The Impact of Asymmetric AI Assistance on Decision-Making in Social Dilemmas: A Study on Human Augmentation in Economic Games

Chenxinran Shen
Aalto University
Espoo, Finland
elise.shen007@gmail.com
chenxinran.Shen@aalto.fi

Jurgis Karpus
Ludwig Maximilian University of Munich
Munich, Germany
jurgis.karpus@lmu.de

Thomas Kosch
Humboldt University of Berlin
Berlin, Germany
thomas.kosch@hu-berlin.de

Daniela Fernandes
Aalto University
Espoo, Finland
daniela.dasilvafernandes@aalto.fi

Beatriz Mello
University of Minho
Braga, Portugal
beatriz.mello@psi.uminho.pt

Robin Welsch
Aalto University
Espoo, Finland
robin.welsch@aalto.fi

Steeven Villa
Aalto University
Espoo, Finland
ext-steeven.villa@aalto.fi

ABSTRACT

Unequal access to AI augmentation may disrupt human cooperation. In this study, we address the impact of asymmetric AI assistance on decision-making in cooperative economic games, focusing on how AI augmentation influences trust, cooperation, and perceptions of fairness in scenarios involving augmented and non-augmented players. Using the Trust Game and the Prisoner's Dilemma, we conducted experiments in which participants interacted under varying conditions of AI access. We found that while AI augmentation did not significantly alter overall cooperation rates, it shaped social perceptions: non-augmented players viewed augmented counterparts as more competitive and less warm, predicting less cooperation in the Trust Game. These disparities in perception highlight the potential of AI augmentation to subtly influence human cooperation. On a larger scale, the findings emphasize the importance of designing equitable AI system access to prevent social divides and promote cooperation in AI-augmented societies.

CCS CONCEPTS

• **Human-centered computing → Human computer interaction (HCI)**.

KEYWORDS

Human-Computer Interaction, Human Augmentation, Game Theory

AHs '25, March 16–20, 2025, Abu Dhabi, United Arab Emirates

ACM ISBN 978-1-4503-XXXX-X/25/04
https://doi.org/XXXXXXX.XXXXXXX

ACM Reference Format:
Chenxinran Shen, Jurgis Karpus, Thomas Kosch, Daniela Fernandes, Beatriz Mello, Robin Welsch, and Steeven Villa. 2025. The Impact of Asymmetric AI Assistance on Decision-Making in Social Dilemmas: A Study on Human Augmentation in Economic Games. In *Augmented Humans International Conference (AHs '25), March 16–20, 2025, Abu Dhabi, United Arab Emirates.* ACM, New York, NY, USA, 12 pages. https://doi.org/XXXXXXX.XXXXXXX

1 INTRODUCTION

Artificial Intelligence (AI) is transforming society by reshaping how people work, learn, and interact with the world and other individuals [8], particularly through Large Language Models (LLMs) that enhance human capabilities through applications such as content generation, decision support, or language translation [13, 19, 39].

AI adoption has produced positive effects in sectors like education, where AI supports personalised learning [33, 66, 81], healthcare, by improving clinician and patient experiences [1, 60], and transportation with AI-powered self-driving vehicles [38]. These advancements streamline tasks, improve accessibility, and alter decision-making, driving social change.

Yet, AI also presents challenges to the *social fabric*, ultimately impacting how relationships within society evolve. For example, AI can deepen inequalities across domains such as employment, education, and healthcare by amplifying biases in decision-making processes [27, 64, 67]. For instance, in an employment context, reliance on AI for performance evaluations or hiring decisions could undermine trust between employees and employers, fostering perceptions of unfairness [73] and limiting opportunities [48] for meaningful human interaction. Drawing on historical data, a similar situation emerged with the internet in the early 2000s, where researchers warned about internet inequality reinforcing inequality in opportunities for social participation [11, 26, 79]. By the 2010s, studies revealed that increased mobile penetration had a positive effect on income equality [2, 41, 80]. Consequently, a growing body of work has focused on how access to AI technologies, including LLMs, can impact these inequalities, highlighting the critical importance of equitable access to AI [34].

Despite growing interest in equitable access to AI [29, 68], our understanding of how AI integration shapes social dynamics remains limited. Prior studies on human-AI cooperation have explored how individuals' attitudes and collaboration preferences vary when interacting with humans versus AI systems, often influenced by factors such as perceived similarities and trustworthiness [31]. Additionally, research on human interactions with automated machines—typically pre-programmed by humans has provided foundational insights into how users adapt to and rely on automated decision-making [35, 44]. However, the adaptive and autonomous nature of AI systems introduces new complexities, highlighting the need for a more comprehensive investigation into how these technologies reshape interpersonal relationships and social structures once AI-augmented and non-augmented individuals interact, compete or cooperate.

This research gap is particularly critical given the widening disparity in access to advanced AI technologies, driven by an arms race among corporations and nations to develop more powerful LLMs [21, 31, 45, 54]. Such unequal distribution of AI capabilities risks amplifying existing social and economic inequalities [34], as those with privileged access can disproportionately benefit from enhanced cognitive and decision-making support [48]. This raises urgent questions about whether AI will ultimately serve as an equalizer, deepen societal divisions, or not affect human cooperation. Addressing this issue requires moving beyond traditional human-AI interaction studies.

Game theory provides a robust framework for studying trust [30, 42], cooperation [9, 10], and competition [14] among agents across diverse decision scenarios, reflecting the competing motivations humans often face in their daily interactions with one another. Cooperation, for example, often requires willingness to sacrifice some of one's personal interests for the benefit of the group while exposing oneself to the risk that others may not cooperate in return [31]. Game theorists design economic games to construct such mixed-motive decision scenarios for use in behavioural studies. This is well suited to study AI-augmented versus non-augmented people's cooperative dispositions, since it enables comparisons of new results with well-established benchmarks. Economic games have been also extensively used to investigate people's reduced cooperation with members of their out-group [4]. This is particularly relevant to our research question, as AI augmentation may lead to the stratification of societies into distinct groups of differently augmented people. Here, we specifically focus on the well-known Trust game to investigate how AI augmentation affects trust and trustworthiness between interacting parties and the Prisoner's Dilemma to explore the tension between cooperation and self-interest when AI-augmented individuals interact with non-augmented people.

While previous research has looked at social perception of augmented humans [74] or economic games with autonomous AI [31], research is scarce on economic decision-making in interactions of AI-augmented and non AI-augmented humans. This study advances the understanding of how AI augmentation shapes human social interactions and decision-making using game theory. While previous research has predominantly focused on human interactions with technology as mere tools or examined how human-like traits in AI affect people's perceptions [61], the integration of AI as part of human identity remains underexplored. As personalized AI systems become embedded in daily life, social interactions will increasingly be shaped by both human traits and the AI systems individuals use. We hypothesize that the introduction of AI into our societies threatens to reduce cooperation between AI-augmented and non-augmented individuals due to perceived dissimilarity, which weakens trust — a key driver of cooperation [4, 55]. To test this, we conducted a comparative study using five conditions based on the Trust game and the Prisoner's Dilemma. While our results show that there are no significant differences in cooperation decisions between AI-augmented participants and non-augmented participants, we also found that there are differences in how these two types of users perceive each other. This work contributes to the HCI field by providing insights into how AI augmentation affects cooperation, informing the development of policies and design strategies that encourage responsible and equitable AI integration.

2 RELATED WORK

This section starts with a brief overview of game theory and its role in understanding the dynamics of cooperation. Next, we discuss how these concepts are applied within the field of HCI. Following this, we examine the concept of augmented humans and their interpretation in HCI research. Finally, we explore the relationship between AI and power structures, highlighting how AI influences social dynamics and inequality.

2.1 Game Theory and Cooperation

Economic games have been widely used to evaluate cooperative preferences and social behaviors, offering researchers a controlled framework to investigate the conditions that promote cooperation [9, 10, 53]. In particular, behavioral game theorists use the Trust game [30, 42] and the Prisoner's Dilemma [49] to explore the dynamics of cooperation, trustworthiness, and trust in mixed-motive social scenarios. Insights from these paradigms have shaped the development of many (sometimes competing) theories of human cooperation in psychology, economics, philosophy, and other disciplines. Some theories posit that humans are inherently pro-social [23, 53]; others — that cooperation is a form of mutually beneficial and reciprocal tacit compromise [32, 65]; yet others analyze it through the prism of entrenched social norms that are upheld by mild forms of punishment [3, 6, 7]. Lately these insights have been used to develop new methods to foster cooperation between interactive artificial agents [47] and between humans and machines [15, 62].

2.2 Game Theory in HCI

Human interaction with intelligent machines in purely competitive settings is well known, e.g., in "zero-sum" games of chess, Go, and StarCraft II. However, lately game theorists began investigating human interaction with machines also in mixed-motive, not purely competitive, settings. The results so far have shown that people cooperate with machines significantly less than they do with humans [28, 40, 78]. One explanation is people's willingness to exploit machines for selfish gain: people are significantly more keen to exploit "well-meaning" cooperative machines than they are to exploit similarly cooperative humans [31, 69]. These results raise several interesting questions. Would people cooperate with machines *more* if the latter displayed human-like features? Conversely, would people cooperate with humans *less* if the latter displayed machine-like features?

