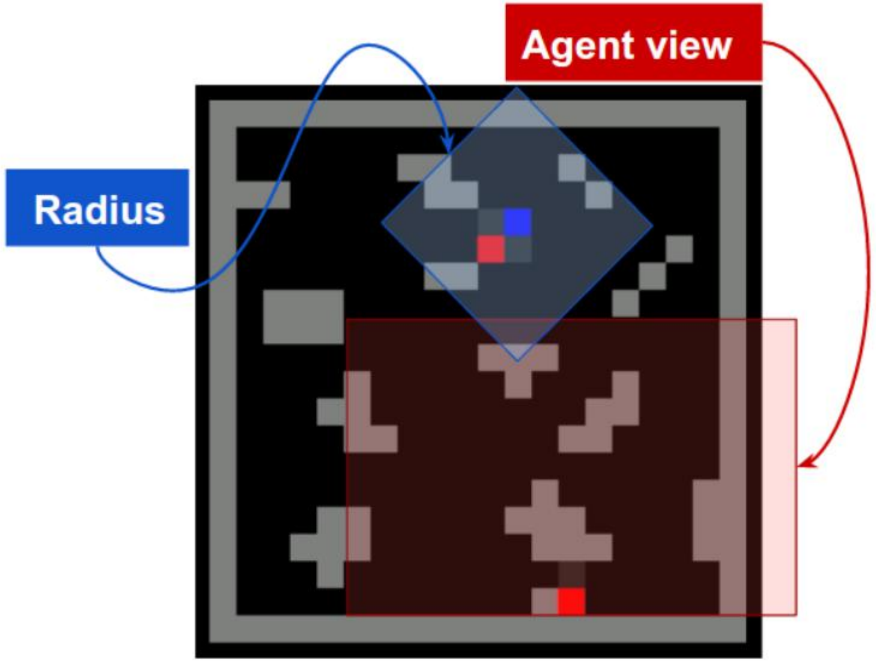
Making artificial intelligence more human

Seminar: Ist künstliche Intelligenz gefährlich?
PD Dr. Ullrich Köthe, SS 2017
Universität Heidelberg
Presentation: Julian Heiss

Picture: http://weknownyourdreamz.com/symbol/sl598741.html

Multi-agent Reinforcement Learning in Sequential social dilemmas[1]

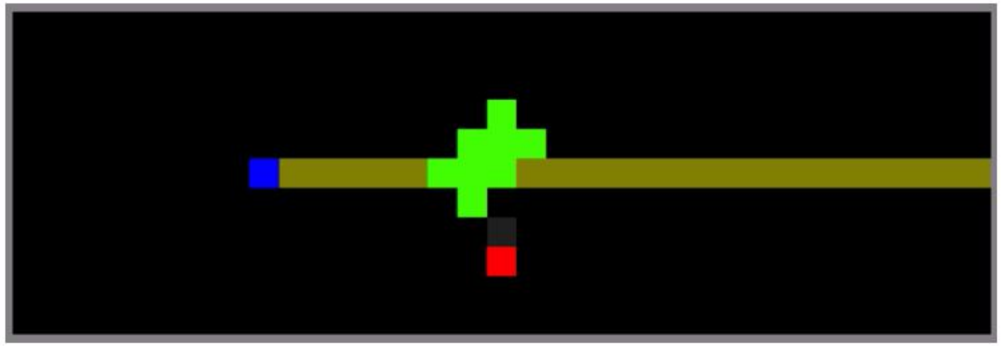
- **Machine-Machine** cooperation.
- In **Wolfpack** game, **learning lone-wolf policy is easier** than learning cooperative pack-hunting policy. This is because the former does not require actions to be conditioned on the presence of a partner within the capture radius.
- Greater network size leads to **more cooperation**.



Source: [1]

[1] Leibo et al. (2017). Multi-agent Reinforcement Learning in Sequential Social Dilemmas. Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AA-MAS 2017).

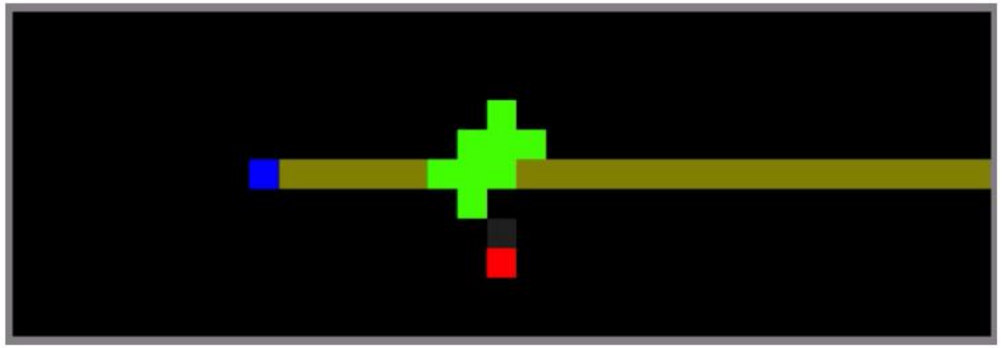
- In the **Gathering** game the situation is reversed. **Cooperative policies are easier** to learn since they need only be concerned with apples and may not depend on the rival player’s actions.
- For Gathering, an **increase in network size** leads to an increase in the agent’s tendency to **defect**



Source: [1]

- **Capacity** for more complex actions leads to **more** cooperative behaviour in the wolfpack, to **less** cooperation in the gathering game.

