

Evidence that Robots Trigger a Cheating Detector in Humans

Alexandru Litoiu, Daniel Ullman, Jason Kim, and Brian Scassellati



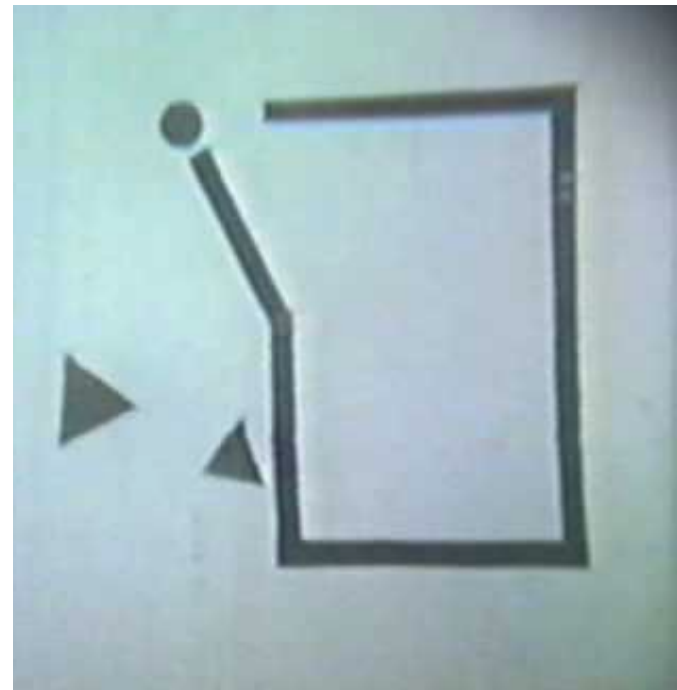
Social Robotics Lab
Yale University





ATTRIBUTIONS OF AGENCY

- Low-level perceptions of movement generate attributions of intentionality
- Interested in generating intentionality from top-down cognitive effects

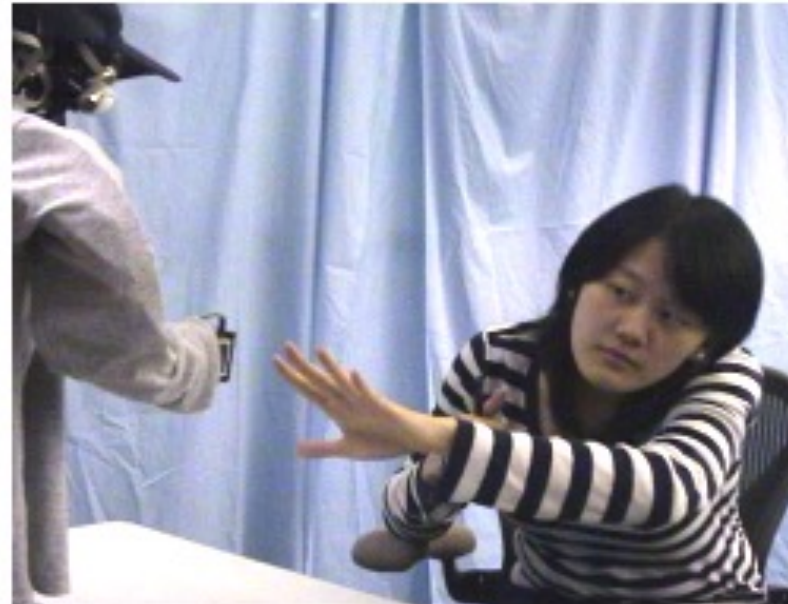


CHEATING ROBOTS ARE MORE ENGAGING

Short et al. showed that a cheating robot:

- Rock-paper-scissors
- Has more intentions and desires
- “Sneaky bastard”

Results were clear, but explanation was not



Short E, Hart J, Vu M, Scassellati B. *No Fair!! An Interaction with a Cheating Robot*. The 5th ACM/IEEE International Conference on Human-Robot Interaction. (2010).



POSSIBLE CAUSES OF EFFECTS (SHORT ET AL.)

