Cheating Robots

Manipulating Perceptions of Robot Agency

Alexandru Litoiu and Daniel Ullman
Yale University

Introduction

How do manipulations of a **robot's agency**, via manipulations of behavior and cognitive abilities, affect people's interactions with the robot?

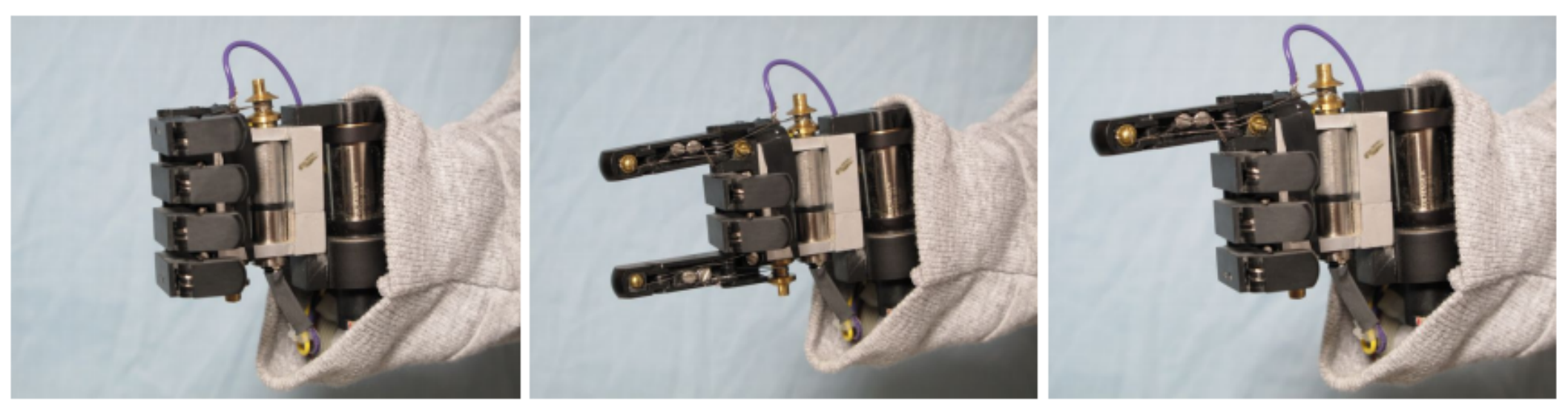
We aim to analyze **human perceptions and behavior** resulting from cheating scenarios to better understand how **apparent agency** affects **human-robot interactions**.

No Fair!! An Interaction with a Cheating Robot

Short et al., HRI 2010

Using a humanoid robot and a simple children's game, we examined the degree to which **variations in behavior** result in **attributions of mental state and intentionality**.

Participants played "rock-paper-scissors" against a robot that either played fairly, cheated verbally by **declaring itself the winner**, or cheated actively by **changing its gesture after seeing its opponent's play**.