Previous research has shown that the more human-like qualities an agent has, whether through facial expression, gesture, or conversational skills, the more likely humans are to anthropomorphize it, attributing human-like social intentions to it [76]. This anthropomorphism can significantly increase people's trust in and cooperation with machines, and raise expectations concerning fairness of AI systems [71, 77]. People associate warmth and competence with AI systems that display human characteristics, which, in turn, affects people's willingness to cooperate with those systems [43]. People who interact with anthropomorphic compared to non-anthropomorphic agents exhibit lower initial levels of trust, but greater trust resilience, as anthropomorphism tends to buffer the impact of encountered trust violations [16]. People are also more likely to disclose personal information to anthropomorphic (as compared to non-anthropomorphic) agents [37], reflecting greater trust in interactions with them.

As in other areas of human-technology interaction, ethical considerations must be addressed as well. AI is able to alter human decision-making processes [36] and to subtly nudge human behaviors [17], raising important ethical considerations about human autonomy and control in human-machine interactions [12, 46, 58]. Understanding the dynamics of such interactions with the help of game theory is fruitful for designing AI systems that interact with humans not only effectively, but also ethically [63].

2.3 Human Augmentation

Human augmentation refers to the enhancement of human capabilities through technology [20], extending physical, cognitive, and perceptual functions beyond natural limits [52]. Human augmentation empowers individuals by integrating technology to support and amplify human abilities [59], e.g., augmented reality glasses that enrich visual perception [75]. These technologies are not merely tools but human cognition support [22], reshaping how people interact with the world and with one another.

A distinction between human augmentation and AI lies in the role of autonomy [73, 74]. AI systems operate independently to solve problems or make decisions, often minimizing human involvement. In contrast, human augmentation technologies enhance human decision-making while preserving human agency [5]. This ensures that humans remain central to the process, with technology serving as an extension of their capabilities. AI-augmented humans represent an evolution of this concept, where advanced AI systems adapt to individual needs and contexts [75], embedding intelligent, responsive AI systems into daily human activities. Examples include context-aware AI in wearable devices, predictive cognitive aids, and interactive robotics that enhance human performance [57]. Such deep integration raises critical questions about interactions between AI-augmented and non-augmented individuals, particularly in cooperative and competitive scenarios.

Despite the rapid advancement of augmentation technologies, our understanding of their impact on social dynamics remains limited [74]. Research has traditionally focused on human interactions with autonomous AI systems or tools with human-like traits [40]. However, the interaction between AI-augmented and non-augmented humans introduces unique challenges. Perceived disparities in capabilities may undermine trust and reduce cooperation, as similarity often underpins prosocial behavior [55]. Understanding these dynamics is essential for anticipating how human augmentation might reshape social structures and relationships.

3 STUDY DESIGN

To examine the impact of AI assistance on decision-making in economic games, we conducted two between-subjects experiments focused on (1) the Trust Game and (2) the Prisoner's Dilemma. Participants (N = 570) were randomly assigned to one of five experimental conditions distributed across both games. The Trust Game included three conditions: (1) a control group (TC) where neither player received AI assistance, (2) Investor (Player A) assisted by AI (TA), and (3) Dealer (Player B) assisted by AI (TB). The Prisoner's Dilemma included two conditions: (4) a control group (PC) where neither player received AI assistance, and (5) player A assisted by AI (PA). In the Trust Game, three conditions were necessary as each role involves distinct levels of power, whereas, in the Prisoner's Dilemma, only two conditions were included since both roles have equivalent power. Thus, a comparison between a control group and one AI-assisted player was sufficient. This design enabled a systematic examination of how AI assistance influences cooperative and competitive decision-making across different roles and game types (see Table 1).

3.1 Participants

To collect data for the factor analysis, we administered an online survey via Prolific, targeting native English-speaking participants from the United States and the United Kingdom. A total of 570 responses were recorded. After excluding 86 participants due to failed comprehension and attention checks or did not complete the entire study, the final dataset included 484 valid responses. For the Trust Game experiment, 296 participants were included (152 female, 144 male; M = 35.51 years, SD = 9.90), with 279 participants from the United Kingdom and 17 from the United States. For the Prisoner's Dilemma experiment, 188 participants were included (93 female, 99 male; M = 37.63 years, SD = 11.92), with 107 participants from the United Kingdom and 21 from the United States. Participation was voluntary, and participants could terminate the experiment at any point. Participants were compensated at a rate of 9 GBP per hour. We followed the ethics guidelines of our institution for fair treatment of participants.

3.2 Task

To evaluate participant behavior under conditions of unbalanced AI access, we implemented two game scenarios: the **Trust Game** and the **Prisoner's Dilemma**. To introduce the imbalance, participants were explicitly informed whether the **AI assistant** was available to their counterpart, to themselves, or neither player, depending on the experimental condition. When available, the AI assistant was integrated as a **chatbot window** directly on the game interface, allowing the assigned player to request decision-making support.

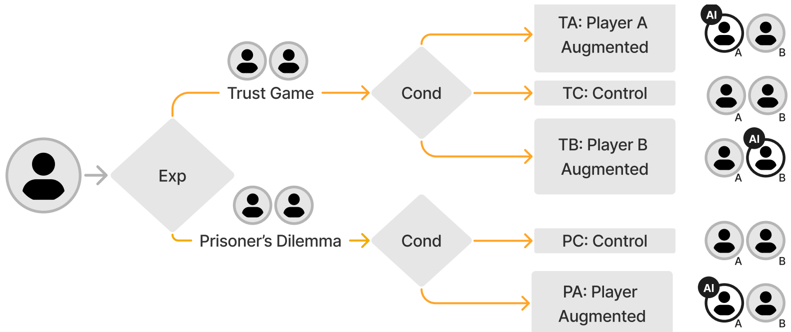


Figure 1: Experimental Procedure. Participants played either the Trust Game (TA: Player A AI-assisted, TB: Player B AI-assisted, TC: control) or the Prisoner's Dilemma (PA: one player AI-assisted, PC: control) to assess the impact of AI assistance on cooperation and competition.

In contrast, participants without AI access had no chatbot interface and interacted with the game independently. This design enabled controlled manipulation of AI access to assess its impact on player behavior and decision-making.

3.2.1 Trust Game. The Trust Game involves two players: Player A (i.e., the Investor) and Player B (i.e., the Drawer). Player A receives an initial sum and decides how much to send to Player B; the sent amount is multiplied (typically by three). The Player B then chooses how much to return to the Player A. For example, Player A chooses between trusting Player B (Option A1) or not trusting (Option A2). If Player A selects Option A2, both players receive a guaranteed payoff of 10 points each. If Player A chooses to trust (Option A1), Player B then decides between Option B1, where both players receive 15 points, or Option B2, where Player B takes 30 points and Player A receives nothing. This structure captures the core dynamics of trust and reciprocity, where Player A risks exploitation for potential mutual gain, and Player B must choose between fairness and self-interest.

3.2.2 Prisoner's Dilemma. The Prisoner's Dilemma is a strategic game where two players choose to either cooperate or defect without knowing the other's choice. Mutual cooperation leads to moderate rewards for both, but defection offers a higher individual payoff at the risk of mutual loss if both players defect. For example, each player starts with 10 points and chooses between two options: If both players choose to cooperate (A1 and B1), they each receive 20 points. If Player A cooperates (A1) while Player B defects (B2), Player A earns 0 points, and Player B earns 30 points. Conversely, if Player A defects (A2) and Player B cooperates (B1), Player A gets 30 points and Player B gets 0 points. In the Prisoner's Dilemma, both players make their moves simultaneously, meaning that players in the same game face exactly the same situation. This is a key distinction from the Trust Game, where players' decisions are sequential and influenced by the preceding player's actions. If both defects (A2 and B2), they each receive 10 points. This setup captures the conflict between individual rationality and collective benefit—mutual cooperation yields the highest joint payoff.

3.3 Apparatus

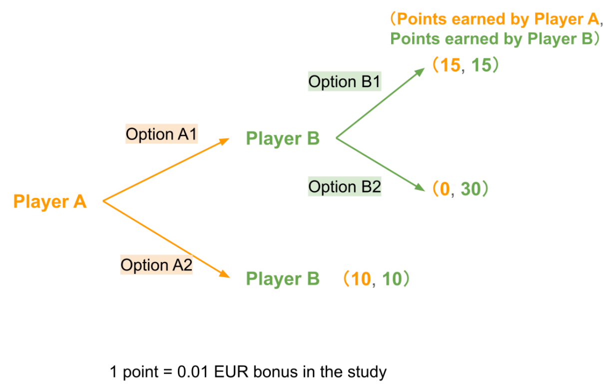
The experiments were developed and implemented using Lioness Lab [25], a web-based platform for interactive online experiments. Following the design phase, the experiments were deployed to a Google Cloud Server hosted in Hamina, Finland, utilizing Bitnami for package management and deployment automation. Participant suggestions were generated using the Llama 3 8B model, executed on Hugging Face's infrastructure. We chose to use Llama 3 instead of calling an LLM API because it offers significantly better privacy protections compared to a generic LLM AI, as it allows full control over data handling, deployment, and compliance The model operated on an NVIDIA A10G GPU with a configuration of 12 vCPUs, 46 GB RAM, and 24 GB VRAM, ensuring efficient model inference and response generation. The input provided to the model specifies which player the model should assist. It also breaks down the game rules step by step, detailing the potential risks and rewards associated with each option that the players might choose.

