When Should I Lead or Follow? Understanding Initiative Levels in Human-AI Collaborative Gameplay

Inês Lobo
INESC-ID and Instituto Superior Técnico, Universidade de Lisboa
Portugal
ines.lobo@tecnico.ulisboa.pt

Janin Koch
ExSitu, Université Paris-Saclay, CNRS, Inria
France
janin.koch@inria.fr

Jennifer Renoux
Center for Applied Autonomous Sensor Systems, Örebro University
Sweden
jennifer.renoux@oru.se

Inês Batina
INESC-ID and Instituto Superior Técnico, Universidade de Lisboa
Portugal
ines.batina@tecnico.ulisboa.pt

Rui Prada
INESC-ID and Instituto Superior Técnico, Universidade de Lisboa
Portugal
rui.prada@tecnico.ulisboa.pt

ABSTRACT

Dynamics in Human-AI interaction should lead to more satisfying and engaging collaboration. Key open questions are how to design such interactions and the role personal goals and expectations play. We developed three AI partners of varying initiative (leader, follower, shifting) in a collaborative game called *Geometry Friends*. We conducted a within-subjects experiment with 60 participants to assess personal AI partner preference and performance satisfaction as well as perceived warmth and competence of AI partners. Results show that AI partners following human initiative are perceived as warmer and more collaborative. However, some participants preferred AI leaders for their independence and speed, despite being seen as less friendly. This suggests that assigning a leadership role to the AI partner may be suitable for time-sensitive scenarios. We identify design factors for developing collaborative AI agents with varying levels of initiative to create more effective human-AI teams that consider context and individual preference.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; • **Computing methodologies** → *Artificial intelligence*.

KEYWORDS

Human-AI Collaboration, Initiative in AI Partners, Collaborative Game, Collaboration Preference

ACM Reference Format:
Inês Lobo, Janin Koch, Jennifer Renoux, Inês Batina, and Rui Prada. 2024. When Should I Lead or Follow? Understanding Initiative Levels in Human-AI Collaborative Gameplay. In *Designing Interactive Systems Conference (DIS '24), July 01–05, 2024, IT University of Copenhagen, Denmark.* ACM, New York, NY, USA, 20 pages. https://doi.org/10.1145/3643834.3661583

DIS '24, July 01–05, 2024, IT University of Copenhagen, Denmark

ACM ISBN 979-8-4007-0583-0/24/07
https://doi.org/10.1145/3643834.3661583

1 INTRODUCTION

Designing dynamic human-computer collaboration that allows users to better understand and work with machines is a key goal of HCI [19]. While mixed-initiative interaction research [19] has provided valuable insights and guidelines into different aspects of human-computer cooperation over the past decades [2], there is no definitive answer on the leadership roles that humans and agents should play in a more collaborative context (i.e., how much initiative agents should have compared to the humans). Collaboration is often referred to in human-AI collaboration as "a coordinated, synchronous activity that is the result of a continued attempt to construct and maintain a shared conception of a problem" [30]. This phrasing has been applied in various domains such as education, healthcare, art and design [45].

With the recent advancements in technology, the possibility of more adaptive human-AI collaboration has become more feasible and practical than before. How to design initiative in interaction, and the role that personal goals and expectations play, is nevertheless still an open question. Especially users' perspectives on AI partners, of themselves, and personal preferences are expected to play a crucial role, but more insights are needed to facilitate the development of more effective human-AI collaboration [16, 25, 34, 38].

Previous research has shown that AI agents in collaborative games can help us better understand how humans and AI interact, as well as uncover agent behaviours that improve human perceptions, performance, and coordination, thereby improving gameplay experience [3, 4, 10, 14–16, 34]. Studies with human-robot teams performing collaborative tasks [17, 23, 26, 38] have also contributed insights into human-AI collaboration, with a few focusing on the role of agent initiative on human perceptions and behaviour. These studies suggest that AI partners with different confidence levels [4], skill levels [3, 14], leadership levels in task planning or execution [17, 23, 38], adaptability to the human strategy [10, 26], can lead to different perceptions of trust, performance, rapport of the AI agent, and therefore can potentially have a significant impact on the overall perceived collaboration.

This paper proposes using game environments like *Geometry Friends* [29] to evaluate the interaction between humans and artificial agents. To better investigate the effects of AI initiative levels on human perception, we designed three different agent behaviours within a collaborative game: 1) a leader agent who acts according to its own plan expecting the human player to follow; 2) a follower agent that aligns its actions to follow the human player's plan; and 3) a shifting agent that changes its initiative depending on whether the human player follows its plan.