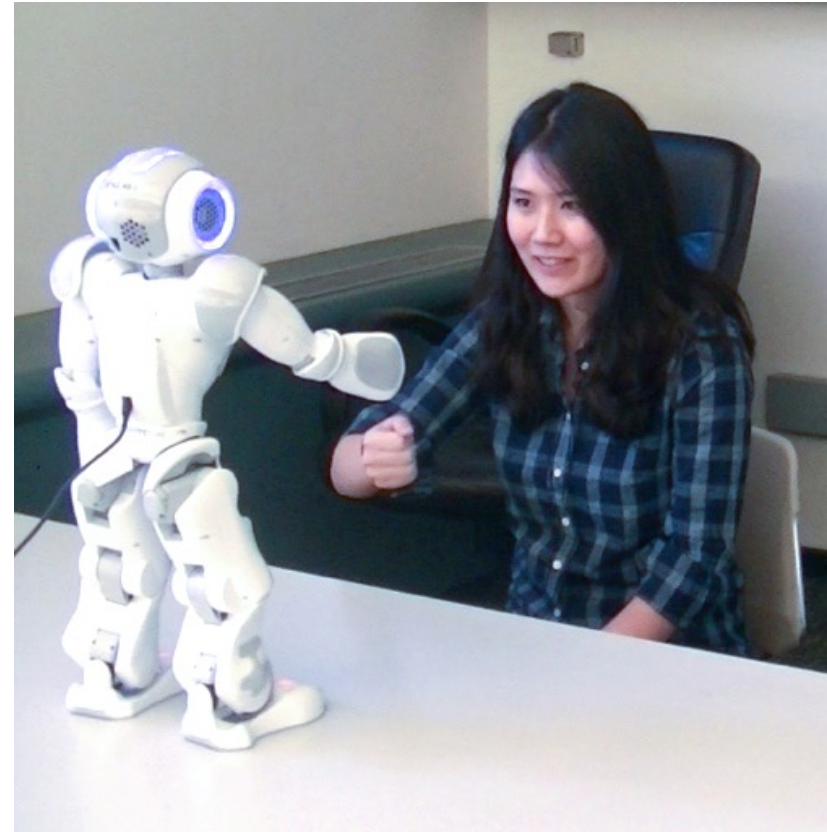
Condition	Utterance	Behavior	
Verbal Cheat	Incorrect "Yes, I win"	Correct	Malfunction
Action Cheat	Incorrect "Yes, I win"	Changed gesture to win	Cheat, Agency

What is the cause of extra attributions in the action cheat?

- Added complexity of the motion?
- Cheating effect?

MAIN DESIGN

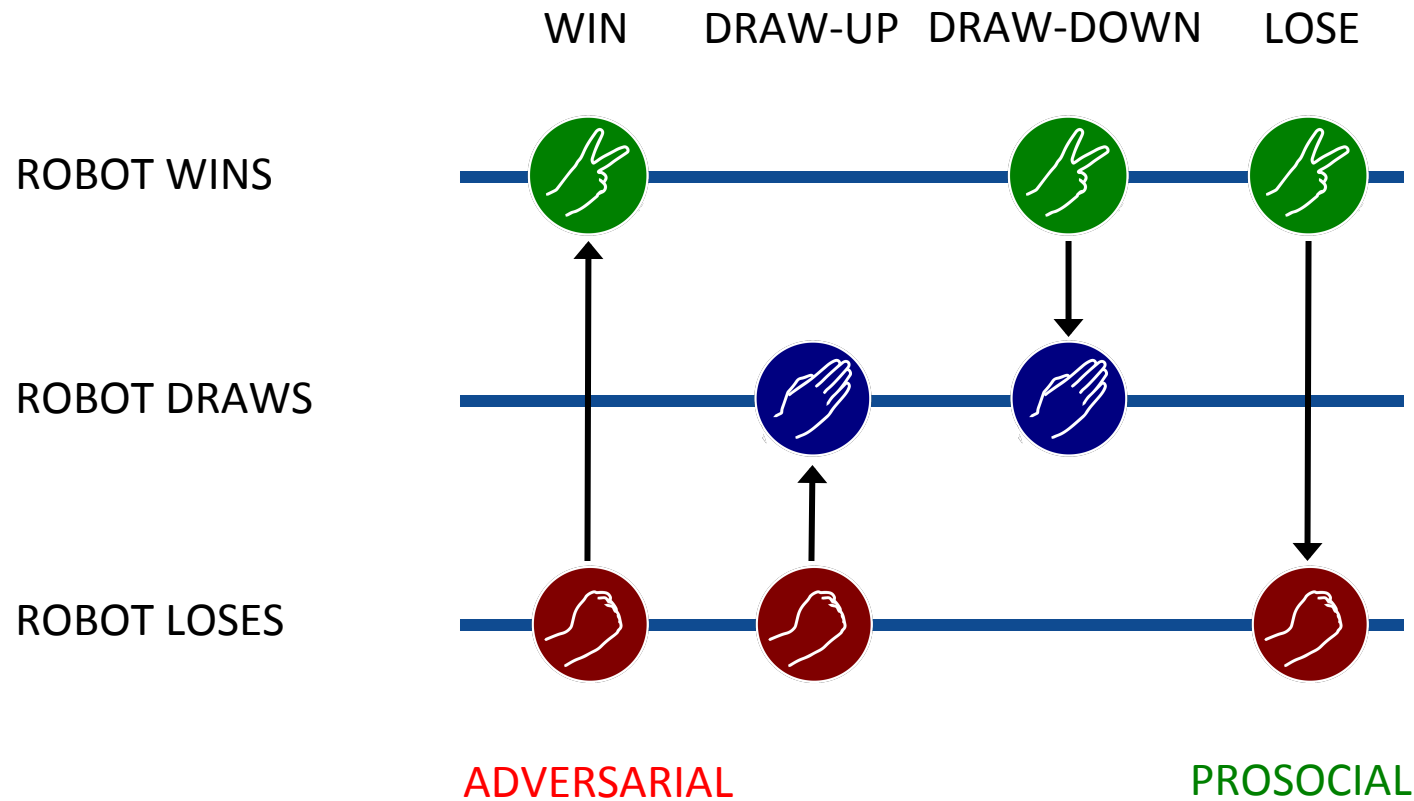
- Maintain amount of motion between conditions
- Vary the “directionality” of cheat from adversarial to prosocial