Nico making rock, paper, scissors gestures

The humanoid robot, Nico

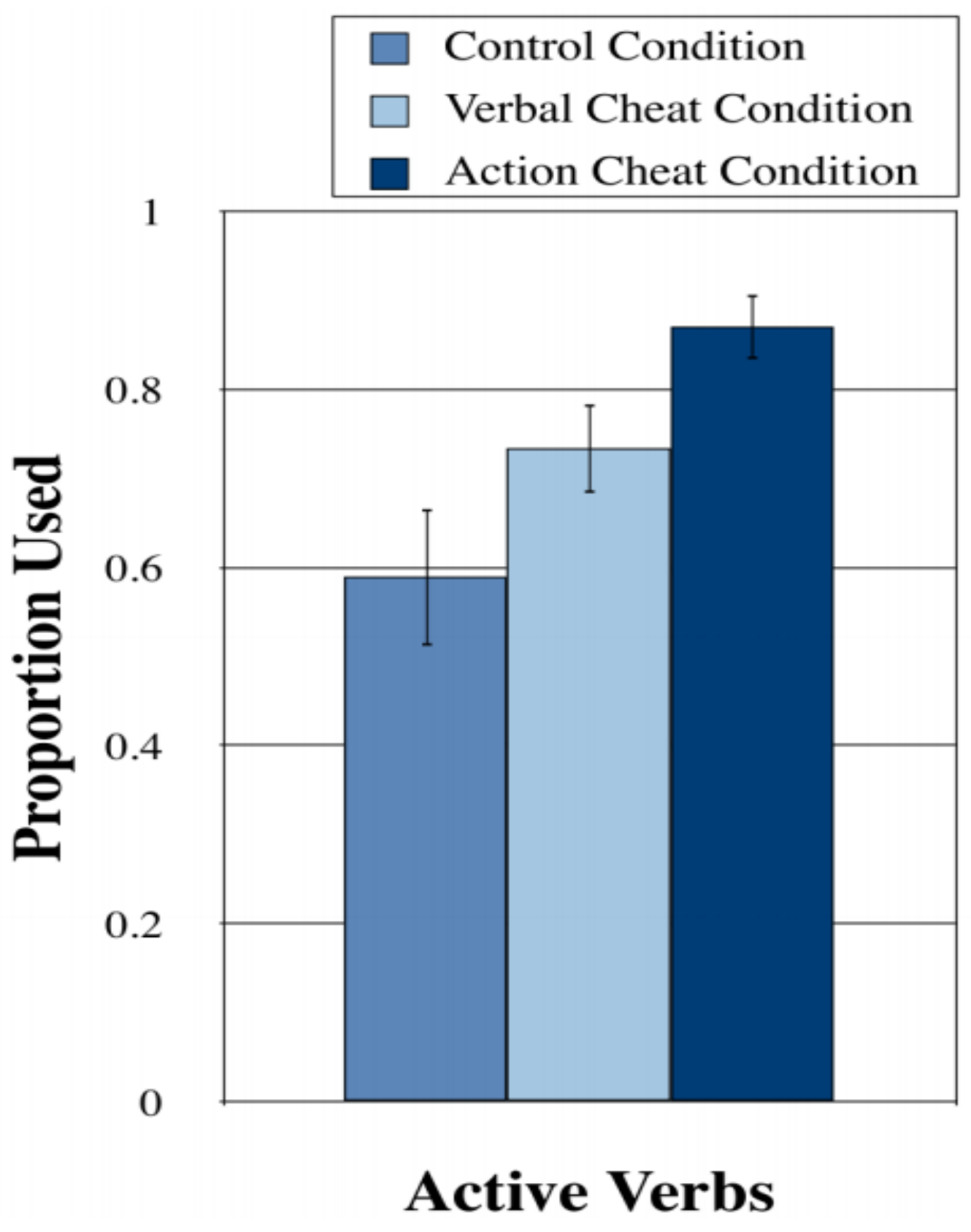
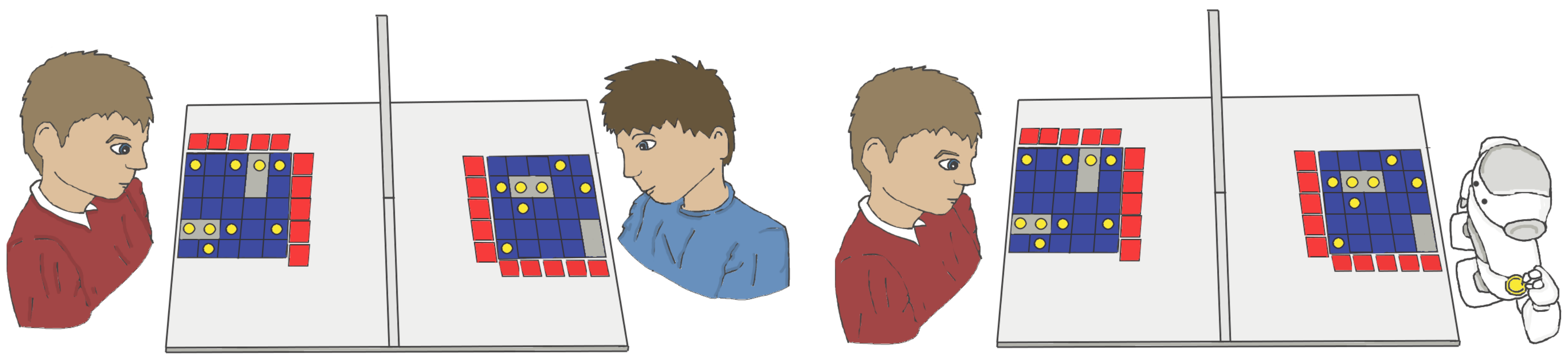


Figure 1. Proportion of active voice verbs used to describe Nico's actions in each condition.