Given the level of initiative of the agent, including its willingness to shift initiative, we intend to answer four research questions:

- **RQ1:** How does an agent's initiative influence the perception of AI partners (agent focus e.g., perceived agent warmth and competence, social identification)?
- **RQ2:** How does an agent's initiative impact the perceived quality of collaboration (interaction focus e.g., satisfaction with team and agent performance, objective performance)?
- **RQ3:** How does an agent's initiative impact the humans' self-perception (user focus e.g., satisfaction with self-performance, perception of played role)?
- **RQ4:** How does agent's initiative affect the overall team perception (team focus e.g., agent preference)?

To address these questions, we conducted a study with a mixed-methods design, using the collaborative game *Geometry Friends*, involving 60 participants across three countries. We present our rationales for designing AI agents with varying levels of initiative, as well as an empirical evaluation of interacting with these agents to assess perceived AI partner warmth and competence, social identification with the team, satisfaction with performance (self, AI partner, and team), and AI partner personal preference based on different levels of initiative.

In the following sections, we begin by briefly discussing relevant literature, focusing on games being used as a research platform in human-AI collaboration, and studies that explore leadership dynamics and mixed-initiative in human-AI teams. Next, we describe the game environment as well as the design of each of the three AI partners. Following this, we present the study methodology and results, and conclude with a discussion of our insights and how to develop collaborative AI agents with varying levels of initiative, with directions for future research.

2 RELATED WORK

In the following, we will briefly introduce recent research on the collaboration between humans and AI. We will also discuss the use of games as a platform for studying this collaboration, as well as existing research on how the initiative of AI agents affects human-AI collaboration.

2.1 Human-AI Collaboration

Human-AI collaboration has been extensively applied within the fields of education, healthcare and art [45], whether through virtual assistants or collaborative robots. Empirical evidence shows that such collaborative work enhances the cognitive capabilities of people involved, allowing them as a team to reach a level of cognitive performance that exceeds the sum of the individuals [5]. With that, collaborative agents may have the opportunity to expand human productivity and performance by collaborating with them rather than replacing them [20, 37, 39, 44].

The existing success in the area of human-AI collaboration enabled the emergence of the human-AI teaming as a subfield, which involves the coordination of humans and agents to solve complex tasks [45]. Several user studies have been conducted in this context, evaluating human-virtual agent teams playing computer games [3, 4, 10, 14–16, 34, 42] and human-robot teams performing collaborative tasks [17, 23, 26, 38], under different team conditions and AI partner features. Their findings suggest that different AI partner behaviours, team compositions, and game environments, influence perceptions of the agent, as well as team performance and coordination. In particular, agents that consider human preferences in their behaviour are usually seen more positively.

However, there has not been much research on the human perspective of AI partners in human-AI teams (e.g., human mental models and expectations) and their impact on the joint action [45]. A study we identified in this regard is using surveys and interviews in the context of multiplayer online games to provide guidelines for the design of collaborative AI agents [45]. Nonetheless, we see an opportunity for further exploration using more controlled and simple game environments.

2.2 Games as a study platform

In the area of collaborative computer games, there are many well-known examples, such as "Overcooked"[1] and "Fireboy and Watergirl"[2]. In particular, several studies have used a simplified environment based on the "Overcooked" game [10, 15]. In one of them, different game environments were found to be a key factor in the emergence of coordination behaviours of human-agent teams (e.g., workload differences, different team fluency) [15]. In another study, Carroll et al. [10] assessed agent performance when teaming up with humans, given different learning algorithms, determining that agents employing imitation learning algorithms were more adaptive and efficient than agents who expected their partners to be optimal.

Schelble et al. [34] further studied the effects of team composition, human-agent teams vs. human-human teams, on shared mental models, performance, and trust, using a collaborative resource-management game where teams had to perform emergency tasks. They highlighted the importance of "action-related communication and explicit shared goals" [34] in human-agent teams. In a similar context, Gero et al. [16] used a collaborative word guessing game to explore how people develop mental models of AI partners, showing that they discussed them more in an out-loud setting when unexpected events happened and suggesting to give humans explanations in these situations as they are more prone to listen. Ashktorab et al. [4] investigated further the effects of an agent's confidence level on user perception (e.g., intelligence, creativity, rapport) using another word game, finding that users had more rapport with AI partners who displayed low confidence during their moves. Using the same game, by varying the agent's learning algorithms and game roles, researchers found that the agent was perceived more positively when its performance was higher [3]. In both of these studies, a few participants were led to believe they

[1]https://www.team17.com/games/overcooked/
[2]https://fireboyand-watergirl.co/

were playing with a human rather than an AI, resulting in higher user perception values.

Whebe et al. [42] also studied how perceptions of AI and humans as teammates in collaborative game play change, stating that even if AI teammates are humanised, they are still perceived as having to play a compliant role, and if they are not competent, they will be judged more harshly. Using a Trivia Game with an AI providing game suggestions, Feng et al. [14] also evaluated human players of various skill levels. They verified that novices trusted the AI agent a lot, whereas experts were able to ignore its bad suggestions. They concluded that in-game suggestions should possibly include a trust measure, similar to the confidence level in [4]. Linked to the downsides of over-relying on AI and its impact on performance, Cabrera et al. [9] studied how showing details of how AI performs, such as metrics and possible failures, benefits decision-making assisted by AI.