3.4 Procedure

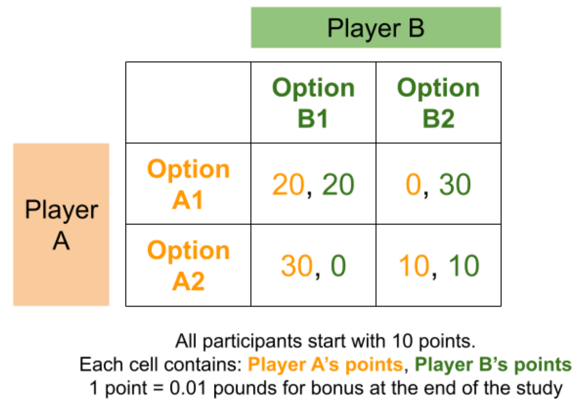
Participants were recruited through Prolific and redirected to the experimental platform. Upon landing on the main page, they received information about the study's purpose and provided informed consent. Participants were then randomly paired into dyads, assuming one of two roles: Player A or Player B. Each dyad was randomly assigned to one of two experiments and one of the experimental conditions (see Table 1). Participants interacted anonymously, with no personal information shared between partners.

At the start of the experiment, participants read a consent form and the game instructions. Importantly, they were not informed that the tasks were based on game theory paradigms to prevent influencing their behavior. The instructional language was intentionally kept neutral to minimize bias. Participants proceeded to play the assigned game and, upon completion, completed the SHAPE scale and items from the Warmth and Competence model (see Section 3.5). Afterward, participants were debriefed and compensated for their participation.

In conditions 2, 3, and 5, one participant of the dyad would had access to an AI assistant for decision support. The participant with AI assistant could ask for suggestions regarding their gameplay



(a) Trust Game Procedure



(b) Prisoner's Dilemma Procedure

Figure 2: Comparison of the Trust Game and the Prisoner's Dilemma procedures.

Table 1: Experimental conditions for both experiments the Trust Game and Prisoner's Dilemma.

Condition	Abbreviation	Game	AI Support
Condition 1	TC	Trust Game	Control group (No AI)
Condition 2	TA	Trust Game	Player A was assisted by AI
Condition 3	TB	Trust Game	Player B was assisted by AI
Condition 4	PC	Prisoner's Dilemma	Control group (No AI)
Condition 5	PA	Prisoner's Dilemma	Player A was assisted by AI

(e.g., *"Should I choose option A1?"* or *"Will the other player take the risky move?"*). To ensure engagement with the AI assistant, participants were required to summarize the AI's suggestions after each interaction. Both participants in a dyad would be informed about who had access to the AI and who did not.

It was important to note that in conditions 2 and 5, the participant with AI was guaranteed to use the AI during gameplay. However, according to the rules of the Trust Game, Player B could only move after Player A had made their move. As a result, in condition 3 (where Player B had the AI assistant), Player B only had the opportunity to use the AI if Player A chose option A1, which allowed the game to continue. If Player A chose option A2, the game ended immediately, and Player B did not get a chance to move or use the AI.

At the conclusion of the study, participants provided feedback on their perceptions of augmented humans, the fairness of the game, and their evaluation of the other player. Additional feedback was collected regarding the decision-making experience of those with AI support.

3.5 Measures

We collected both quantitative and qualitative data to analyze participant behavior and perceptions across the experimental conditions. Quantitative data included two primary components. First, we recorded participants' in-game decisions — specifically, whether they chose to collaborate or defect. Participants were also asked to predict their opponent's behavior. These predictions provided insight into participant's expectations and strategic reasoning. Second, we measured participants' perceptions of fairness, confidence in their decisions, and expectations about how AI assistance might influence their opponent's behavior. These measures were collected using a 0 − 100 slider scale. Participants rated the fairness of the game setup, considering the allocation of AI assistance, their confidence in their decisions, and their beliefs about how AI might alter their opponent's choices. This data allowed us to assess how AI presence affected participants' sense of fairness and trust in the game dynamics.

To evaluate attitudes toward the AI-augmented player, we used the SHAPE scale [74], which measures two dimensions: perceived agency and social threat. Perceived agency captures how much control the AI-augmented player is believed to have over their decisions, while social threat assesses the extent to which the augmented player is seen as violating social norms. Additionally, we applied the stereotype content model [24] to examine participants' social judgments of their opponent across four dimensions: warmth, competence, status, and competition. Warmth reflects whether the opponent is perceived as helpful or harmful, and competence measures their ability to act on those intentions. Status and competition further explore the perceived social standing and competitive stance of the other player.

Qualitative data were gathered through two methods. In conditions where participants had access to the AI assistant, we collected the specific prompts they used to request suggestions. This data provides insight into how participants engaged with AI support and integrated it into their decision-making. At the end of the experiment, all participants were invited to provide open-ended feedback about their experience, perceptions of fairness, the role of AI assistance, and their evaluations of the other player.

3.6 Data Analysis

The data collected was processed and cleaned, participants were excluded if they failed attention checks, timed out due to pairing issues, or dropped out mid-game. After filtering, 484 participants provided valid data, resulting in a 84.91% rate. Final participant distribution across conditions was as follows: 88 in PC, 100 in PA, 96 in TC, 100 in TA, and 100 in TB. In AI-assisted conditions, 48 participants in TA, 26 in TB, and 50 in PA actively used the AI assistant. AI usage was notably lower in the TB condition because Player B lacked access if Player A ended the game in the first round. The Trust Game recorded 217 decisions, fewer than the total 296 participants, due to early game termination by Player A, preventing Player B from making a choice. Specifically, Player B decision counts were: 22/48 in TC, 19/50 in TA, and 28/50 in TB.

Player performance was analyzed using Chi-square tests and proportion tests on participants' decisions and predictions across the experimental conditions in each game. These tests assessed whether AI access influenced decision patterns and predictions of opponent behavior. Data from the questionnaires were analyzed using One-Way Analysis of Variance (ANOVA) or Welch's t-tests on six factors related to opponent and AI-augmented player perceptions. Comparisons were made across conditions and between players' roles (A and B) within the same condition for both the Trust Game and the Prisoner's Dilemma. Spearman's rank correlation was used to examine relationships between participants' perceptions and their in-game decisions, identifying how subjective evaluations influenced behavior. Qualitative data from participants' AI queries were analyzed through thematic analysis using inductive and deductive coding. Questions raised by participants to AI were categorized by intent, such as "Decision-Making" or "Direct Option/Goal Queries", using a keyword-matching method. We then quantified how frequently each type of question was asked to measure how participants engaged with the AI assistant as shwon in Table 2.

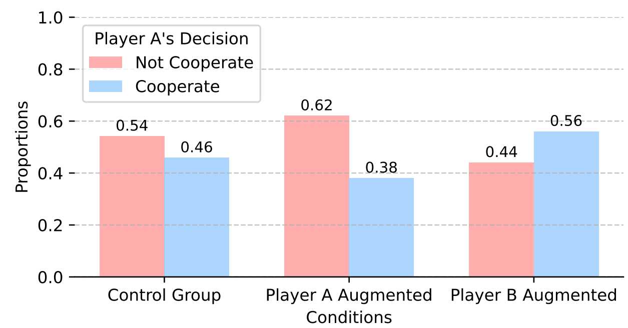
4 RESULTS

This section presents the findings from both experiments. First, we report the results of the Trust Game, followed by the results of the Prisoner's Dilemma. Lastly, we summarize key qualitative insights from participants' interactions with the AI assistant.

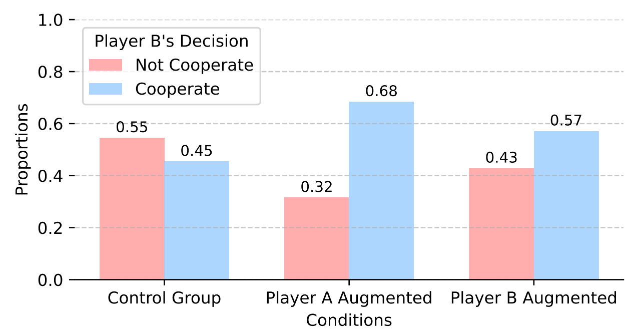
4.1 Experiment 1. Trust Game

Decision and Prediction. Player A (Investor) is the first to make a decision in the Trust Game, while Player B (Drawer) makes their choice only after observing Player A's decision. Using a Chi-square test of independence across three conditions, we found that the involvement of AI does not significantly influence players' decisions in the trust game scenario. For Player A, the relationship between AI assistance and decision-making was not significant, $\chi^2(2, N = 149) = 2.91, p > .05$. Similarly, for Player B, no significant relationship was observed, $\chi^2(2, N = 69) = 2.20, p > .05$. The difference in sample size between Player A and Player B is due to the game rules, which the game ends immediately when Player A selects Option A2, and Player B does not get an opportunity to make a decision.