EXPERIMENTAL CONDITIONS

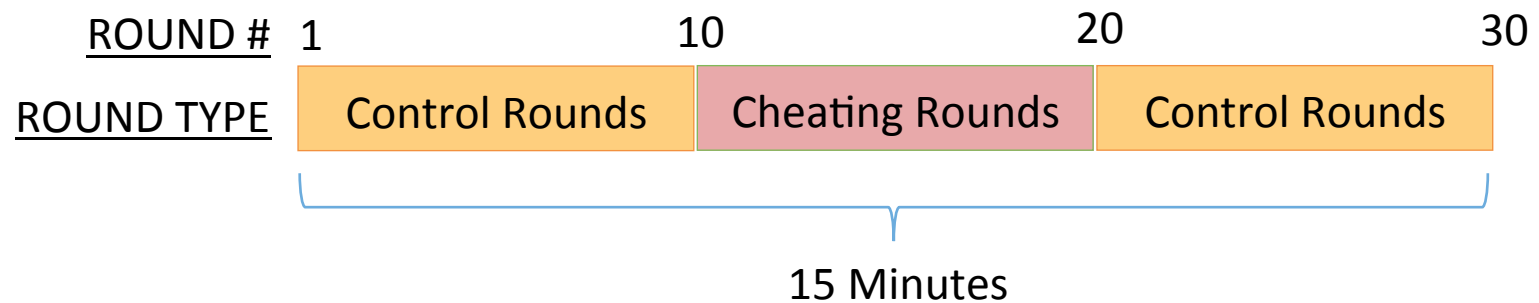
PARTICIPANT THROWS





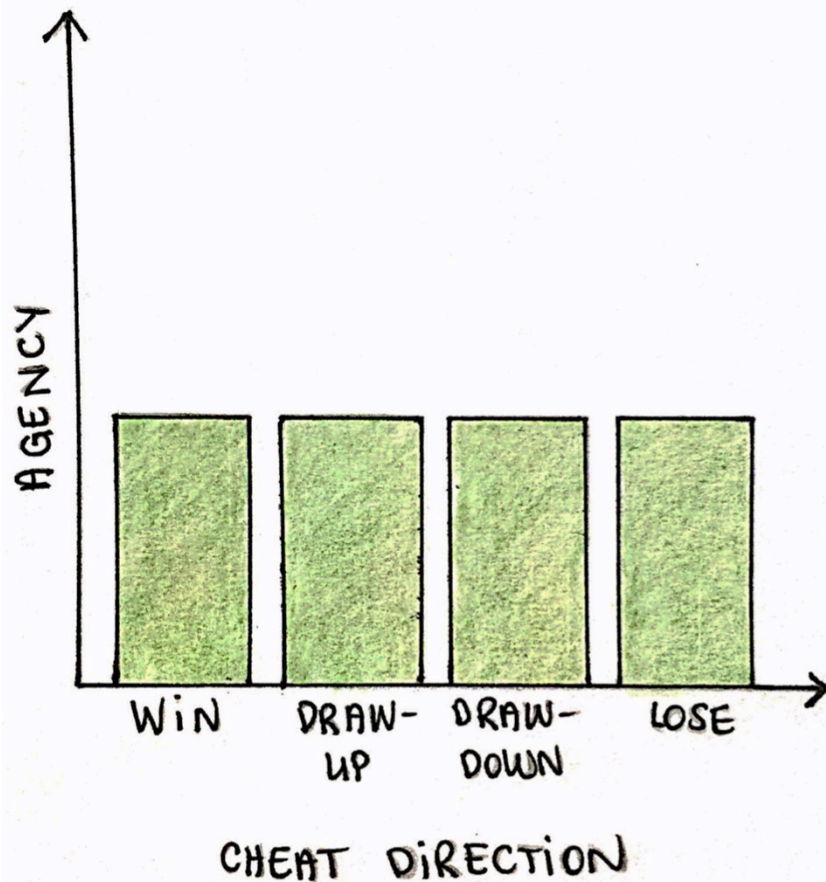
EXPERIMENTAL DETAILS

- 83 participants
- Between-participant design
- 30 rounds of rock-paper-scissors
 - 2 cheat occurrences in the “Cheating Rounds”

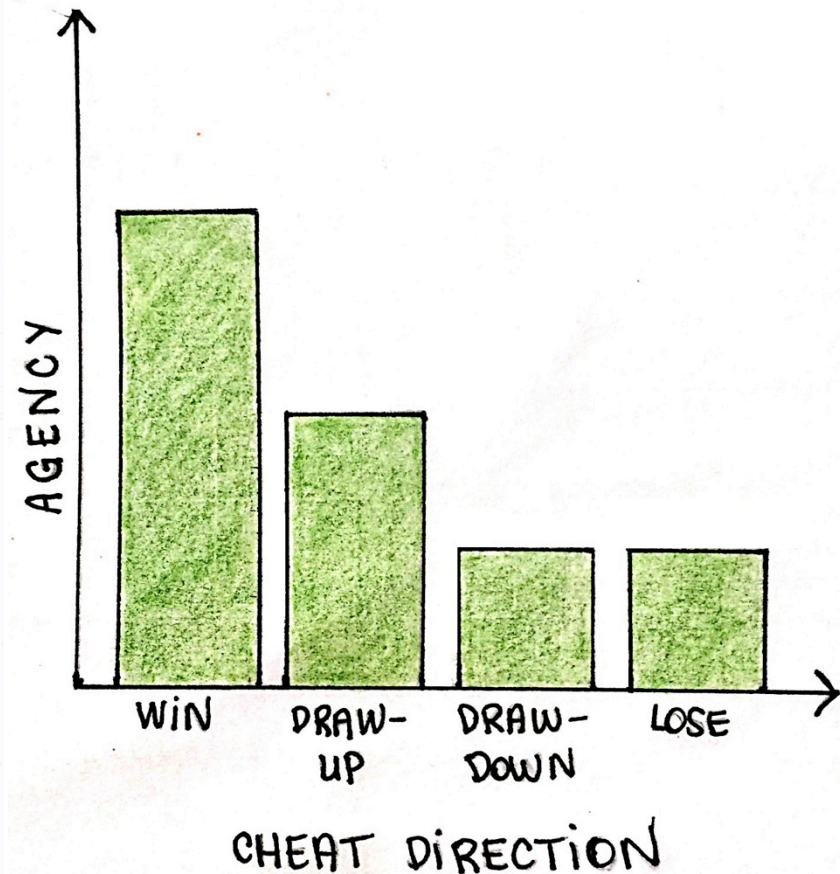


EXPECTATIONS OF RESULTS

If the cause is motion...