2.3 Understanding Initiative of AI Agents in Human-AI Collaboration

Several studies have explored leadership dynamics and initiative roles in human-robot collaboration. For example, Gombolay et al [17] investigated the impact of a robot's autonomy level on the user's situational awareness of the team's schedule during collaborative task planning. They found that as the robot's autonomy increased, the user's situational awareness decreased. They also found that humans preferred to work with a robot that planned the task considering their preferences. Moreover, Lei et al. [23] studied the effect of robot status, or leadership level, on responsibility attributions, finding that users felt less responsible for the task performance as the robot status increased. Finally, Van Zoelen et al. [38] studied the development of leading and following behaviours in a collaborative task, in which humans had to move between two physical points while holding the leash of a robot with a conflicting behaviour. Their results varied considerably, stating that the evolution of leading and follower behaviours is not the same for every individual and that the "diversity of leadership development will need to be taken into account when designing collaborative interactions" [38] of humans and agents.

In HCI, research on mixed-initiative interaction has been investigating how leadership roles can shift between humans and agents for decades [19]. This concept entails that "initiative switches between the user and the system" [22], and has been explored in diverse contexts, such as collaborative writing with AI systems [22] and AI-assisted game level design and generation [12]. In the latter scenario, researchers found that many users preferred having full control over level design rather than relying on AI suggestions. Nikolaidis et al.'s work [26] is a practical example of mixed-initiative in human-robot collaboration. They explored the influence of different adaptability strategies of a robot on the users' own adaptability level, finding that participants adapted well to the robot employing a mutual adaptation strategy (i.e., if the human is adaptable, the robot takes initiative and follows optimal strategy; if not, the robot follows human initiative). Users perceived it to be more trustworthy compared to the robot using a fixed optimal strategy.

Leadership dynamics within human-AI teams, as well as the adaptiveness of the agent's leadership role given human preferences, could be further explored with games. In particular, understanding the reasons for humans preferring to lead or to follow in human-AI teams is essential for a more effective collaboration. A mixed-methods design appears to be a good methodology to do so, as seen in other human-AI collaboration studies [25, 34, 38].

3 GAME AND AGENT DESIGN

In this section, the game environment and game levels are introduced. Furthermore, we detail the behaviour, design rationale, and implementation of the agents.

3.1 Game environment

3.1.1 Game Description. Geometry Friends [29] is a two-player collaborative puzzle platformer game with a 2D physics-based environment, having gravity and friction. The game was developed in C# using the *Microsoft XNA framework*, designed for PC and Xbox 360 game development. *XNA* provides a game cycle structure and essential classes for handling input, graphics, and sound. For the 2D physics, the game employs the *Farseer Physics Engine*, "a collision detection system with realistic physics responses" [3] designed for the *XNA framework*. This engine has been applied in additional studies with games requiring realistic physics simulations [11, 24]. The game can be played locally with a keyboard, *Wii* remotes or game controllers, which were used for the study. The controllers were integrated using *SharpDX*, an open-source wrapper for the *Microsoft DirectX API*.

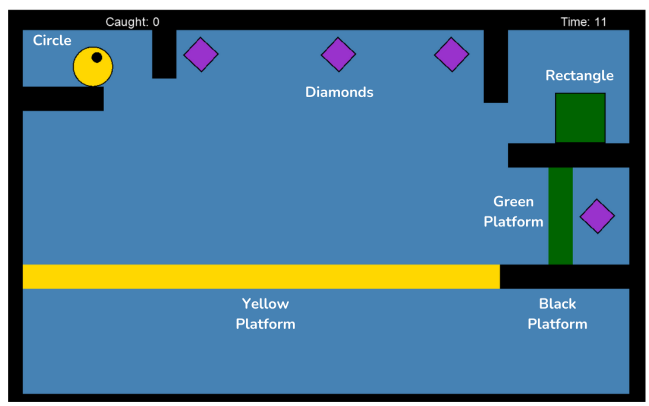

Figure 1: A *Geometry Friends* level: the yellow circle and the green rectangle collaboratively collect the purple diamonds. The circle collides with the green and black platforms, while the rectangle collides with the yellow and black platforms.

3.1.2 Characters (Circle and Rectangle) and Actions. Players control one of two characters, a yellow circle capable of jumping and rolling, and a green rectangle that can slide and morph its shape to a vertical or a horizontal rectangle. The goal of the game is for players to collaborate and collect a set of purple diamonds positioned throughout the levels as quickly as possible, as seen in Figure 1. Each level can feature different platform types, such as standard black platforms that collide with both characters, yellow

[3] https://www.nuget.org/packages/FarseerPhysics

platforms that exclusively collide with the green rectangle, and green platforms that only collide with the yellow circle.

