Multiplayer Space Invaders: A Platform for Studying Evolving Fairness Perceptions in Human-Robot Interaction

Houston Claure
houston.claure@yale.edu
Yale University
New Haven, CT, USA

Kate Candon
kate.candon@yale.edu
Yale University
New Haven, CT, USA

Olivia Clark
olivia.clark@aya.yale.edu
Yale University
New Haven, CT, USA

Marynel Vázquez
marynel.vazquez@yale.edu
Yale University
New Haven, CT, USA

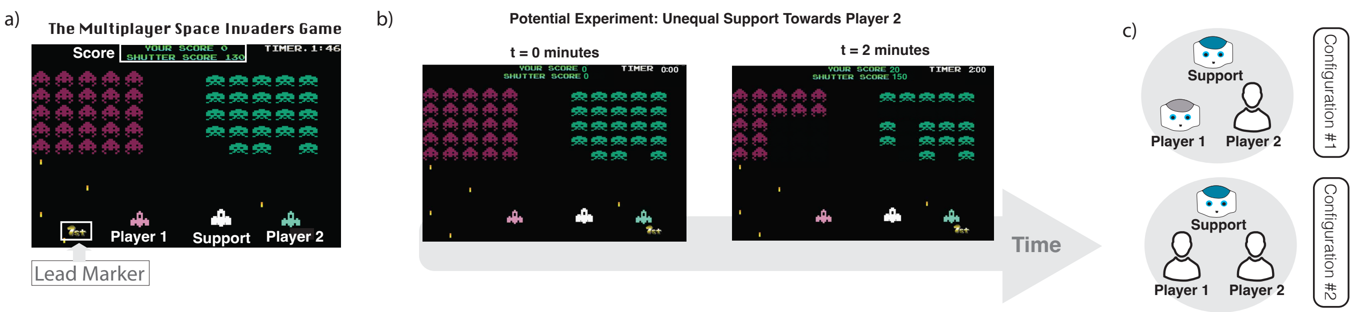

Figure 1: (a) The Multiplayer Space Invaders game is a platform where two players aim to earn points by eliminating as many enemies as possible. A third player, in a supporting role, is responsible for evenly or unevenly distributing assistance between the two main players. (b) An example of a possible experiment where one player receives a disproportionate level of assistance from the supporting player over time. (c) Two possible configurations involving multiple robots or humans that could influence fairness judgments.

ABSTRACT

Current methods of measuring fairness in human-robot interaction (HRI) research often gauge perceptions of fairness at the conclusion of a task. However, this methodology overlooks the dynamic nature of fairness perceptions, which may shift and evolve as a task progresses. To help address this gap, we introduce a platform designed to help investigate the evolution of fairness over time: the Multiplayer Space Invaders game. This three-player game is structured such that two players work to eliminate as many of their own enemies as possible while a third player makes decisions about which player to support throughout the game. In this paper, we discuss different potential experimental designs facilitated by this platform. A key aspect of these designs is the inclusion of a robot that operates the supporting ship and must make multiple decisions about which player to aid throughout a task. We discuss how capturing fairness perceptions at different points in the game could give us deeper insights into how perceptions of fairness fluctuate in response to different variables and decisions made in the game.

HRI '24 Companion, March 11–14, 2024, Boulder, CO

ACM ISBN 979-8-4007-0323-2/24/03
https://doi.org/10.1145/3610978.3640687

CCS CONCEPTS

• **Human-centered computing → Empirical studies in collaborative and social computing**; **Empirical studies in HCI**.

KEYWORDS

Human-Robot Interaction; Fairness; Resource Allocation

ACM Reference Format:
Houston Claure, Kate Candon, Olivia Clark, and Marynel Vázquez. 2024. Multiplayer Space Invaders: A Platform for Studying Evolving Fairness Perceptions in Human-Robot Interaction. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction (HRI '24 Companion), March 11–14, 2024, Boulder, CO, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3610978.3640687

1 INTRODUCTION

The increased use of robots across various domains, such as hospitals and schools, raises questions about how robots should handle situations in which fairness considerations are necessary. Consider a robotic teacher in a classroom. What would happen if the robotic teacher focused a disproportionate amount of attention on a few students? What effects would that have on the way the students both perform and behave with one another? What is the best strategy for the robot to fairly distribute support to the students?

These considerations highlight the fundamental role of fairness in settings where humans interact and collaborate with one another. When individuals perceive unfair treatment, their reactions can be intense and sometimes irrational [17]. This includes going far as making decisions that go against their own rational self-interest [2]. More recent work has highlighted how even unfair distribution decisions from AI algorithms can trigger fairness judgments leading to strong responses from humans [6]. Taken together, these works emphasize human sensitivity to the allocation of resources, particularly in relation to others. This extends even when the allocation decisions are made by AI systems.