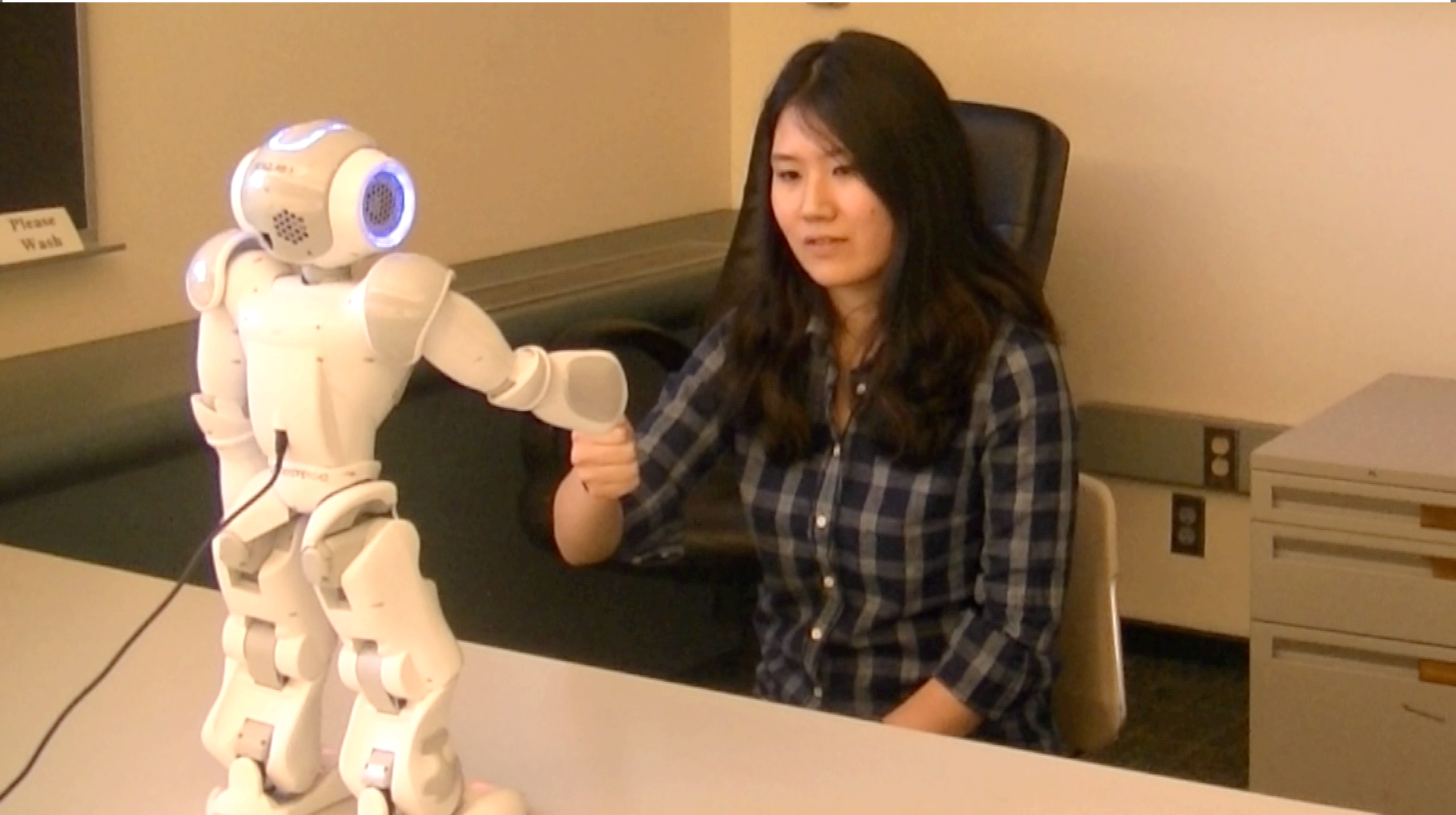


If the cause is a high-level cognitive effect...