→ Increasing capacity does not automatically make the algorithm more cooperative.



Source: [1]

- Shooting a beam might still be favourable, e.g. so that not both go for the same apple.
- Still need to improve cooperation.
- Possible ways to go at it:
  - – **Learn reward** function for game.
  - – Talk before you shoot. **Communication** is key.

Cooperating with Machines

Source: [2] Crandall et al. (2017). Cooperating with Machines. Computing Research Repository (CoRR), abs/1703.0. http://arxiv.org/abs/1703.06207

- **Motivation**: Need algorithms to be able to cooperate, not just compete in special areas.

- **Goal**: AI algorithm cooperating with people/machines as good as humans cooperate (in arbitrary two-player repeated interactions).

- **Conditions** for successful algorithm: Generality, flexibility (associates), learning speed (human-machine)

- **M-M** and **H-M** cooperation.

- Standard ML algorithms could not bring players to cooperate effectively long-term.

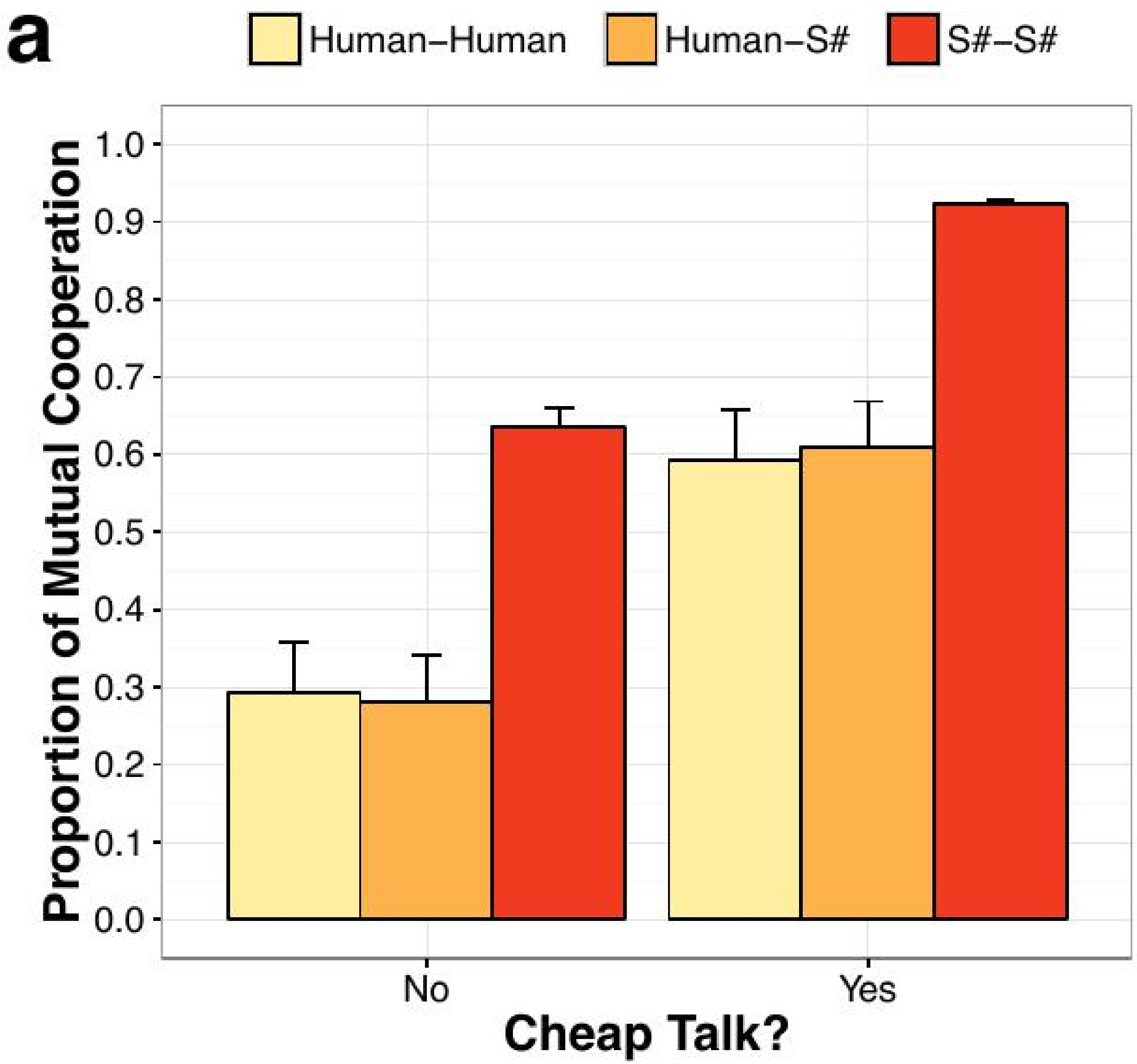
- Idea: Introduce element of **communication**.
  - Helps to create shared representations.
  - **Cheap talk**: "Cheap talk refers to non-binding, unmediated, and costless communication"

- Cheap talk: Feedback and Planning.
- Difficulties: Some algorithm do not have easy **understandable representations**. But works with S++.
- 19 possible sentences (different categories).