Yet, fairness remains underexplored in HRI. This can be attributed to the lack of viable tools to study fairness. The various decision-making scenarios encountered by robots, each governed by its unique set of rules and interpretations of fairness [11], further complicates the development of such tools. Take, for instance, the scenario involving a robotic teacher. In situations where the robot allocates more resources to students who require additional help to reach the same level as their peers, this unequal resource distribution might be considered fair. Conversely, if the robot favors students who do not need extra assistance, its actions could be perceived as unfair. In order to recreate and investigate such scenarios, researchers often resort to using vignettes [3] and online simulations [10] to gauge human judgments of perceived unfair actions by robots. However, these methods capture the fairness judgments of humans only at the end of an interaction, overlooking the nuanced and dynamic nature of fairness perceptions that evolve over time [9, 15]. This limitation highlights the need for more sophisticated tools that can capture the finer details of how perceptions of fairness change throughout the course of an interaction with robots.

In this work, we present a new tool for exploring fairness in HRI. The Multiplayer Space Invaders game is a platform where two main players focus on obtaining the highest score while a third player focuses on supporting the main players. This dynamic mirrors real-world situations where a robot must balance its assistance among multiple participants in a given task. We envision this platform being used to expand our understanding of fairness by allowing researchers to expose people to different contexts where resources are distributed unequally. This game requires the supporting player to continuously make decisions about which player to assist throughout the course of the task. Tracking how unfair humans judge the decisions of the support player to be allows us to explore fairness not as a static concept but as a dynamic one that changes over time. We believe that this tool will significantly enhance our understanding of fairness in HRI, providing researchers with the means to create and study a variety of contexts and interactions.

2 FAIRNESS IN HRI

People deeply care how they are treated in relation to others in similar situations [1]. With robots increasingly in situations where they will have to make decisions about the distribution of resources, questions have been raised on how humans will judge a robot's decisions [4]. This is especially relevant in multi-party settings such as hospitals, schools, and manufacturing floors, where robots must decide on the distribution of both tangible resources like tools and intangible resources like attention or gaze. Adhering to human standards of fairness is crucial in these contexts, as deviations can have negative social repercussions [16]. For instance, how a robot allocates resources among group members can profoundly affect their sense of inclusion and participation. Research by Mutlu et al. has highlighted that unequal visual attention from a robot can induce feelings of exclusion in group members [20]. Similarly, Jung et al. demonstrated that unequal resource distribution can lead to interpersonal tensions within a group [16]. These studies indicate that fairness violations by robots not only influence human perceptions of the robots but also significantly impact the dynamics of human interactions. Furthermore, these results suggest that robots violating fairness norms shape not only the way people perceive the robot [10] but also how they interact with one another [16].

2.1 Tools for Investigating Fairness in HRI

In order to explore fairness in HRI, researchers have drawn tools from fields such as psychology and have repurposed well-known video games. Some HRI researchers have leveraged variations of established economic games [13, 14, 21, 22, 24], such as the prisoner's dilemma [5] and the ultimatum game [23], to negotiate the distribution of a reward between a human and a robot. Others have used video vignettes of robots committing unexpected behaviors to expose participants to unique scenarios where a robot's behavior can be perceived as unfair [3]. More recently, the use of video games, such as Tetris, has been shown as a viable way to study fairness perceptions within multi-human groups[10].

However, across all of these works, fairness judgments of a robot's actions are captured at the conclusion of the interaction with a robot. This method of measuring fairness ignores the work in organizational psychology which highlights that a person's perceptions of fairness evolve as they gain more information about the context in which they find themselves [15]. Motivated by prior work in HRI on dynamic trust [9, 12], we argue that in order to gain a more nuanced understanding of these evolving perceptions, it is crucial to design experiments that involve multiple decision-making points. By doing so, researchers can observe and measure how participants' views on fairness change throughout the interaction with the robot. This approach can offer a more detailed and accurate picture of fairness in HRI, reflecting the continuous and evolving nature of human perceptions in response to a robot's decisions.