Participants in the action cheat group use a higher ratio of active voice to passive voice verbs than those in the other two groups. There was a significant interaction between study group and this ratio, $F(2,49) = 6.686$, $p = 0.003$.

Smart Human, Smarter Robot



We investigated to what extent the **type of agent** (human or robot) and the **type of behavior** (honest or dishonest) affected perceived agency and trustworthiness in the context of a competitive game.

The human and robot in the dishonest manipulation received lower attributions of trustworthiness as predicted, but the **robot was perceived to be more intelligent than the human**.

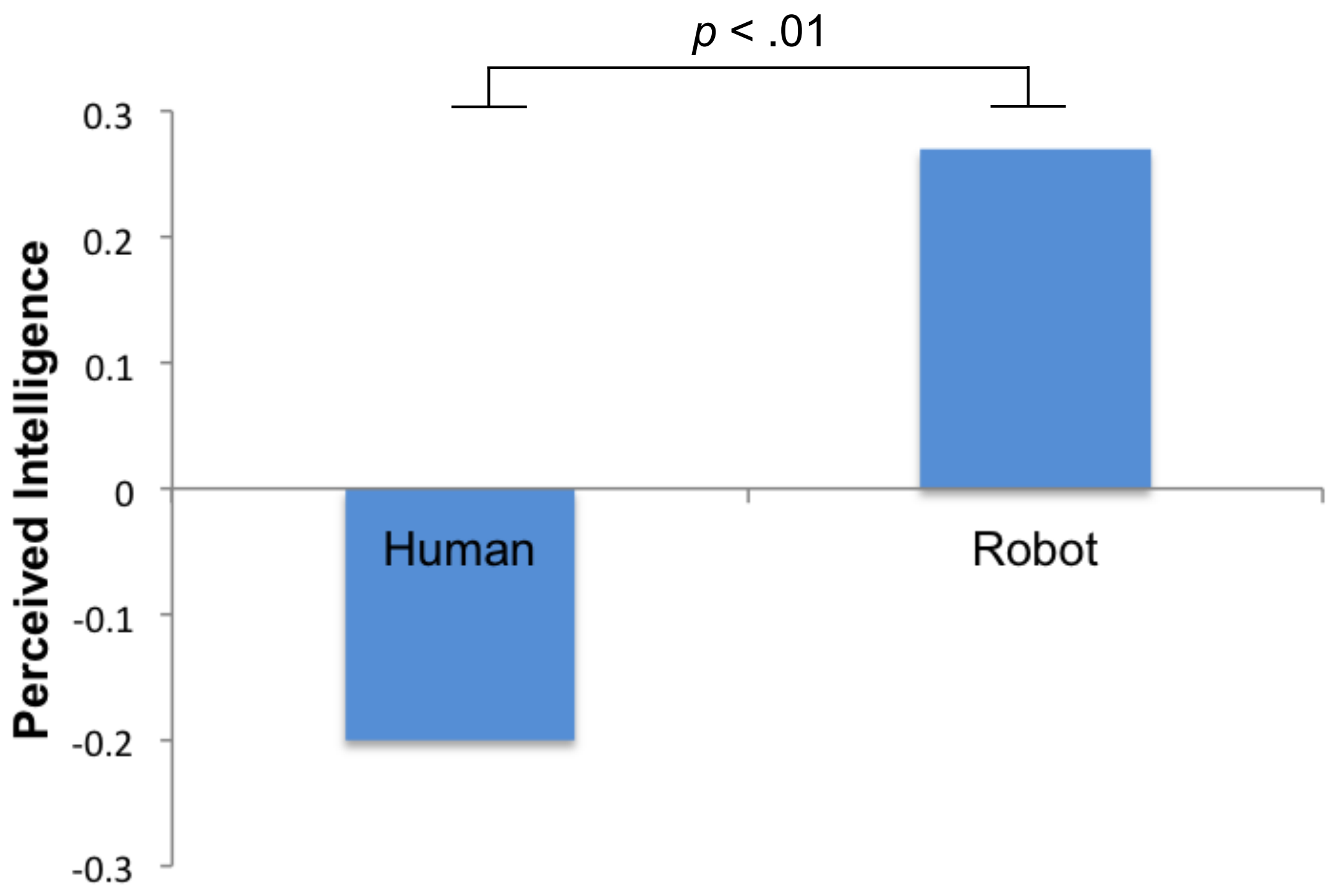


Figure 2. Mean difference of perceived intelligence between dishonest and honest manipulations.