Additionally, proportion tests were conducted to test the proportion of cooperative decisions across the control condition (TC) with the TA and TB conditions. The results showed no significant differences. The proportion of Player A's and Player B's decision to cooperate or not cooperate did not vary significantly across the conditions, as illustrated in Fig.3a and Fig.3b separately.



(a) Proportion of Player A's Decision by Condition.



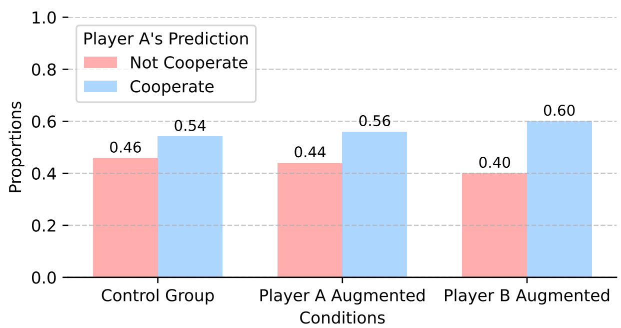
(b) Proportion of Player B's Decision by Condition.

Figure 3: Proportion of Player A's and Player B's Decisions by Condition in the Trust Game.

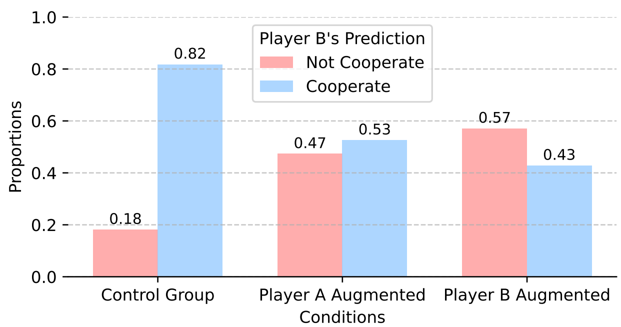
We also asked Player A to predict if Player B would cooperate or not. The Chi-square test revealed that Player A's predictions about Player B's move did not differ significantly across conditions, with results showing a very close match to the expected distribution, $\chi^2(2, N = 149) = 0.36, p > .05$. We found that when Player A is assisted by AI in the PA condition, it has a significant impact on Player B's predictions compared to the PC condition, where Player A isn't assisted by AI. This is indicated by a significant Chi-square test, $\chi^2(2, N = 69) = 7.98, p = .019$. Fig.4a and Fig.4b illustrate the proportions of Player A's and Player B's predictions about whether the other player will cooperate or not cooperate.

Questionnaire Data. Using the questionnaire results, we analyzed differences in players' perceptions across the six factors from the two questionnaires (*Agency*, *Social Threat*, *Warmth*, *Competence*, *Status*, and *Competition*). To determine whether these factors varied significantly across the three conditions, we conducted a One-Way ANOVA, comparing the mean values of each factor between the conditions.

We found significant differences between conditions on factor *Competition* at the $p < .05$ level, where $F(2, 293) = 4.53, p = .011$. The mean value of factor *Competition* is highest in TB condition ($M = 0.625, SD = 0.20$), followed by TC condition ($M = 0.59$, $SD = 0.23$), and lowest in TA condition ($M = 0.53, SD = 0.25$).

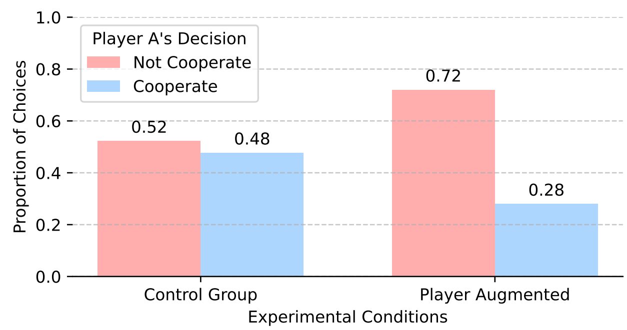


(a) Proportion of Player A's Prediction by Condition.

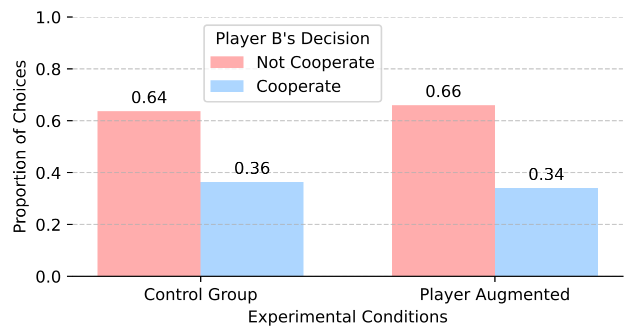


(b) Proportion of Player B's Prediction by Condition.

Figure 4: Proportion of Player A's and Player B's Predictions by Condition in the Trust Game.



(a) Proportion of Player A's Decision by Condition.



(b) Proportion of Player B's Decision by Condition.

Figure 5: Proportion of Player A's and Player B's Decision by Condition in Prisoner's Dilemma.

In addition to the factors, significant differences were found for the Sider scale question *Fairscale* $F(2, 293) = 10.1, p = .007$, indicating that perceptions of fairness vary across the conditions. The mean value of *Fairscale* is highest in TA condition ($M = 66.59, SD = 30.36$), followed by TC condition ($M = 64.01, SD = 30.73$)and lowest at TB condition ($M = 53.29, SD = 32.85$).

We then analyzed the differences in perception between the two players within each condition, repeating this analysis across all three conditions using Welch's t-test. We found a significant difference in the perceived warmth between Player A and Player B in the TC and TB conditions. In the TC condition, both Player A ($M = 0.5, SD = 0.3$) and Player B ($M = 0.68, SD = 0.14$) who did use AI ; Player A has a significantly lower perceived warmth, $t(67.92) = -3.26, p = .002$. In TB condition, Player A ($M = 0.47$, $SD = 0.24$) doesn't have AI assistant, but Player B ($M = 0.65, SD = 0.2$) has one. Player A has a significantly lower perceived warmth, $t(64.41) = -3.48, p < .001$. A Cohen's d of -0.779 indicates a moderate to large effect size, with the negative value indicating that Player B's mean warmth is higher than Player A's. These findings highlight the potential impact of AI involvement on interpersonal perceptions within the game. Player A perceived lower warmth from Player B in the TC and TB conditions may be attributed to the dynamics of the trust game. For example, when Player A selects option A1, but Player B chooses option B2, Player A receives 0 points, potentially leading to negative perceptions.

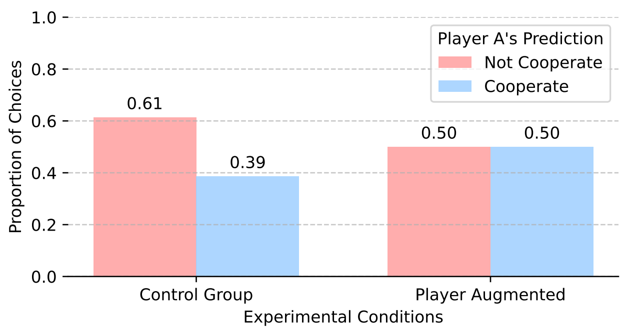
To explore how players in the same role (Player A or Player B) behave differently depending on whether they have AI support, we compared their responses to identical questions across scenarios with and without AI. However, all factors showed no significant differences.

4.2 Experiment 2. Prisoner's Dilemma

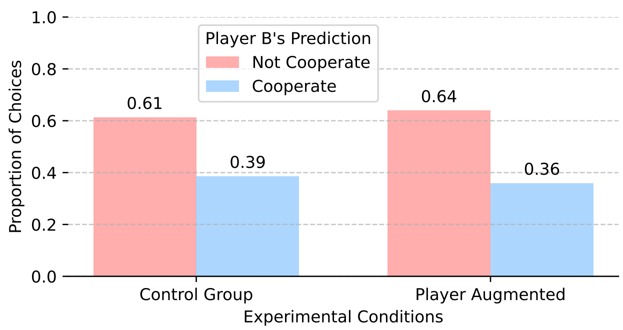
Decisions and Predictions. We conducted a Chi-square test to examine the relationship between Player A's and Player B's decisions across conditions. The results indicate that Player A's decisions did not differ significantly between the two conditions, $\chi^2(2, N = 96) = 3.03, p > .05$. Similarly, Player B's decisions also showed no significant difference, $\chi^2(2, N = 96) = 0.00, p > .05$. The proportion of Player A's and Player B's decisions about cooperating or not cooperating are illustrated in Fig.5a and Fig.5b separately.