3 THE MULTIPLAYER SPACE INVADERS GAME

We introduce the Multiplayer Space Invaders (Fig. 1a) game as a viable platform to explore how fairness perceptions evolve in a multi-party setting. This game extends the well-known single-player Space Invaders game, which has a rich history in psychology [19] and HRI [7, 8, 18], to a multiplayer setting, which can be either collaborative or competitive. Developed using the Phaser game framework, which is based on HTML5, our game is designed to run seamlessly in a web browser. Our game involves three players, and each player commands an individual spaceship differentiated by color. The adversaries in the game are represented as alien spaceships and organized into two distinct clusters on the display. The alien adversaries were programmed to reappear after being eliminated, so enemies could be destroyed continuously until the game ended. Two players (red and green spaceships) are tasked with eliminating as many enemies on their respective sides as possible. A third player (white spaceship) can support one of the two players by moving to help eliminate the cluster of enemies on the left or right side. Importantly, we have designed the game to be compatible with robotic integration. Leveraging the Robot Operating System (ROS), we can easily incorporate a physical robot participating as a player in the game. This feature allows for real-time interaction between a human and a robot player, responding to events within the game. This integration not only enhances the gaming experience but also provides a rich environment for studying dynamic human-robot interactions and fairness perceptions.

4 FUTURE RESEARCH DIRECTIONS

We envision the Multiplayer Space Invaders platform being used for investigating various questions around the social impact of a robot's resource distribution decisions. One of the primary features that sets this tool apart in fairness research is its provision of multiple decision points. Here, the player in the supportive role (controlling the white ship) must continually decide how to allocate support between the other players. By having participants continuously report their fairness judgments towards the support player's distribution of assistance, we can get a clearer understanding of how fairness perceptions can evolve. To facilitate this exploration, we can have a robot control the white ship, with two human participants as players 1 and 2. We can envision simple experimental designs where the robot disproportionately allocates its support towards one player (Fig 1b), as in [16]. We can also envision more complex designs where a robot supports the player who is in need of the most support.

The Multiplayer Space Invaders game also provides a new platform where researchers can evaluate various algorithms that aim to capture fairness in a robot's decision-making. This game extends the groundwork laid by prior work, such as the study by Claure et al.[10], which demonstrated the effectiveness of using games as a tool for assessing fair algorithms for HRI. Having such a platform makes it easier to compare the effects of different algorithms aimed at including factors such as fairness into a robot's decisions and capture how humans respond to the robot's actions.

Furthermore, this setup allows for future experiments where we examine the different variables that shape fairness perceptions. For instance, we can examine fairness perceptions in different contexts, such as competitive versus cooperative gaming environments. It is plausible that the context of the interaction – whether players are competing against each other or collaborating – could significantly influence how fairness is perceived with regards to the robot's actions. Moreover, by varying the configuration of the agents and the nature of their interaction (Fig 1c), we can uncover how different scenarios impact perceptions of fairness. This approach not only enables us to track fairness perceptions over time but also allows us to measure various performance metrics within the game. Finally, the Multiplayer Space Invaders Game provides a platform where we can investigate how different verbal and nonverbal behaviors from a robot will affect how fairness judgments are formed. If a robot demonstrates more transparency in situations where it acts unfairly, we can hypothesize that this would impact how humans judge the robot's actions. Consequently, this provides us with valuable insights into how perceptions of fair or unfair treatment by a robot can affect human performance and behavior in a game setting.

5 CONCLUSION

In this paper, we introduce the Multiplayer Space Invaders game as a platform for capturing fairness perceptions over time in HRI contexts. The platform is a three-player game where two players work to eliminate enemy spaceships while a third support player is charged with distributing its support to the two other players. This recreates the well-observed scenario in multi-party HRI contexts involving a robot that must allocate resources across humans [16, 20]. Using this game, researchers can explore how perceptions of fairness evolve over time from the perspective of the player who is benefiting from the support and from the player who suffers from the lack of support. Additionally, this platform can be used to answer questions about the social consequences of a robot disproportionately distributing its support towards one player. We hope this platform enables research that facilitates more intentional decision-making about resource allocation, considering dynamic perceptions of fairness.

6 ACKNOWLEDGEMENTS

The authors thank the NSF for partially supporting this work under Grant No. IIS-2106690. Any opinions and findings presented in this paper are those of the authors and do not necessarily reflect the views of the National Science Foundation.