Speech ID	Text	Speech ID	Text
0	Do as I say, or I'll punish you.	10	We can both do better than this.
1	I accept your last proposal.	11	Curse you.
2	I don't accept your proposal.	12	You betrayed me.
3	That's not fair.	13	You will pay for this!
4	I don't trust you.	14	In your face!
5	Excellent!	15	Let's always play `<action pair>`.
6	Sweet. We are getting rich.	16	This round, let's play `<action pair>`.
7	Give me another chance.	17	Don't play `<action>`.
8	Okay. I forgive you.	18	Let's alternate between `<action pair>` and `<action pair>`.
9	I'm changing my strategy.		

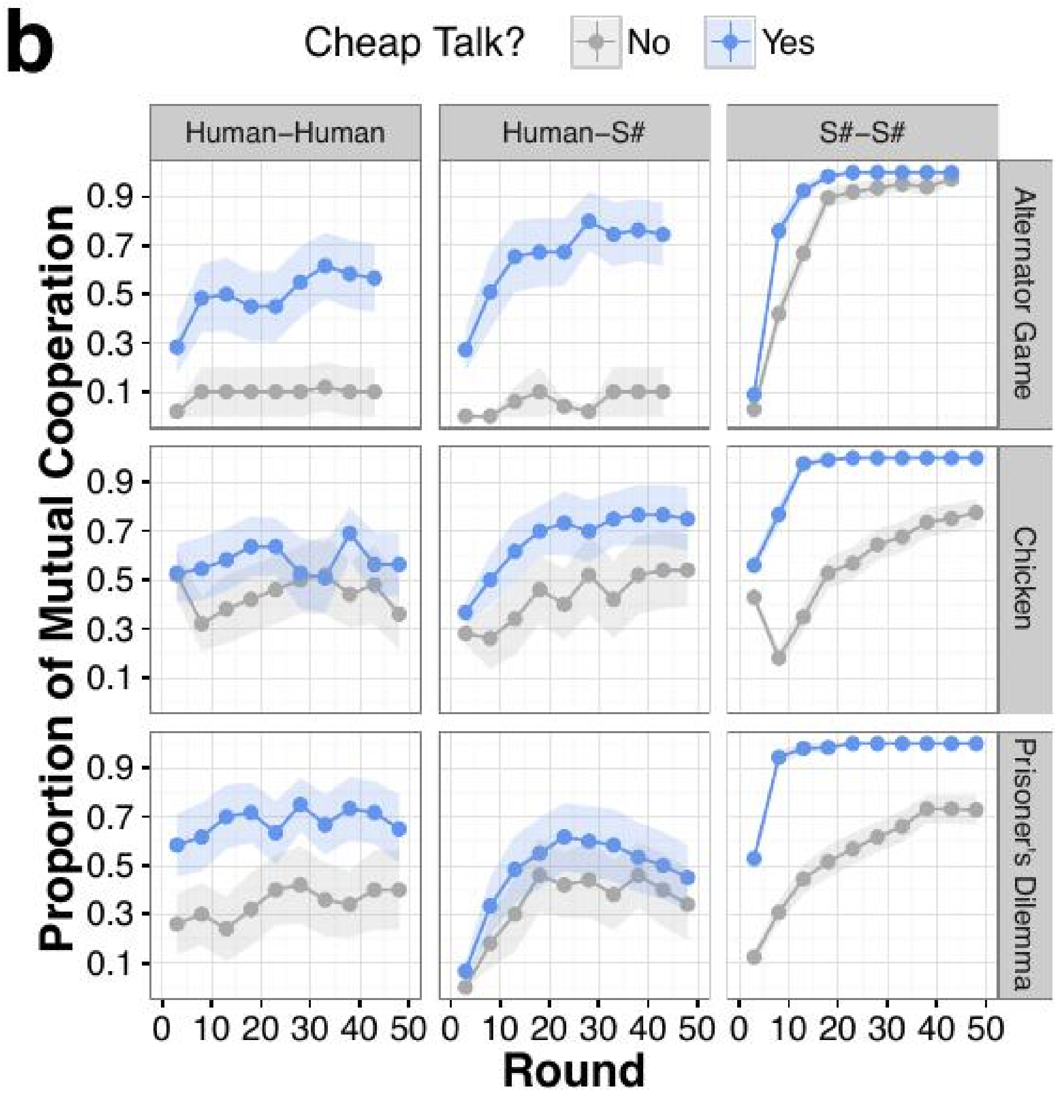
Source: [2]

- **Results**: Development of S# (Extension to S++)
  - S++ brings generality with it. Also fast convergence.
  - Communication via cheap talk is not the only, but on of the main features.
  - Info from communication reduces set of experts.


a
Human–Human
Human–S#
S#–S#
Proportion of Mutual Cooperation
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
No
Yes
Cheap Talk?