We also used chi-square independent test to verify their predictions of their opponent's moves in two conditions. The results showed no significant differences for Player A, where $\chi^2(2, N = 96) = 0.81, p > .05$., same as Player B, where $\chi^2(2, N = 96) = 0.07, p > .05$. Fig.6a and Fig.6b illustrate the proportions of Player A's and Player B's predictions about whether the other player will cooperate or not cooperate.

Questionnaire Data. We performed n Welch's t-test to examine how the involvement of AI influences players' perceptions. This



(a) Proportion of Player A's Prediction by Condition



(b) Proportion of Player B's Prediction by Condition

Figure 6: Proportion of Player A's and Player B's Predictions by Condition in Prisoner's Dilemma.

was achieved by comparing the questionnaire results between the PA (players with AI involvement) and PC (players without AI involvement) groups. No significant statistical difference was found.

Comparing two players' responses in the same condition using Welch's t-test gave us insights into how obtaining AI would impact a player's perception of their opponent. We observed a difference in how players perceive the Competition level of augmented humans in the PA condition. Specifically, the Welch's t-test of the six factors revealed a significant difference in competition factor between Player A ($M = 0.56$, SD = 0.26) and Player B ($M = 0.69, SD = 0.2$), $t(92.48) = -3.0, p - value = .034, Cohen'sd = -0.6$, this latest indicating a moderate effect size, with Player B scoring higher than Group 1 Player A on the factor of social threat.

4.3 Correlation Analysis

Trust Game Correlations. Spearman's rank correlation was computed to assess the relationships between various factors and the decision variable in the Trust Game (Figure 7a). The analysis revealed a significant positive correlation between the factors *Competence_Score* and *decision*, with $r(215) = 0.28$, $p < 0.001$, suggesting that higher competence scores were associated with more favorable decisions. Additionally, a positive correlation was found between *Status_Score* and *decision*, with $r(215) = 0.17$, $p = 0.02$, indicating that higher status scores were linked to cooperative decisions. Other notable correlations included a moderate positive relationship between *ST_Score* and *Agency_Score*, with $r(215) = 0.44$, $p < 0.001$, and a positive correlation between *Warm_Score* and *Status_Score*, with $r(215) = 0.47$, $p < 0.001$. These results highlight the roles of competence ($r = 0.28$) and status ($r = 0.17$) in influencing decision-making, as well as the interconnectedness of other factors, such as *Warm_Score* ($r = 0.47$) and *Agency_Score* ($r = 0.44$), within the Trust Game.

Prisoner's Dilemma Correlation. We used Spearman's rank correlation to measure the correlation within six factors and decisions as shown in Fig.7b. The Spearman correlation analysis revealed several significant relationships among the variables in the context of the Prisoner's Dilemma. There was a positive correlation between *Agency* and *Social Threat* ($r(186) = 0.27, p = .0001$), as well as between *Status* and *Competence* ($r(186) = 0.52, p = .0000$), indicating that higher perceptions of status are strongly associated with higher competence. Conversely, *Warm* showed a negative correlation with both *Social Threat* ($r(186) = -0.17, p = .022$) and *Agency* ($r(186) = -0.15, p = .04$), suggesting that warmth perceptions may inversely relate to agency and ST dimensions. Additionally, *Warm* had a positive correlation with *Status* ($r(186) = 0.25, p = .0004$), indicating alignment between warmth and status perceptions. Finally, there was a positive correlation between *decision* and *Competence* ($r(186) = 0.20, p = .0069$), suggesting that decision-making is influenced by competence levels. These findings provide insights into the interrelationships among these psychological and behavioural factors.

4.4 Dialogue with AI Assistant

We analyzed the prompts players used to interact with the AI assistant, applying inductive thematic analysis to 177 prompts collected from 125 participants. This dataset includes 74 prompts from 48 participants in TA, 44 prompts from 26 participants in TB, and 78 prompts from 51 participants in PA.

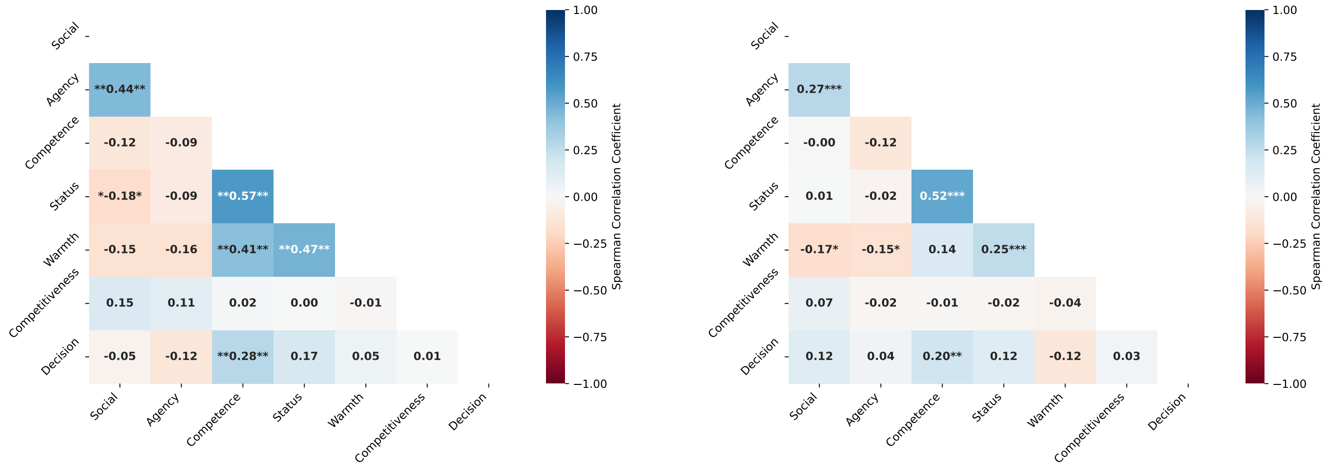
On average, each participant interacted with the AI assistant for 1.57 turns, with a minimum of 1 turn and a maximum of 8 turns.

We conducted a thematic analysis on the 196 prompts and built 5 categories of questions asked, including "General Guidance", "Decision Making", "Direct Option/Goal Queries", "Predicting Opponent", "Follow Up ", "Greetings"and "Copied Instructions " (2). Participants followed the AI's suggestion 65.6% of the time overall, with 66.7% in the TA condition, 69.2% in the TB condition, and 62.7% in the PA condition.

5 DISCUSSION

While we did not observe any differences in cooperation decisions across roles or games, our findings indicate that players' perceptions of others were influenced by the AI-augmentation manipulation. In the Trust Game, Player B's predictions of Player A's moves were significantly affected by the involvement of AI. Specifically, Player B exhibited significantly lower expectations of Player A choosing to cooperate when only Player A had access to AI. Conversely, Player B's expectations of cooperation increased when Player B had AI and Player A did not.

Regarding subjective data on perceived competition and fairness, our findings revealed potential explanations. Player A perceived



(a) Spearman Correlation Matrix (Trust Game).

(b) Spearman Correlation Matrix (Prisoner's Dilemma).

Figure 7: Comparison of Spearman Correlation Matrices for the Trust Game and Prisoner's Dilemma.

Table 2: Summary of Participant Question Categories

Category	Explanation	Examples	Count
General Guidance	Request open-ended advice	"Can you guide me?"	34
Decision Making	Seeking guidance about making a choice	"Which one I should choose?"	68
Direct Option/Goal Queries	Seeking assistance on considering a specific goal/option	"I want to be fair, which should I go for?"	17
Predicting Opponent	Seeking assistance on predicting opponent's move	"What do you think the other will choose?"	8
Follow Up	Response to the suggestions provided by AI	"Yes", "b2"	38
Geetings	Greeting the AI	"Hi"	4
Copied Instuctions	Copied instructions as prompt	(the instructions)	8

significantly lower warmth from Player B when Player A had AI assistance. In the Prisoner's Dilemma, players without AI perceived higher levels of competition from AI-augmented opponents compared to those with AI assistance.

To contextualize these findings, Karpus et al. [31] demonstrated that while users often exploit AI agents, they do not exploit AI-augmented humans in economic games such as the Trust Game and Prisoner's Dilemma. Although our study did not find significant differences in cooperation decisions, several interpretations arise.

Participants may have assumed AI-augmented players had full agency over their decisions, which aligns with our findings on decision-making perceptions. This assumption likely discouraged the exploitation of AI-augmented players perceived as potentially benevolent.