REFERENCES

[1] Sheldon Alexander and Marian Ruderman. 1987. The role of procedural and distributive justice in organizational behavior. *Social justice research* 1, 2 (1987), 177–198.
[2] Maureen L Ambrose, Mark A Seabright, and Marshall Schminke. 2002. Sabotage in the workplace: The role of organizational injustice. *Organizational behavior and human decision processes* 89, 1 (2002), 947–965.
[3] Thomas Arnold and Matthias Scheutz. 2018. Observing robot touch in context: How does touch and attitude affect perceptions of a robot's social qualities?. In *2018 13th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 352–360.
[4] Edmond Awad, Sohan Dsouza, Richard Kim, Jonathan Schulz, Joseph Henrich, Azim Shariff, Jean-François Bonnefon, and Iyad Rahwan. 2018. The moral machine experiment. *Nature* 563, 7729 (2018), 59–64.
[5] Robert Axelrod. 1980. Effective choice in the prisoner's dilemma. *Journal of conflict resolution* 24, 1 (1980), 3–25.
[6] Solon Barocas, Moritz Hardt, and Arvind Narayanan. 2017. Fairness in machine learning. *Nips tutorial* 1 (2017), 2.
[7] Kate Candon, Zoe Hsu, Yoony Kim, Jesse Chen, Nathan Tsoi, and Marynel Vázquez. 2022. Perceptions of the Helpfulness of Unexpected Agent Assistance. In *Proceedings of the 10th International Conference on Human-Agent Interaction*. 41–50.
[8] Kate Candon, Helen Zhou, Sarah Gillet, and Marynel Vázquez. 2023. Verbally Soliciting Human Feedback in Continuous Human-Robot Collaboration: Effects of the Framing and Timing of Reminders. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*. 290–300.
[9] Vivienne Bihe Chi and Bertram F Malle. 2023. People dynamically update trust when interactively teaching robots. In *Proceedings of the 2023 ACM/IEEE International Conference on Human-Robot Interaction. HRI*, Vol. 23. 554–564.
[10] Houston Claure, Yifang Chen, Jignesh Modi, Malte Jung, and Stefanos Nikolaidis. 2020. Multi-armed bandits with fairness constraints for distributing resources to human teammates. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. 299–308.
[11] Jason A Colquitt, Jerald Greenberg, and J Greenberg. 2003. Organizational justice: A fair assessment of the state of the literature. *Organizational behavior: The state of the science* (2003), 159–200.
[12] Munjal Desai, Mikhail Medvedev, Marynel Vázquez, Sean McSheehy, Sofia Gadea-Omelchenko, Christian Bruggeman, Aaron Steinfeld, and Holly Yanco. 2012.

Effects of changing reliability on trust of robot systems. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*. ACM, 73–80.

[13] Cinzia Di Dio, Federico Manzi, Shoji Itakura, Takayuki Kanda, Hiroshi Ishiguro, Davide Massaro, and Antonella Marchetti. 2020. It does not matter who you are: fairness in pre-schoolers interacting with human and robotic partners. *International Journal of Social Robotics* 12, 5 (2020), 1045–1059.

[14] Marlena R Fraune, Steven Sherrin, Selma Šabanović, and Eliot R Smith. 2019. Is human-robot interaction more competitive between groups than between individuals?. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 104–113.

[15] David A Jones and Daniel P Skarlicki. 2013. How perceptions of fairness can change: A dynamic model of organizational justice. *Organizational psychology review* 3, 2 (2013), 138–160.

[16] Malte F Jung, Dominic DiFranzo, Solace Shen, Brett Stoll, Houston Claure, and Austin Lawrence. 2020. Robot-Assisted Tower Construction—A Method to Study the Impact of a Robot's Allocation Behavior on Interpersonal Dynamics and Collaboration in Groups. *ACM Transactions on Human-Robot Interaction (THRI)* 10, 1 (2020), 1–23.

[17] Michael Koenigs and Daniel Tranel. 2007. Irrational economic decision-making after ventromedial prefrontal damage: evidence from the Ultimatum Game. *Journal of Neuroscience* 27, 4 (2007), 951–956.

[18] Jamie Large, Graham Stodolski, and Marynel Vázquez. 2020. Studying Human-Agent Interactions in Space Invaders. In *Proceedings of the 8th International Conference on Human-Agent Interaction*. 245–247.

[19] Michael Mateas. 2003. Expressive AI: Games and Artificial Intelligence.. In *DiGRA Conference*, Vol. 15. Citeseer.

[20] Bilge Mutlu, Toshiyuki Shiwa, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. 2009. Footing in human-robot conversations: how robots might shape participant roles using gaze cues. In *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*. 61–68.

[21] Mayada Oudah, Vahan Babushkin, Tennom Chenlinangjia, and Jacob W Crandall. 2015. Learning to interact with a human partner. In *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 311–318.

[22] Eduardo Benítez Sandoval, Jürgen Brandstatter, Utku Yalcin, and Christoph Bartneck. 2021. Robot likeability and reciprocity in human robot interaction: Using ultimatum game to determinate reciprocal likeable robot strategies. *International Journal of Social Robotics* 13, 4 (2021), 851–862.

[23] Paul G Straub and J Keith Murnighan. 1995. An experimental investigation of ultimatum games: Information, fairness, expectations, and lowest acceptable offers. *Journal of Economic Behavior & Organization* 27, 3 (1995), 345–364.

[24] Debora Zanatto, Massimiliano Patacchiola, Jeremy Goslin, Serge Thill, and Angelo Cangelosi. 2020. Do humans imitate robots? An investigation of strategic social learning in human-robot interaction. In *2020 15th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 449–457.