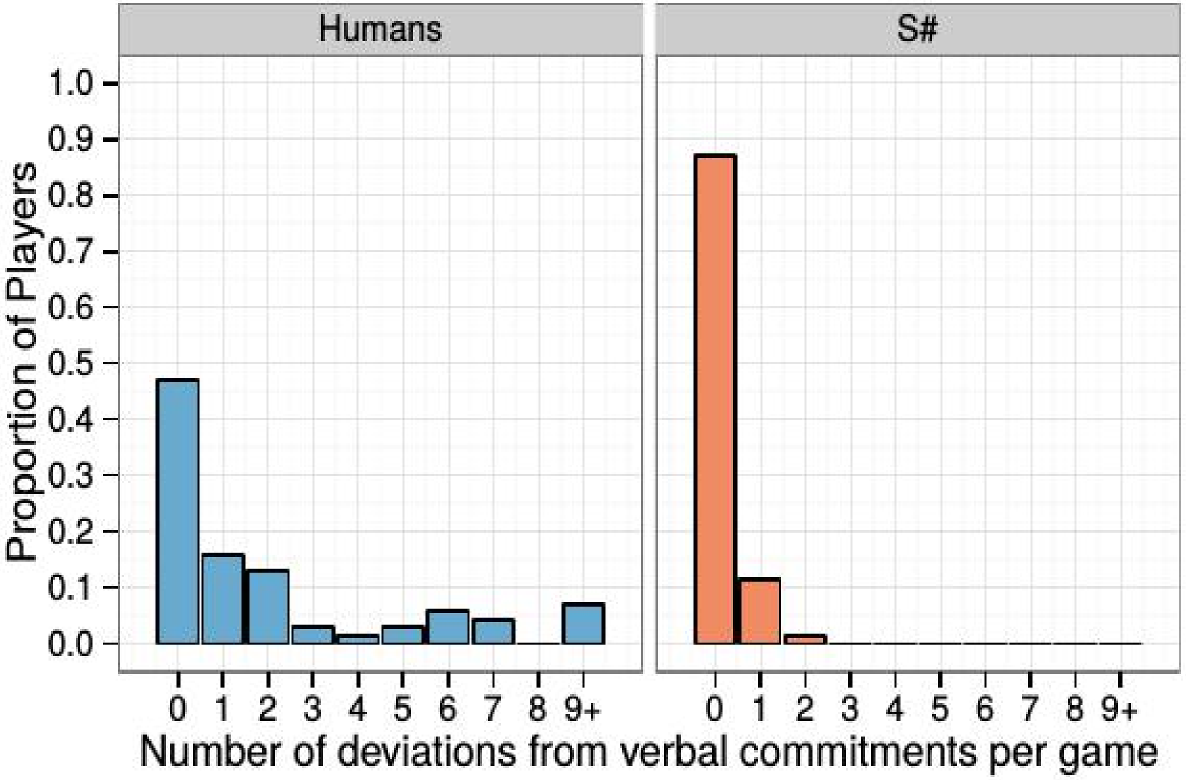
Source: [2]

- To forge mutually cooperative relationships, players must do two things: **Establish** cooperative behaviour and **maintain** it.
- Cheap talk helps with establishing (especially for humans)
- **Loyalty** is a reason for M-M pairs outperforming humans. Also **Honesty.**