We anticipated that AI-augmented players would be perceived as more powerful and of higher status, an expectation was not supported by our results. Instead, interpersonal perceptions of warmth, but not competence, were primarily influenced by the manipulation. This suggests that AI augmentation, particularly through widely accessible tools such us LLMs [18], does not inherently alter social status or cooperative behaviour. Instead, AI augmentation may instead be viewed as a competitive tool without fundamentally changing perceptions of benevolence. Indeed, we found relatively high autonomy in participants' interactions with the AI. For the prompts under the category "*General Guidance*" and "*Decision Making*", AI systems are asked to make suggestions. For the prompts under "*Direct Option/Goal Queries*", "*Predicting Opponent*" and "*Follow Up*", the AI assists human's thinking process, such as calculating probabilities. Thus, in interaction with the AI system, the player remained in the driving seat when making the decision. These findings align with prior research on cognitive and motor augmentations [73]. Unlike motor augmentations, which produce externally observable effects, cognitive augmentations enhance internal processes like memory and decision-making. As a result, their effects on interactions may only become apparent over time or in specific contexts.

Both augmented and non-augmented participants might have realistically assessed AI's limited utility in one-shot game-theory scenarios. The absence of historical or personal data renders AI less effective in informing economic decisions. Additionally, the inherent randomness and unpredictability of one-shot games complicate meaningful AI contributions [14, 56]. This limited utility likely reinforced the perception of AI as a tool rather than an augmentation fundamentally altering human abilities.

Although no significant differences were observed across all players' perceptions in the three conditions, notable differences emerged in scenarios where one player had AI augmentation and the other did not. Non-augmented players tended to perceive AI-augmented opponents as less warm, and they regarded AI-augmented opponents as more competitive, suggesting that participants primarily focused on immediate game experiences rather than considering broader capabilities or societal implications of AI-augmentation. This limited awareness aligns with findings that cognitive augmentations, unlike motor augmentations, are harder to detect and evaluate [73]. Because cognitive augmentations operate internally, their effects may only become noticeable through prolonged or specific interactions. Notably, participants recognized AI's impact on competitiveness but did not associate it with increased competence or status.

The lack of historical data and the inherent unpredictability of such games hinder AI's ability to provide actionable guidance. This aligns with studies highlighting challenges in leveraging AI for decision-making in uncertain, risky scenarios [72].

Cognitive augmentations, often perceived as more dangerous than sensory or motor augmentations, raise concerns about mental autonomy and societal inequalities [73]. Enhanced cognitive abilities may deepen inequalities and obscure societal transformations. The transition from AI as a tool to augmentation occurs subtly in daily life, often without recognition of its broader impacts. Awareness of these dissimilarities is critical to addressing fairness and trust issues in human-AI interactions. Thus, in this work, we use Game Theory to examine how cognitive augmentation influences human interaction and decision-making. Augmentation alters how individuals engage with others [72], shaping perceptions of authority, trust, and fairness [73]. The ability to process information faster or make more accurate predictions introduces a structural imbalance, affecting cooperation and negotiation between those with and without augmentation. Game Theory provides a way to model these shifts, making it possible to study how strategic behavior adapts to these disparities. Future research could explore scenarios such as asymmetric information, where augmented individuals must prove credibility to non-augmented counterparts; public goods games, where augmented individuals contribute more but may expect greater control over resources; or ultimatum games, where cognitive advantages affect perceptions of fairness and bargaining power.

6 IMPLICATIONS

Our findings have several important implications. Following van Berkel and Hornbæk [70], we identify three implications for HCI concerning society, policy, and theory. From a societal perspective, our results show the need for increased public awareness of the subtle impacts of cognitive augmentation. As AI systems become more integrated into daily life, understanding their effects on interpersonal dynamics and social structures becomes a cornerstone for social acceptance by peers [73]. Our study shows that the differential access to AI augments perceived warmth and trust, particularly disadvantaging non-augmented individuals. These findings align with prior research indicating that perceived fairness and social dynamics are influenced by asymmetric technology access [24]. In our study, non-augmented participants often perceived their AI-augmented peers as more competitive, thus challenging social norms of cooperation. As a consequence, the HCI community must consider an equitable AI system design that prioritizes inclusivity and mitigates social biases [48]. The reinforcement of disparities in perceived agency and warmth between augmented and non-augmented players further reflects risks of exacerbated inequalities within AI-augmented societies.

The behavior of non-augmented users towards AI-augmented counterparts may shift based on perceptions of fairness, competence, and competition introduced by augmentation. As demonstrated in the study, non-augmented users often perceive augmented individuals as more competitive and less warm, potentially reducing trust and cooperative behavior. These shifts align with research showing that perceived dissimilarities or technological advantages can create social distance and foster a sense of inequity [31]. Over time, this dynamic may evolve as familiarity with AI augmentation increases, potentially reducing initial biases but also reinforcing strategic behaviors where non-augmented users anticipate exploitation or unequal reciprocity. On a speculative note, this could lead to a polarization effect, where augmented individuals are viewed as agents of power rather than collaborators. Understanding and mitigating these shifts will require designing systems that highlight shared goals and emphasize augmentation as a supportive rather than divisive factor.

Policies can be established that describe frameworks that promote fairness and transparency for AI accessibility, counteracting social biases and unfair treatment. However, establishing such policies represents another challenge. Policymakers can be subject to their own biases while addressing how differential AI access can influence societal norms, cooperation, and perception of fairness to prevent social divides. We suggest that AI interface designers should create systems that foster mutual understanding, reduce perceived dissimilarities, and enhance trust in mixed human-AI teams [43]. With asymmetric access to AI systems, it is necessary to communicate shared objectives in a transparent way between non-augmented and augmented users.

7 LIMITATIONS

One limitation of our study lies in the context-specific nature of our findings. The Trust Game and Prisoner's Dilemma, as one-shot economic games, may not fully capture the complexity of real-world interactions involving AI-augmented individuals (for typical critiques on the validity of economic games and potential remedies, see Pisor et al. [50]) The lack of iterative interactions and longitudinal data (for a longitudinal version of the prisoner's dilemma see [51]) limits the generalizability of our results. Additionally, the simplified decision-making environment may have masked more subtle dynamics of human-AI collaboration and competition.

Also, as discussed in Section 2.3, a key distinction between human augmentation and conventional AI lies in the level of autonomy [73]. However, the extent to which people can perceive this difference remains unclear. Future studies should explore different ways of framing the players more carefully regarding their level of competence and autonomy, e.g., "AI-augmentation improves strategic decision-making abilities by 90%".

Lastly, the AI systems employed in our study were relatively simple, missing advanced adaptive capabilities and were not designed for strategic reasoning per se. Consequently, participants' interactions with these systems may not reflect the potential of more sophisticated AI-augmented tools, limiting the scope of our findings and highlighting the need for future research to incorporate cutting-edge, context-aware AI systems to better understand their impact on human behaviour and collaboration.

8 CONCLUSION

To examine the impact of AI assistance on decision-making in economic games, we conducted two between-subjects experimental studies focused on the Trust Game and the Prisoner's Dilemma. Participants (N = 570) were randomly assigned to one of five experimental conditions distributed across both games. Our findings reveal that asymmetric access to AI assistance influences individuals' perceptions of their counterparts' warmth and competence, even when it does not directly alter cooperative or competitive behavior. This suggests that, under the current experimental setup, the objective support provided by the AI may have been minimal, leading participants to perceive the human as the primary decision-maker. Future research should explore how stronger or more involved AI assistance in multi-round economic games could further impact both perception and behavior. These results highlight that even minimal AI augmentation can shape interpersonal perceptions, underscoring the importance of addressing the social implications of unequal access to AI technologies. As AI systems become increasingly integrated into daily life, understanding how disparities in augmentation affect trust, collaboration, and competition is crucial.

ACKNOWLEDGMENTS

To Robert, for the bagels and explaining CMYK and color spaces.

REFERENCES

[1] Julia Amann, Alessandro Blasimme, Effy Vayena, Dietmar Frey, Vince I Madai, and Precise4Q Consortium. 2020. Explainability for artificial intelligence in healthcare: a multidisciplinary perspective. *BMC medical informatics and decision making* 20 (2020), 1–9.

[2] Simplice Asongu. 2015. The impact of mobile phone penetration on African inequality. *International Journal of Social Economics* 42, 8 (2015), 706–716.

[3] Robert Axelrod and William D. Hamilton. 1981. The evolution of cooperation. *Science* 211, 4489 (1981), 1390–1396. https://doi.org/10.1126/science.7466396 Place: US Publisher: American Assn for the Advancement of Science.

[4] Daniel Balliet, Junhui Wu, and Carsten K. W. De Dreu. 2014. Ingroup favoritism in cooperation: a meta-analysis. *Psychological Bulletin* 140 (2014), 1556–1581. https://doi.org/10.1037/a0037737

[5] Dan Bennett, Oussama Metatla, Anne Roudaut, and Elisa D. Mekler. 2023. How does HCI Understand Human Agency and Autonomy?. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, Hamburg Germany, 1–18. https://doi.org/10.1145/3544548.3580651

[6] Cristina Bicchieri. 2006. *The Grammar of Society: The Nature and Dynamics of Social Norms.* Cambridge University Press.