Participants rated the **robot higher on intelligence than the human**, $F(1, 177) = 8.89$, $p < .01$, $\eta_p^2 = .05$.

People perceived the **robot as more intelligent when it was dishonest**, while the **human was rated as less intelligent when dishonest**. There was a significant interaction effect between player type and behavior type, $F(1, 177) = 6.63$, $p = .01$, $\eta_p^2 = .04$.

Disambiguating Attributions of Intelligence and Agency

We aim to disambiguate the following possible **causes of increased attributions of agency to a cheating robot**. Is this phenomenon attributable to participants perceiving **added complexity** (Complexity), **perceiving intentionality** (Goal Detection), or perceiving the **specific intention, to beat him or her** (Cheat Detection)?

In a between-participant design, we disambiguate between these causes by testing the cases of the **robot cheating to win** (WIN), **cheating to lose** (LOSE), **cheating to tie from a winning position** (TIE-1D) and **cheating to tie from a losing position** (TIE-1U).

Nao robot making rock, paper, scissors gestures

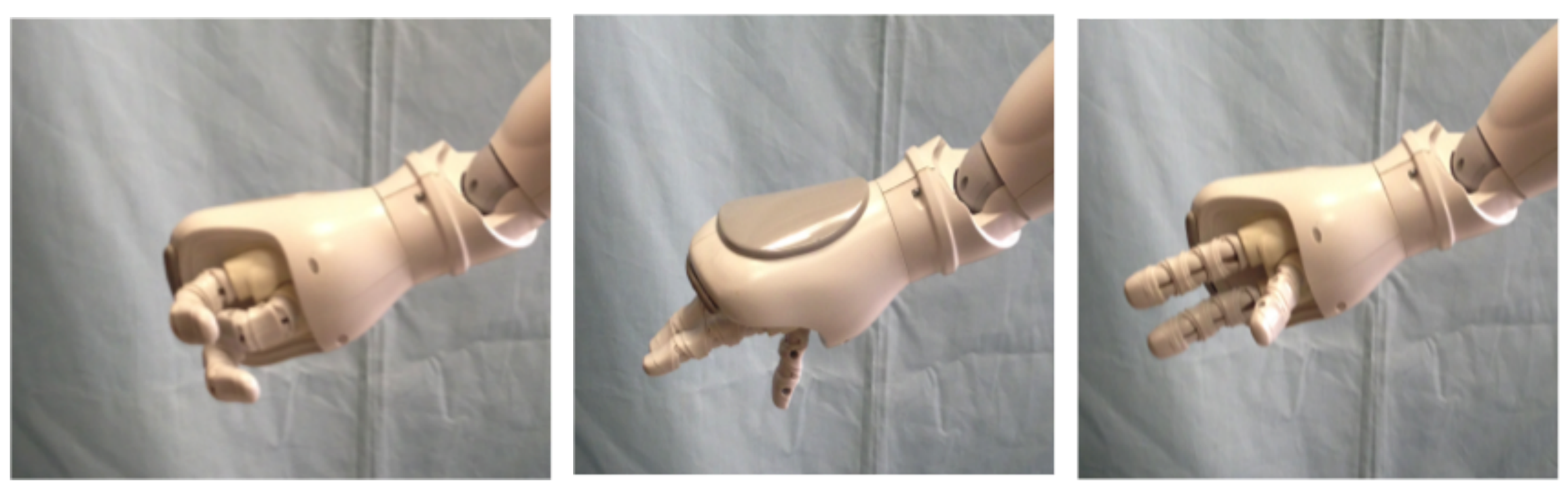


Figure 3. Explanatory causes of high attributions of agency, and their expected results in the strength of agency in the four testing conditions.

Cause	WIN	TIE-1U	TIE-1D	LOSE
Complexity	Strong	Strong	Strong	Strong
Goal Detection	Strong	Weak	Weak	Weak
Cheat Detection	Strong	Weak	None	None

We recorded 80 participants, split evenly across the four conditions. We collected Likert responses, long answer responses, as well as heart rate and galvanic skin response throughout the experiment.

The data has not yet been analyzed - this experiment is ongoing.

This material is based upon work supported by the National Science Foundation award #1117801 (Manipulating Perceptions of Robot Agency) and award #1139078 (Socially Assistive Robots).

scazlab.yale.edu