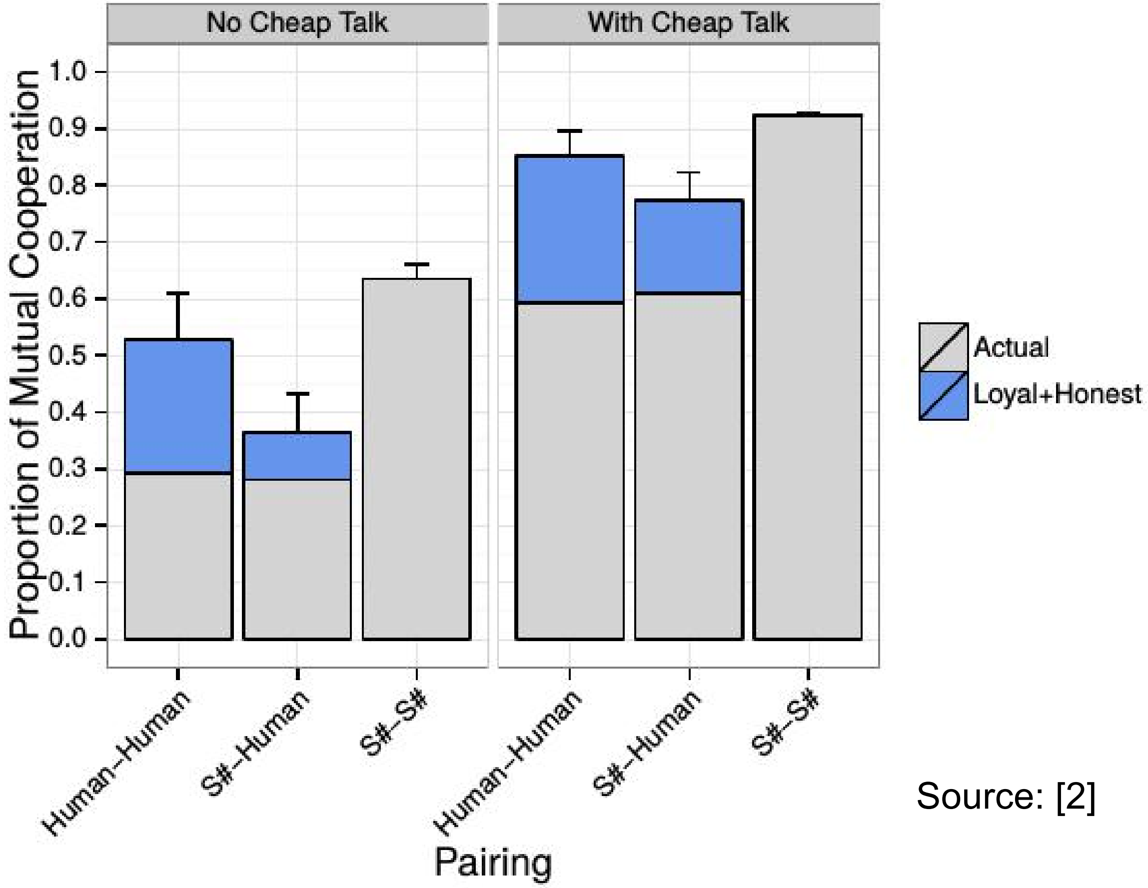
Since verbal commitments by S# are derived from its **intended behaviour**, it does what it says.

Unlike "a sizeable portion" of the human participants.



Source: [2]

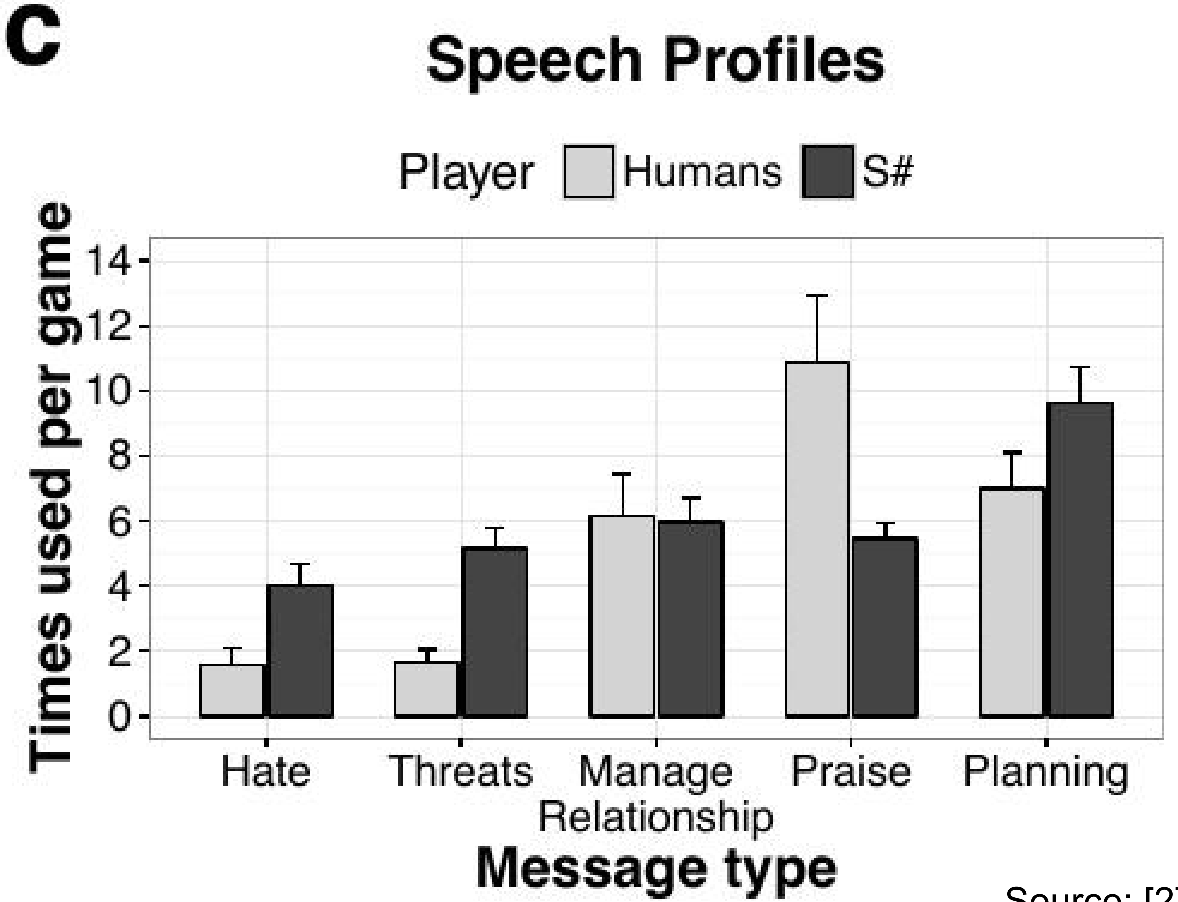
Over **all games** played, a human player had a positive net gain due to betrayals in **just two** interactions.