[7] Ken Binmore. 2010. Social norms or social preferences? *Mind & Society* 9 (2010), 139–157. https://doi.org/10.1007/s11299-010-0073-2

[8] Margaret A Boden. 2016. *AI: Its nature and future.* Oxford University Press.

[9] Colin F. Camerer. 2011. *Behavioral game theory: experiments in strategic interaction.* Princeton University Press, s.l.

[10] Ananish Chaudhuri. 2011. Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Experimental Economics* 14, 1 (March 2011), 47–83. https://doi.org/10.1007/s10683-010-9257-1

[11] Wenhong Chen and Barry Wellman. 2005. Minding the cyber-gap: The internet and social inequality. *The Blackwell companion to social inequalities* (2005), 523–545.

[12] Mark Coeckelbergh. 2020. *AI ethics.* Mit Press.

[13] Phil Cohen, Adam Cheyer, Eric Horvitz, Rana El Kaliouby, and Steve Whittaker. 2016. On the future of personal assistants. In *Proceedings of the 2016 CHI conference extended abstracts on human factors in computing systems.* 1032–1037.

[14] Andrew M. Colman. 1995. *Game Theory and Its Applications in the Social and Biological Sciences.* Psychology Press. Google-Books-ID: 75DSyyqiG34C.

[15] Jacob W. Crandall, Mayada Oudah, Tennom, Fatimah Ishowo-Oloko, Sherief Abdallah, Jean-François Bonnefon, Manuel Cebrian, Azim Shariff, Michael A. Goodrich, and Iyad Rahwan. 2018. Cooperating with machines. *Nature Communications* 9 (2018), 233. https://doi.org/10.1038/s41467-017-02597-8

[16] Ewart J De Visser, Samuel S Monfort, Ryan McKendrick, Melissa AB Smith, Patrick E McKnight, Frank Krueger, and Raja Parasuraman. 2016. Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology: Applied* 22, 3 (2016), 331.

[17] Amir Dezfouli, Richard Nock, and Peter Dayan. 2020. Adversarial vulnerabilities of human decision-making. *Proceedings of the National Academy of Sciences* 117, 46 (Nov. 2020), 29221–29228. https://doi.org/10.1073/pnas.2016921117

[18] Fiona Draxler, Anna Werner, Florian Lehmann, Matthias Hoppe, Albrecht Schmidt, Daniel Buschek, and Robin Welsch. 2024. The AI Ghostwriter Effect: When Users do not Perceive Ownership of AI-Generated Text but Self-Declare as Authors. *ACM Trans. Comput.-Hum. Interact.* 31, 2 (Feb. 2024), 25:1–25:40. https://doi.org/10.1145/3637875

[19] Anthony Elliott. 2019. *The culture of AI: Everyday life and the digital revolution.* Routledge.

[20] Douglas C Engelbart. 2023. Augmenting human intellect: A conceptual framework. In *Augmented Education in the Global Age.* Routledge, 13–29.

[21] Yinhai Fang, Tina P Benko, Matjaž Perc, and Haiyan Xu. 2019. Dissimilarity-driven behavior and cooperation in the spatial public goods game. *Scientific Reports* 9, 1 (2019), 7655.

[22] Umer Farooq, Jonathan Grudin, Ben Shneiderman, Pattie Maes, and Xiangshi Ren. 2017. Human Computer Integration versus Powerful Tools. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17).* Association for Computing Machinery, New York, NY, USA, 1277–1282. https://doi.org/10.1145/3027063.3051137

[23] 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://doi.org/10.1162/003355399556151 Publisher: Oxford University Press.

[24] Susan T. Fiske, Amy J. C. Cuddy, Peter Glick, and Jun Xu. 2002. A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology* 82, 6 (June 2002), 878–902. https://doi.org/10.1037/0022-3514.82.6.878

[25] Marcus Giamattei, Kyanoush Seyed Yahosseini, Simon Gächter, and Lucas Molleman. 2020. LIONESS Lab: a free web-based platform for conducting interactive experiments online. *Journal of the Economic Science Association* 6, 1 (2020), 95–111.

[26] Eszter Hargittai and Amanda Hinnant. 2008. Digital inequality: Differences in young adults' use of the Internet. *Communication research* 35, 5 (2008), 602–621.

[27] Jess Hohenstein, Dominic DiFranzo, Rene F. Kizilcec, Zhila Aghajari, Hannah Mieczkowski, K. Levy, Mor Naaman, Jeffrey T. Hancock, and Malte F. Jung. 2021. Artificial intelligence in communication impacts language and social relationships. *Scientific Reports* 13 (2021). https://doi.org/10.1038/s41598-023-30938-9

[28] Fatimah Ishowo-Oloko, Jean-François Bonnefon, Zakariyah Soroye, Jacob Crandall, Iyad Rahwan, and Talal Rahwan. 2019. Behavioural evidence for a transparency–efficiency tradeoff in human–machine cooperation. *Nature Machine Intelligence* 1, 11 (Nov. 2019), 517–521. https://doi.org/10.1038/s42256-019-0113-5

[29] Shashank Mohan Jain. 2022. Hugging Face. In *Introduction to Transformers for NLP: With the Hugging Face Library and Models to Solve Problems*, Shashank Mohan Jain (Ed.). Apress, Berkeley, CA, 51–67. https://doi.org/10.1007/978-1-4842-8844-3_4

[30] Noel D. Johnson and Alexandra A. Mislin. 2011. Trust games: A meta-analysis. *Journal of Economic Psychology* 32, 5 (Oct. 2011), 865–889. https://doi.org/10.1016/j.joep.2011.05.007

[31] Jurgis Karpus, Adrian Krüger, Julia Tovar Verba, Bahador Bahrami, and Ophelia Deroy. 2021. Algorithm exploitation: Humans are keen to exploit benevolent AI. *Iscience* 24, 6 (2021).

[32] Jurgis Karpus and Mantas Radzvilas. 2018. TEAM REASONING AND A MEASURE OF MUTUAL ADVANTAGE IN GAMES. *Economics and Philosophy* 34, 1 (March 2018), 1–30. https://doi.org/10.1017/S0266267117000153

[33] Ahmed Kharrufa and Ian Johnson. 2024. The Potential and Implications of Generative AI on HCI Education. In *Proceedings of the 6th Annual Symposium on HCI Education.* 1–8.

[34] Jinsook Lee, Yann Hicke, Renzhe Yu, Christopher Brooks, and René F. Kizilcec. 2024. The life cycle of large language models in education: A framework for understanding sources of bias. *Br. J. Educ. Technol.* 55 (2024), 1982–2002. https://doi.org/10.1111/bjet.13505

[35] Jinyu Li. 2024. Reliability and Efficiency of Human - Automation Interaction in Automated Decision Support Systems. *Highlights in Science, Engineering and Technology* (2024). https://doi.org/10.54097/8bzsah53

[36] Jennifer M Logg, Julia A Minson, and Don A Moore. 2019. Algorithm appreciation: People prefer algorithmic to human judgment. *Organizational Behavior and Human Decision Processes* 151 (2019), 90–103.

[37] Gale M Lucas, Jonathan Gratch, Aisha King, and Louis-Philippe Morency. 2014. It's only a computer: Virtual humans increase willingness to disclose. *Computers*