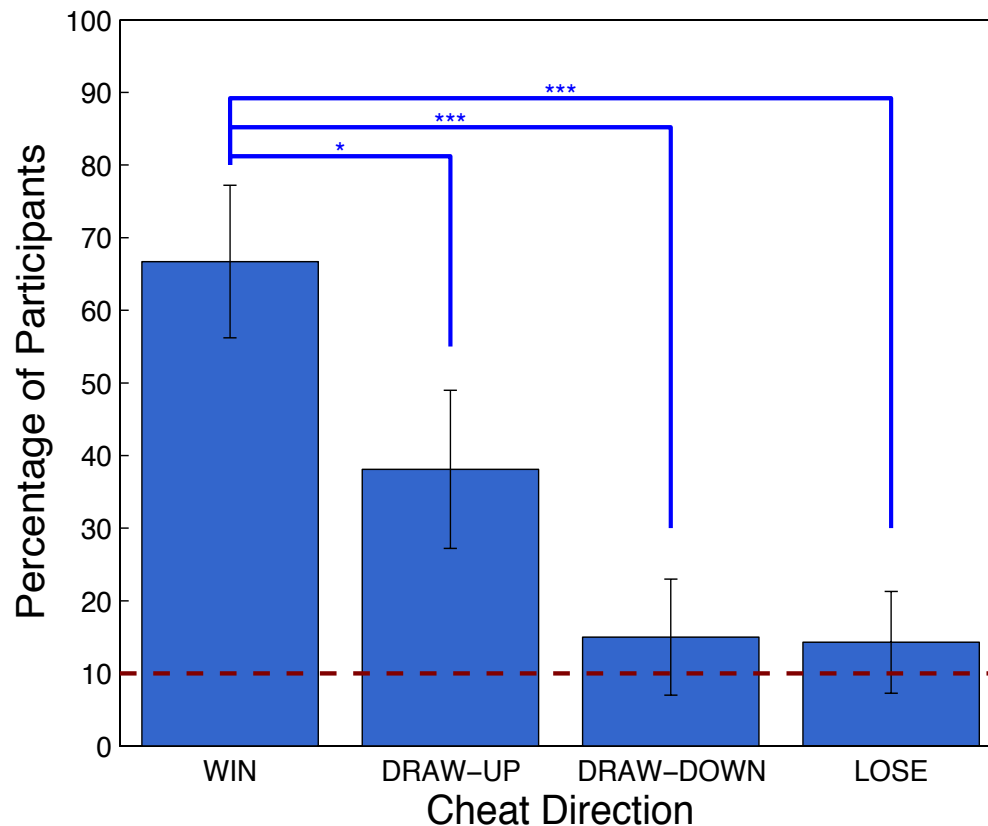
PARTICIPANT REACTIONS





PARTICIPANT ENGAGEMENT