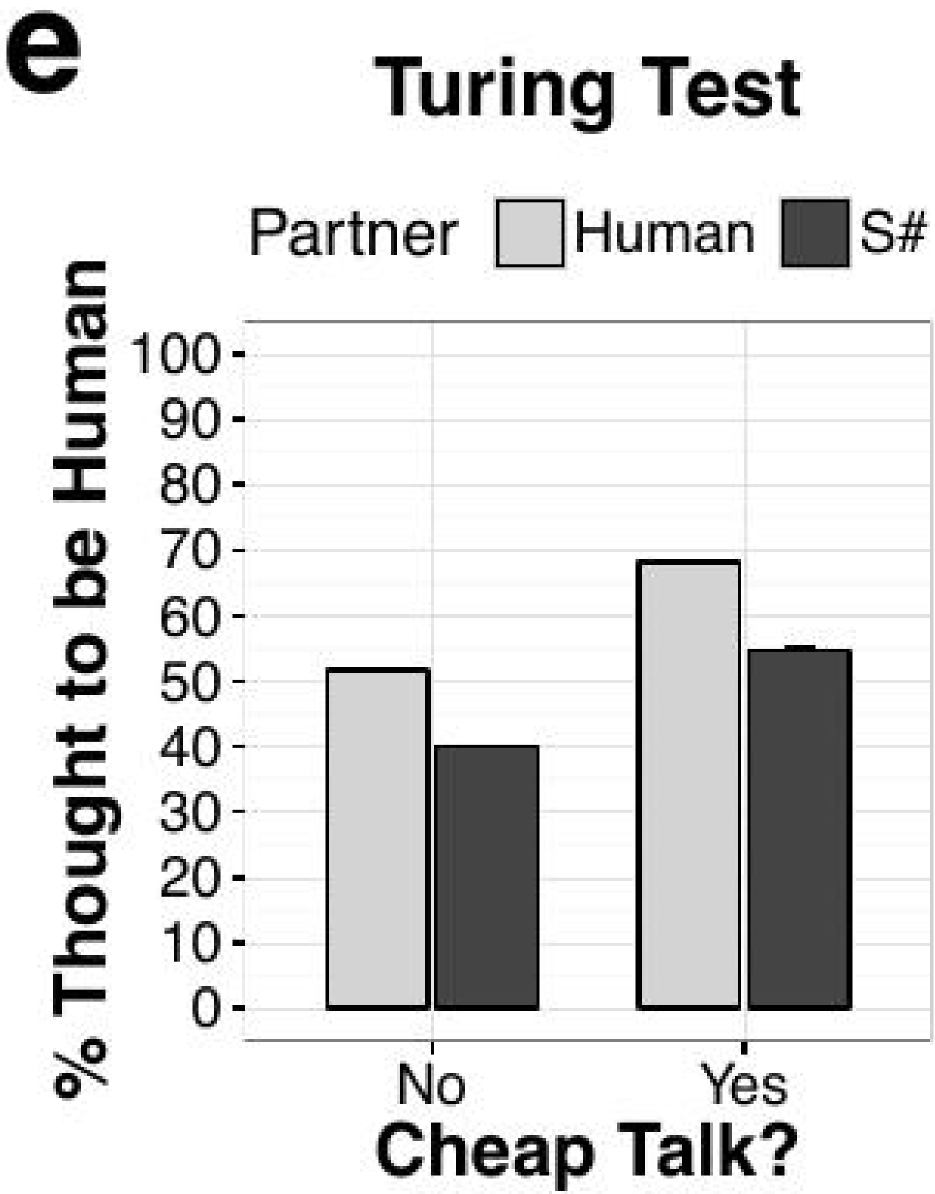
Source: [2]

This graph does not necessarily imply that the AI is more evil than humans.

Maybe threats are just a more "effective" way to ensure cooperation.



Source: [2]


e
Turing Test
Partner
Human
S#
% Thought to be Human
100
90
80
70
60
50
40
30
20
10
0
No
Yes
Cheap Talk?

