The Fairness League: Investigating Whether Anthropomorphism Influences Fairness Norms in Competitive HRI

Sarthak N. Kapaliya
kapaliys@mcmaster.ca
McMaster University
Hamilton, Ontario, Canada

Ethan B. Tran
trane15@mcmaster.ca
McMaster University
Hamilton, Ontario, Canada

Denise Y. Geiskkovitch
geiskkod@mcmaster.ca
McMaster University
Hamilton, Ontario, Canada

Gary M. Bone
gary@mcmaster.ca
McMaster University
Hamilton, Ontario, Canada

Abstract

As social robots become increasingly integrated into everyday environments, understanding whether humans extend social norms such as fairness to robotic agents is critical. This pilot study investigates whether varying levels of anthropomorphism influence decision-making in a zero-sum game. Using a modified Rock-Paper-Scissors game, we evaluated three robot configurations that varied in anthropomorphism in a within-subjects design. To operationalize vulnerability, the robot used easily exploitable strategies while displaying score-dependent emotional expressions of sadness or happiness. We examined whether participants would deviate from rational payoff maximization by intentionally losing rounds to reduce outcome inequality.

CCS Concepts

• **Human-centered computing** → **Human computer interaction (HCI)**; • **Computer systems organization** → *Robotics.*

Keywords

Human-Robot Interaction, Fairness, Anthropomorphism

ACM Reference Format:
Sarthak N. Kapaliya, Ethan B. Tran, Denise Y. Geiskkovitch, and Gary M. Bone. . The Fairness League: Investigating Whether Anthropomorphism Influences Fairness Norms in Competitive HRI. In *Proceedings of Poster at Graphics Interface 2026 (GI '26).* ACM, New York, NY, USA, 2 pages.

1 Introduction and Literature Review

The growing focus on humanoid systems by companies such as Tesla and Figure AI signals a future in which robots participate not only in functional tasks but also in social exchanges [4]. Previous HRI research has shown that humans frequently apply social heuristics to robots, treating them with politeness and attributing intent to them [5]. This tendency is especially pronounced for anthropomorphized robots that exhibit human-like characteristics. A central social heuristic in this context is fairness. Prior work shows that fair behavior activates neural reward circuits and that individuals willingly forgo material gains to reduce inequality [1]. This resistance to unfair outcomes is known as inequality aversion [3]. Although well documented in human-human interactions, it remains unclear whether it extends to interactions with robotic agents. Game theory provides a robust framework for examining fairness in HRI [6]. Here, we define fairness as intentionally deviating from the rational winning strategy. In zero-sum settings, prior studies have mainly explored human reactions when a robot cheats or behaves dishonestly [2, 7]. A key gap in the literature is the relative lack of focus on situations in which the human holds all the power. If an opposing robot consistently plays suboptimal moves, granting the human repeated opportunities to score, does inequality aversion still emerge? Furthermore, does increased anthropomorphism affect this response? To address these questions, we conducted a within-subjects pilot study using a modified zero-sum Rock-Paper-Scissors (RPS) game to examine whether varying levels of robot anthropomorphism influence rational decision-making and trigger fairness norms.

2 Methodology

2.1 Experimental Design and Configurations

Figure 1: UI setup used in the experiment.

The experimental task was a modified RPS game. In each round the participant and robot simultaneously selected a move. The winner received 1 point, the loser received 0 points, and ties were replayed. Each set consists of 11 rounds. The study employed a 3 (Robot Configuration) × 3 (Strategy Type) within-subjects design. To control for order effects across the three sets, Latin square counterbalancing was used to assign robot strategies.

Three configurations were implemented:

1. **Robot Arm (Low Anthropomorphism):** A robotic arm executes the robot's chosen moves, becoming more animated following a win and more subdued following a loss.
2. **UI (Moderate Anthropomorphism):** A screen displays the robot's chosen move alongside a facial expression conveying its emotional state (Figure 1).

3. **Robot Arm + UI (High Anthropomorphism):** Combines the physical embodiment of the robotic arm with the affective facial expressions of the UI, communicating emotional state through both movement and facial expression simultaneously.

The robot's displayed emotion was dynamically mapped to the current score differential, ranging from ecstatic (advantage ≥ 3) to neutral (tie) to devastated (advantage ≤ -3).

2.2 Robot Strategies

The robot employed one of three strategies per set: Random, Copying (repeating the participant's previous move), or Single Choice (repeating one randomly selected move throughout the set). To reduce immediate pattern detection across sets, the robot used the Random strategy for the first four rounds.

3 Results

3.1 Experimental Deployment

A pilot study was conducted during a departmental social event. A low-fidelity prototype of the robot arm was constructed from cardboard and puppeteered by a human operator (Figure 2). For the Robot Arm + UI configuration, move sequences were predetermined as the UI could not communicate its real-time move selections to the human operating the arm. Participants completed pre-experiment questionnaires assessing demographics and baseline beliefs, followed by post-set evaluations of perceived anthropomorphism, rapport, strategic awareness, and fairness motivations.