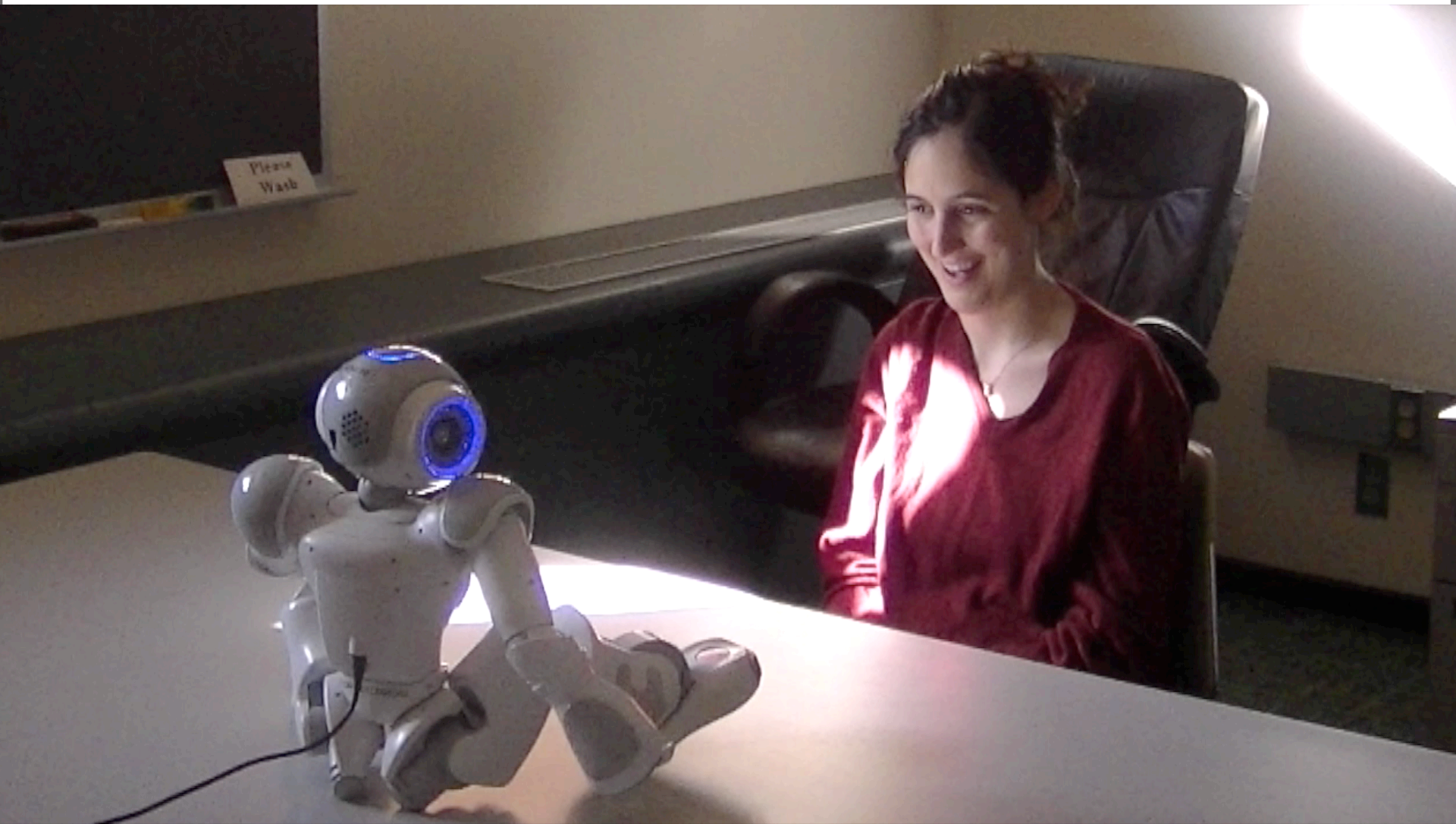
Emitted an Utterance Following Gesture Change