Source: [2]

“*The machine-learning algorithm learned to be **loyal.***”

(J. Crandall, Author)

This is open for discussion.

- **Big picture**: Added a new mechanism to the algorithm.
  Mimicking humans.
  Feedback by other player is used as part of the input.

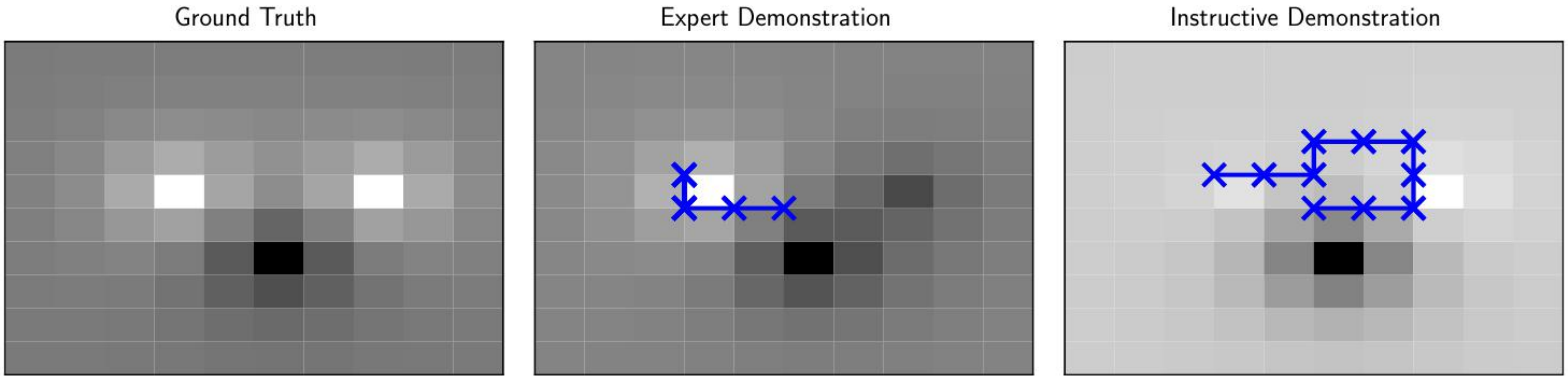
CIRL - Cooperative inverse reinforcement learning[3]

"*If we use, to achieve our purposes, a mechanical agency with whose operation we cannot interfere effectively [. . .] we had better be quite sure that the purpose put into the machine is the purpose which we really desire.*" (Norbert Wiener, 1960)[3]

Value Alignment Principle: Highly autonomous AI systems should be designed so that their goals and behaviors can be assured to align with human values throughout their operation.[4]

Sources: [3] Hadfield-Menell, D., Dragan, A., Abbeel, P., & Russell, S. (2016). Cooperative Inverse Reinforcement Learning, (Nips). Retrieved from http://arxiv.org/abs/1606.03137
[4] https://futureoflife.org/2017/02/03/align-artificial-intelligence-with-human-values/ and https://futureoflife.org/ai-principles/

- A **CIRL** problem is a cooperative partial information game:
  – 2 agents, human and robot.
  – Both rewarded according to the **human**'s reward function.
  – But robot does not initially know what this is.
- Difference to IRL:
  – Observating agent (robot) is optimizing reward for the **human**.
  – Acting Agent might act **suboptimal** to be better at explaining.



Source: [3]

- Contribution: Optimal policy pair can be obtained by solving a **POMDP.**

- "Returning to Wiener’s warning, we believe that the best solution is not to put a specific purpose into the machine at all, but instead to design machines that provably converge to the right purpose as they go along."

- Problem with Value Alignment: **What Do We Want?**

- Understanding what “we” want is a big challenge for AI research.
  - Difficult to **encode** human values in a programming language.
  - Humanity does not agree on **common values**.
  - And the parts we do agree on **change with time**.

- Are human values the **best values** there can be?

“Friendly” AGI via Human Emotion: the Vital Link[5]