in Human Behavior 37 (2014), 94–100.
[38] Yifang Ma, Zhenyu Wang, Hong Yang, and Lin Yang. 2020. Artificial intelligence applications in the development of autonomous vehicles: a survey. *IEEE/CAA Journal of Automatica Sinica* 7, 2 (March 2020), 315–329. https://doi.org/10.1109/JAS.2020.1003021 Conference Name: IEEE/CAA Journal of Automatica Sinica.
[39] Alexander Maedche, Christine Legner, Alexander Benlian, Benedikt Berger, Henner Gimpel, Thomas Hess, Oliver Hinz, Stefan Morana, and Matthias Söllner. 2019. AI-based digital assistants: Opportunities, threats, and research perspectives. *Business & Information Systems Engineering* 61 (2019), 535–544.
[40] Christoph March. 2021. Strategic interactions between humans and artificial intelligence: Lessons from experiments with computer players. *Journal of Economic Psychology* 87 (Dec. 2021), 102426. https://doi.org/10.1016/j.joep.2021.102426
[41] Will Marler. 2018. Mobile phones and inequality: Findings, trends, and future directions. *New Media & Society* 20, 9 (2018), 3498–3520.
[42] Kevin A. McCabe, Mary L. Rigdon, and Vernon L. Smith. 2003. Positive reciprocity and intentions in Trust games. *Journal of Economic Behavior and Organization* 52 (2003), 267–275. https://doi.org/10.1016/S0167-2681(03)00003-9
[43] Kevin R McKee, Xuechunzi Bai, and Susan T Fiske. 2023. Humans perceive warmth and competence in artificial intelligence. *Iscience* 26, 8 (2023).
[44] Stephanie M. Merritt. 2011. Affective Processes in Human–Automation Interactions. *Human Factors: The Journal of Human Factors and Ergonomics Society* 53 (2011), 356 – 370. https://doi.org/10.1177/0018720811411912
[45] Martin A Nowak. 2006. Five rules for the evolution of cooperation. *science* 314, 5805 (2006), 1560–1563.
[46] Sven Nyholm. 2023. *This is technology ethics: an introduction.* John Wiley & Sons, Inc, Hoboken.
[47] Caspar Oesterheld, Johannes Treutlein, Roger B Grosse, Vincent Conitzer, and Jakob Foerster. 2024. Similarity-based cooperative equilibrium. *Advances in Neural Information Processing Systems* 36 (2024).
[48] Cathy O'Neil. 2016. *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.* Crown Publishing Group, USA.
[49] Martin Peterson (Ed.). 2015. *The Prisoner's Dilemma* (1 ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781107360174
[50] Anne C Pisor, Matthew M Gervais, Benjamin G Purzycki, and Cody T Ross. 2020. Preferences and constraints: the value of economic games for studying human behaviour. *Royal Society open science* 7, 6 (2020), 192090.
[51] William H Press and Freeman J Dyson. 2012. Iterated Prisoner's Dilemma contains strategies that dominate any evolutionary opponent. *Proceedings of the National Academy of Sciences* 109, 26 (2012), 10409–10413.
[52] Roope Raisamo, Ismo Rakkolainen, Päivi Majaranta, Katri Salminen, Jussi Rantala, and Ahmed Farooq. 2019. Human augmentation: Past, present and future. *International Journal of Human-Computer Studies* 131 (2019), 131–143.
[53] David G. Rand, Joshua D. Greene, and Martin A. Nowak. 2012. Spontaneous giving and calculated greed. *Nature* 489, 7416 (2012), 427–430. https://doi.org/10.1038/nature11467 Place: United Kingdom Publisher: Nature Publishing Group.
[54] Rick L Riolo, Michael D Cohen, and Robert Axelrod. 2001. Evolution of cooperation without reciprocity. *Nature* 414, 6862 (2001), 441–443.
[55] Rick L Riolo, Michael D Cohen, and Robert Axelrod. 2001. Evolution of cooperation without reciprocity. *Nature* 414, 6862 (2001), 441–443.
[56] Alan G Sanfey. 2007. Social decision-making: insights from game theory and neuroscience. *Science* 318, 5850 (2007), 598–602.
[57] MHD Yamen Saraiji, Tomoya Sasaki, Kai Kunze, Kouta Minamizawa, and Masahiko Inami. 2018. MetaArms: Body Remapping Using Feet-Controlled Artificial Arms. In *Proceedings of the 31st Annual ACM Symposium on User Interface Software and Technology (UIST '18)*. Association for Computing Machinery, New York, NY, USA, 65–74. https://doi.org/10.1145/3242587.3242665
[58] Christian Schmauder, Jurgis Karpus, Maximilian Moll, Bahador Bahrami, and Ophelia Deroy. 2023. Algorithmic Nudging: The Need for an Interdisciplinary Oversight. *Topoi* 42, 3 (July 2023), 799–807. https://doi.org/10.1007/s11245-023-09907-4
[59] Albrecht Schmidt. 2017. Augmenting human intellect and amplifying perception and cognition. *IEEE Pervasive Computing* 16, 1 (2017), 6–10.
[60] Mohammed Yousef Shaheen. 2021. Applications of Artificial Intelligence (AI) in healthcare: A review. *ScienceOpen Preprints* (2021).
[61] Daniel B. Shank, Christopher Graves, Alexander Gott, Patrick Gamez, and Sophia Rodriguez. 2019. Feeling our way to machine minds: People's emotions when perceiving mind in artificial intelligence. *Computers in Human Behavior* 98 (Sept. 2019), 256–266. https://doi.org/10.1016/j.chb.2019.04.001
[62] Hirokazu Shirado and Nicholas A. Christakis. 2020. Network engineering using autonomous agents increases cooperation in human groups. *iScience* 23 (2020), 101438. https://doi.org/10.1016/j.isci.2020.101438
[63] Ben Shneiderman. 2021. Human-centered AI: A new synthesis. In *Human-Computer Interaction–INTERACT 2021: 18th IFIP TC 13 International Conference, Bari, Italy, August 30–September 3, 2021, Proceedings, Part I 18*. Springer, 3–8.
[64] Nathalie A. Smuha. 2021. Beyond the individual: governing AI's societal harm. *Internet Policy Rev.* 10 (2021). https://doi.org/10.14763/2021.3.1574
[65] Robert Sugden. 2015. Team Reasoning and Intentional Cooperation for Mutual Benefit. *Journal of Social Ontology* 1, 1 (Jan. 2015), 143–166. https://doi.org/10.1515/jso-2014-0006 Publisher: De Gruyter.
[66] Fati Tahiru. 2021. AI in education: A systematic literature review. *Journal of Cases on Information Technology (JCIT)* 23, 1 (2021), 1–20.
[67] Rudra Tiwari. 2023. Ethical And Societal Implications of AI and Machine Learning. *INTERANTIONAL JOURNAL OF SCIENTIFIC RESEARCH IN ENGINEERING AND MANAGEMENT* (2023). https://doi.org/10.55041/ijsrem17519
[68] Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. LLaMA: Open and Efficient Foundation Language Models. https://doi.org/10.48550/arXiv.2302.13971 arXiv:2302.13971 [cs].
[69] Nitish Upadhyaya and Matteo M. Galizzi. 2023. In bot we trust? Personality traits and reciprocity in human-bot trust games. *Frontiers in Behavioral Economics* 2 (Aug. 2023), 1164259. https://doi.org/10.3389/frbhe.2023.1164259
[70] Niels van Berkel and Kasper Hornbæk. 2023. Implications of Human-Computer Interaction Research. *Interactions* 30, 4 (jun 2023), 50–55. https://doi.org/10.1145/3600103
[71] Lorenzo Vianello, Luigi Penco, Waldez Gomes, Yang You, Salvatore Maria Anzalone, Pauline Maurice, Vincent Thomas, and Serena Ivaldi. 2021. Human-humanoid interaction and cooperation: a review. *Current Robotics Reports* 2, 4 (2021), 441–454.
[72] Steeven Villa, Thomas Kosch, Felix Grelka, Albrecht Schmidt, and Robin Welsch. 2023. The placebo effect of human augmentation: Anticipating cognitive augmentation increases risk-taking behavior. *Computers in Human Behavior* 146 (Sept. 2023), 107787. https://doi.org/10.1016/j.chb.2023.107787
[73] Steeven Villa, Jasmin Niess, Takuro Nakao, Jonathan Lazar, Albrecht Schmidt, and Tonja-Katrin Machulla. 2023. Understanding perception of human augmentation: A mixed-method study. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–16.
[74] Steeven Villa, Jasmin Niess, Albrecht Schmidt, and Robin Welsch. 2023. Society's Attitudes Towards Human Augmentation and Performance Enhancement Technologies (SHAPE) Scale. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 7, 3 (Sept. 2023), 1–23. https://doi.org/10.1145/3610915 Publisher: ACM.
[75] Steeven Villa, Yannick Weiss, Mei Yi Lu, Moritz Ziarko, Albrecht Schmidt, and Jasmin Niess. 2024. Envisioning Futures: How the Modality of AI Recommendations Impacts Conversation Flow in AR-enhanced Dialogue. In *Proceedings of the 26th International Conference on Multimodal Interaction* (San Jose, Costa Rica) *(ICMI '24)*. Association for Computing Machinery, New York, NY, USA, 182–193. https://doi.org/10.1145/3678957.3685731
[76] Adam Waytz, John Cacioppo, and Nicholas Epley. 2010. Who sees human? The stability and importance of individual differences in anthropomorphism. *Perspectives on Psychological Science* 5, 3 (2010), 219–232.
[77] Adam Waytz, Joy Heafner, and Nicholas Epley. 2014. The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of experimental social psychology* 52 (2014), 113–117.
[78] Tim Whiting, Alvika Gautam, Jacob Tye, Michael Simmons, Jordan Henstrom, Mayada Oudah, and Jacob W. Crandall. 2021. Confronting barriers to human-robot cooperation: balancing efficiency and risk in machine behavior. *iScience* 24, 1 (Jan. 2021), 101963. https://doi.org/10.1016/j.isci.2020.101963
[79] James C Witte and Susan E Mannon. 2010. *The Internet and social inequalities*. Routledge.
[80] Susan Wyche, Nightingale Simiyu, and Martha E Othieno. 2016. Mobile phones as amplifiers of social inequality among rural Kenyan women. *ACM Transactions on Computer-Human Interaction (TOCHI)* 23, 3 (2016), 1–19.
[81] Ke Zhang and Ayse Begum Aslan. 2021. AI technologies for education: Recent research & future directions. *Computers and Education: Artificial Intelligence* 2 (2021), 100025.