Baseline
utterance
prevalence
in non-
cheating
rounds



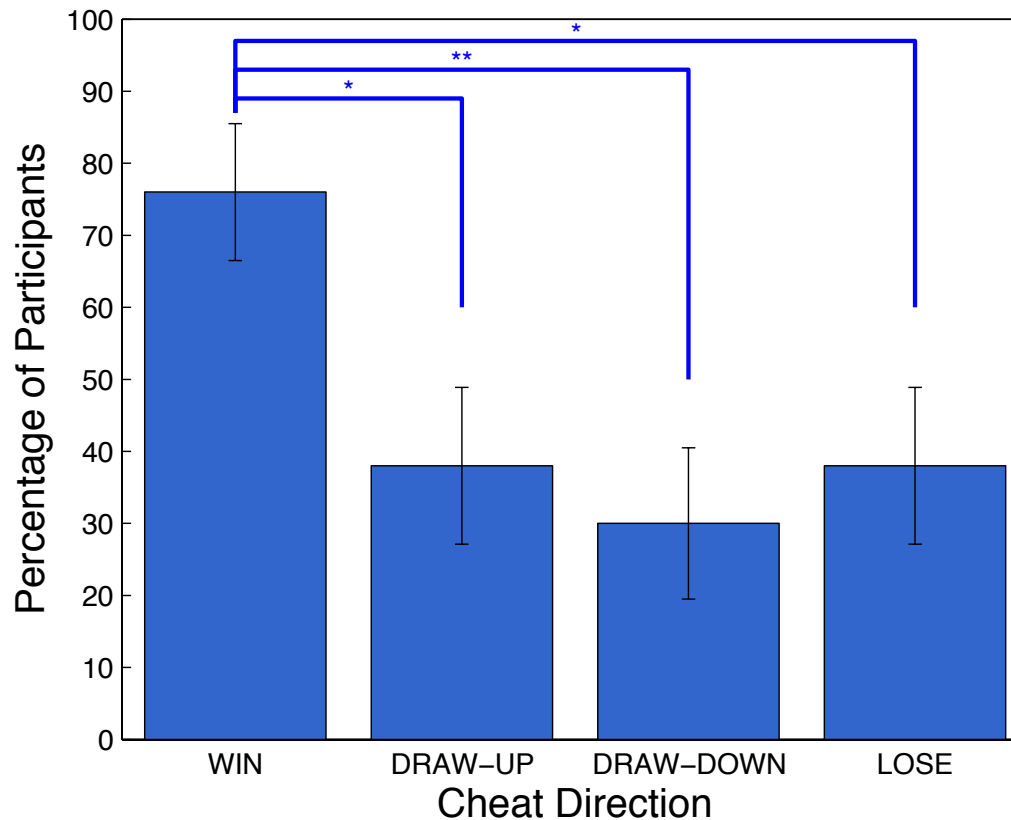
LASTING EFFECTS





CHEAT SALIENCE

Volunteered that the Robot Changed its Gesture in the Post-Study Response

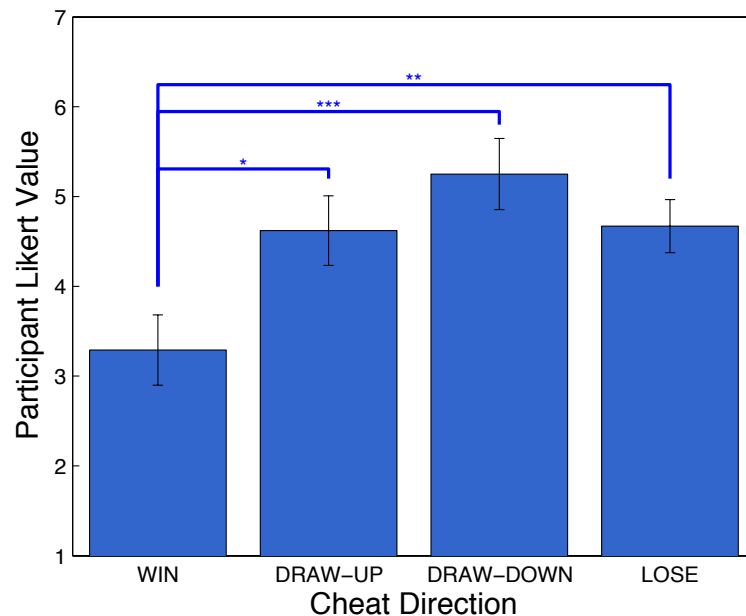


* represents $p < 0.05$, ** represents $p < 0.01$, *** represents $p < 0.001$. Error bars represent standard error.

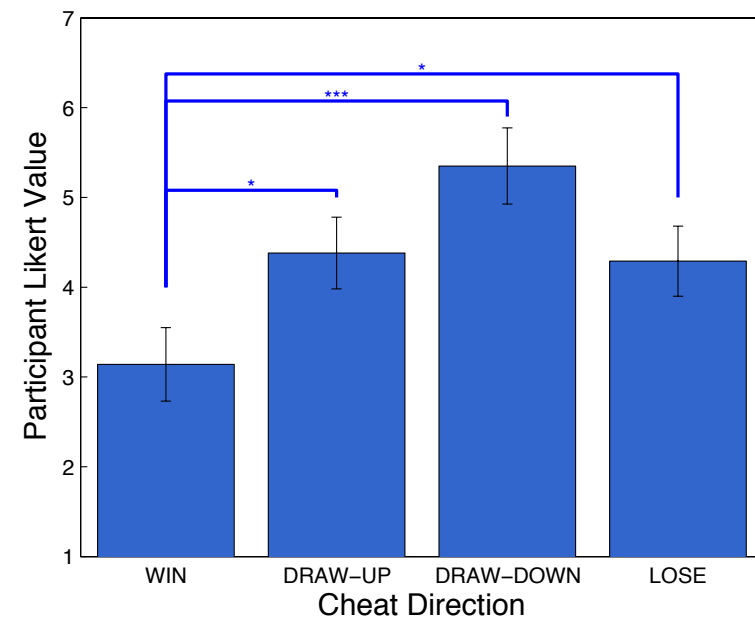


ATTRIBUTIONS – FAIRNESS AND HONESTY

“Fair” Likert Question



“Honest” Likert Question



* represents $p < 0.05$, ** represents $p < 0.01$, *** represents $p < 0.001$. Error bars represent standard error.



EVIDENCE THAT ROBOTS TRIGGER A CHEATING DETECTOR IN HUMANS

- Cause of effect in Short et al.
 - Not the added complexity of the motion
 - Not the rule-breaking behavior
 - **Instead, it is specifically the adversarial cheat**
- In line with expected results of cheating detector that is triggered by robots



THANK YOU!

**MORE RESULTS AND QUESTIONS AT POSTER
SESSION**