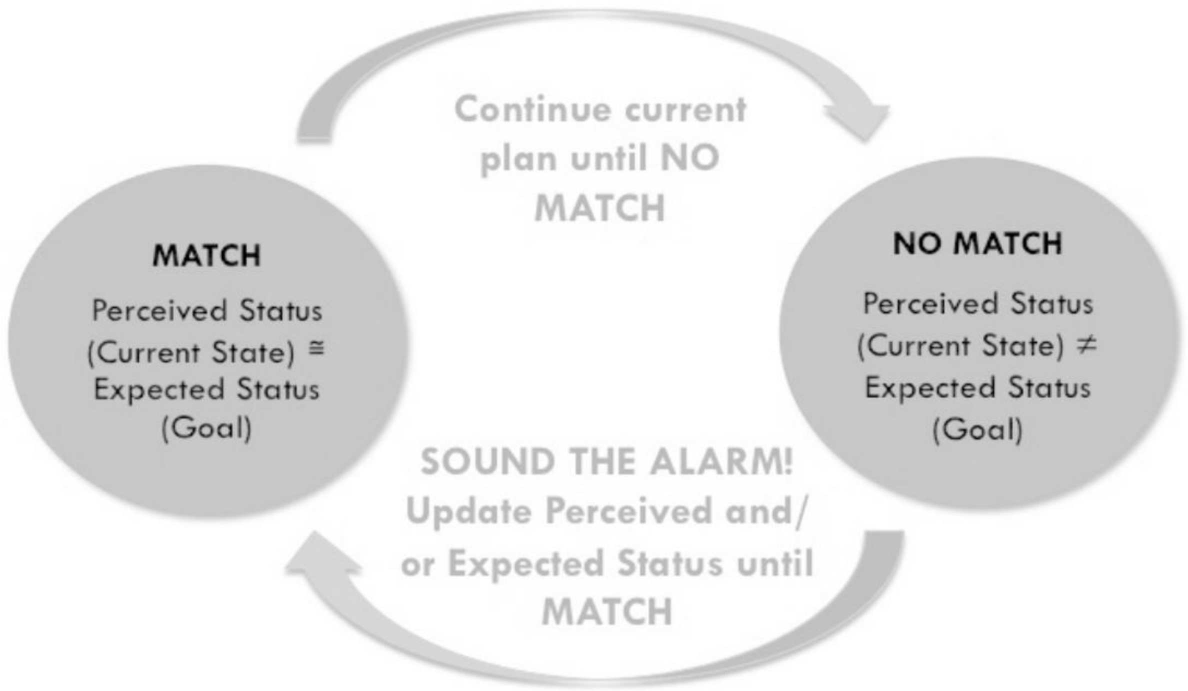
- Consider trade-off situation, ethical dilemma.
- How does a busy AGI even **become aware** that a situation calls for an ethical action or decision?
- Control by human intercession **not feasible**.
- **Recognition** of a problem as first step.

[5] Dietsch, J. (2014). “Friendly” AGI via Human Emotion: the Vital Link. AAAI 2014 Fall Workshop.

- **Decision making**: Limited computational capacity (always information overflow) → distillation/filtering of info (just like humans).

- Make decision based on: memory, pattern recognition, prediction, evaluation.

  → AGI confronted with **same problems as humans.**



Source: [5]

- **Difference Engine**: Expected an perceived states. Situation with great discrepency between expected and perceived will receive attention. → **Homeostasis**
- Valuation of this disparity? Emotions and Needs.

- Include humans in the needs of the AGI.
- Needs of AGI? minimally physical, social needs, data security?
- Needs **distinction** between self and others.
- Who is me? Make "We" and "Me" **inseparable**, so that it includes the human team.
  It is critical than humans are **innate members** of the AGI ingroup.

- **Arguments against** Linking Human Emotion and AGI in **Meta-Beings**: Privacy, Freedom, Individuality; who dominates, whose needs dominate?

- "Multi-individual homo communicatus, joined through our technologies"[5]

- **Communication** might be a key in linking.

Take away

- Role of communication as mechanism in AI algorithms.
- Idea of teaching the AI human values via IRL.

➡ Mechanisms and algorithms can be used to introduce concepts of cooperation into AI.

- Might not be sufficient.
- Maybe resolve the problem by going from "it"/"us" to just **"us"**.

➡ **Meta-beings**

Picture: http://weknownyourdreamz.com/symbol/sl598741.html

Sources

① Leibo et al. (2017). Multi-agent Reinforcement Learning in Sequential Social Dilemmas. Proceedings of the 16th International Conference on Autonomous Agents and Multiagent Systems (AA-MAS 2017).

② Crandall et al. (2017). Cooperating with Machines. Computing Research Repository (CoRR), abs/1703.0. Retrieved from http://arxiv.org/abs/1703.06207.

③ Hadfield-Menell, D., Dragan, A., Abbeel, P., & Russell, S. (2016). Cooperative Inverse Reinforcement Learning, (Nips). Retrieved from http://arxiv.org/abs/1606.03137.

④ Ariel Conn. https://futureoflife.org/2017/02/03/align-artificial-intelligence-with-human-values/
and https://futureoflife.org/ai-principles/

⑤ Dietsch, J. (2014). “Friendly” AGI via Human Emotion: the Vital Link. AAAI 2014 Fall Workshop.

- james barrat 2013 ai
- harming humans. asimow

Whitaker Paper: Modeling of Donation games

- **Social brain hypothesis**
- **Social Heuristics Hypothesis**:
  behaviours that support success in regular social interactions become intuitive and automatic (type-1, intuitive), unless they are moderated by reflective type-2 (cognitive) processes that represent learning to update a type-1 heuristic.

- Using "social comparison", reputation.
- Avoid free riders.
- results showed that evolution favours the strategy to donate to those who are at least as reputable as oneself

Big picture: Introduced a score