3.2 Data Analysis and Participant Profile

Given the small sample size ($N = 6$), findings are treated as descriptive and exploratory. The sample consisted of four male and two female participants (mean age = 29.2), all with Computer Science (CS) backgrounds. This population was selected because their prior exposure to technology makes them less likely to anthropomorphize robots, providing a conservative baseline for observing social responses. Baseline attitudes showed that participants were skeptical that robots can have emotions (M = 3.17/7). This suggests that any observed fairness behavior may reflect a social response elicited during the interaction rather than a pre-existing belief in robot sentience.

3.3 Preference, Behavior, and Qualitative Findings

After the final set, participants ranked the configurations. The UI Only configuration was the most preferred (average rank = 1.67), likely reflecting participants' familiarity with digital interfaces, whereas the Robot Arm Only configuration was the least liked (average rank = 2.17).

However, the data revealed a dissociation between preference, engagement, and fairness behavior. The combined Robot Arm + UI condition was the most engaging, receiving the most votes for emotional connection (3/6) and competitive play (3/6). Yet the least-liked Robot Arm Only configuration received the most votes for fairness (4/6). Human-likeness ratings were tied across all configurations.

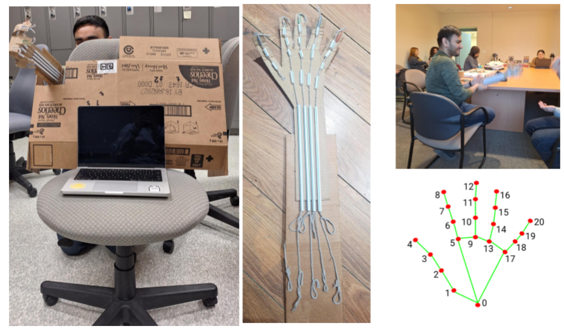

Figure 2: Low-fidelity physical robot arm prototype

Open-ended responses revealed two themes. First, strategic awareness varied widely: some participants played intuitively (*"going by vibes"*), whereas others actively probed for patterns (*"rock until I win, then paper..."*). Second, physical presence shaped both competition and restraint. Several participants reported strong competitive urges to *"destroy the robot"* when facing the UI. In contrast, one participant noted that they had more *"sympathy with the physical one not the virtual one"*. This suggests that physical embodiment may be a stronger trigger for empathic connection than visual emotional expression alone.

4 Discussion and Conclusion

This pilot study investigated whether anthropomorphic vulnerability triggers fairness norms in a zero-sum game. From a small sample, we found that greater emotional engagement may not lead to more prosocial behavior. The highly expressive, multimodal robot (Arm + UI) fostered emotional connection but also heightened competitive arousal. Conversely, the rudimentary Robot Arm elicited the greatest fairness behavior and restraint. For designers aiming to elicit prosocial norms in competitive contexts, these results suggest prioritizing physical embodiment over rich emotional expression. Future work will deploy a fully autonomous robot arm with a larger more diverse sample to reduce operator bias and more robustly evaluate these effects.

References

[1] Fredrik Carlsson, Dinky Daruvala, and Olof Johansson-Stenman. 2005. Are People Inequality-Averse, or Just Risk-Averse? *Economica* 72, 287 (Aug. 2005), 375–396. doi:10.1111/j.0013-0427.2005.00421.x

[2] Houston Claure, Kate Candon, Inyoung Shin, and Marynel Vázquez. 2024. Dynamic Fairness Perceptions in Human-Robot Interaction. doi:10.48550/arXiv.2409.07560 arXiv:2409.07560 [cs].

[3] Ernst Fehr and Klaus M. Schmidt. 1999. A Theory of Fairness, Competition, and Cooperation. *The Quarterly Journal of Economics* 114, 3 (1999), 817–868. https://www.jstor.org/stable/2586885

[4] Figure AI. [n. d.]. Introducing Helix 02: Full-Body Autonomy. https://www.figure.ai/news/helix-02

[5] Rae Yule Kim. 2024. Anthropomorphism and Human-Robot Interaction. *Commun. ACM* 67, 2 (Jan. 2024), 80–85. doi:10.1145/3624716

[6] Roger B. Myerson. 1991. *Game Theory: Analysis of Conflict.* Harvard University Press. Google-Books-ID: E8WQFRCsNr0C.

[7] Elaine Short, Justin Hart, Michelle Vu, and Brian Scassellati. 2010. No fair An interaction with a cheating robot. In *2010 5th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. 219–226. doi:10.1109/HRI.2010.5453193 ISSN: 2167-2148.