3.1.3 Collaboration Challenges. The game includes several collaboration challenges related to situational awareness, motion control and task planning [32]. Due to that, *Geometry Friends* has been used for an AI competition, in which participants have to develop their own agents for the game, solving individual and collaborative puzzles [27]. There are several implementations of agents for *Geometry Friends* [21, 28, 31, 36], making the game suitable for AI research and development likewise. The game also serves as an effective framework for developing design principles for collaborative AI partners. It is simple to create and has a restricted range of actions that allows for easy storage and analysis of game data, making it a great platform for our studies.

3.2 Agents and Levels

3.2.1 AI Partners. For our study, we developed three different AI partners for the circle character to interact with the human player who, in turn, controlled the rectangle character. Their behaviour represented three levels of initiative:

- **Leader without shift (L)** - the agent always acts according to its own plan for gathering diamonds in the levels, expecting the human player to comply and follow the plan
- **Follower without shift (F)** - the agent consistently aligns its actions with the human player's plan, never taking the initiative to pursue its own plan
- **Shifting between leader and follower (S)** - the agent initially assumes a leader role, but it will give up its initiative and assume a follower role if the human player does not follow the action from the agent's plan (and vice versa, if the human player is slow to act on their plan the agent assumes the leader role again)

3.2.2 Behaviour and Design Rationale by Levels of Initiative. We designed the behaviour of our three agents throughout the game in the following way:

- **Agent L** had a predefined plan with the order to collect the diamonds. To communicate this plan to the rectangle controlled by the human, the circle agent moved close or positioned itself under the target diamond. It then jumped up and down in the same position until the rectangle got closer and could stabilise the circle on its surface. Following this, the circle jumped towards its target, unless the rectangle moved away from it, prompting the circle to go back to its initial behaviour of communicating its goal.
- **Agent F** followed the human player's plan to collect the diamonds. When the rectangle, played by the human, stopped its movement, the circle agent was triggered to jump on its surface. Moreover, the circle also followed the rectangle's movement until it stopped. After stabilising the circle on the rectangle's surface, the human player could carry the circle to their desired position. If the rectangle stopped moving near a diamond, the circle agent would perceive this diamond as the human's current target and jump towards it.
- **Agent S** started the level by acting with a leader behaviour similar to Agent L. If the human player did not promptly follow its plan (longer than 2 seconds) - detected by recognising the proximity of the rectangle to the circle - the circle agent would switch to a follower behaviour equivalent to Agent F. If the human player stopped its movement for a certain period of time (2 seconds, as well) while the circle was in follower mode, without being close to any target diamond, the circle would shift back to leader mode. Then, the plan to collect the diamonds would also shift depending on which character was leading. This initiative shift was not communicated in an explicit way to the player.

Given the limited actions and lack of explicit communication mechanisms between the circle and rectangle characters, simple nonverbal communications acts [40] had to be defined to allow the human player and agents to "understand" each other and build a shared plan. This is in line with Heider and Simmel [18] who studied how people attribute motives and actions to geometric figures by simply interpreting their movements in terms of human behaviour. A common nonverbal communication strategy involves the use of "deictic gestures (i.e., pointing) [...] to direct shared attention between people and signal toward specific objects" [40]. In line with this, we used gestures such as jumping up and down for the circle and stopping in place for the rectangle to signal interest and indicate target goals based on the actions available to the game characters. Conversely, we chose not to define explicit communication acts for the human-controlled characters (one could imagine e.g., making slight left-and-right motions) to communicate the player's goals. The rationale is that the human player should not be required to learn how to trigger the agent in follower mode.

Another communication strategy we applied is known as "behavioural mirroring" and aims to provide a "subconscious signal to convey a listener's attention" [40]. When the circle agent in follower mode was "attentive" to the human player's plan, the circle would follow (or mirror) the movement of the rectangle, for example, moving left when the rectangle moved left. In addition, the predefined plan of the leader agent was intentionally designed not to align with the plan we expected the human players to follow (i.e., the optimal plan). This decision was made to allow participants to easily recognise the leader agent's communication mechanism of jumping up and down, as we assumed it would not be evident if the agent and the human shared the same game plan.

3.2.3 AI Agent Implementation. Implementation of the agents combined reinforcement learning with a rule-based approach. It is based on a version of an agent submitted to the *Geometry Friends* competition[4], adapted for the circle agent and the corresponding study conditions. In leader mode, the planning of the level is based on a predefined order of diamonds. In follower mode, it is based on the proximity of the rectangle to the diamonds. Agents employ a Q-learning policy for motion control with a discrete state space (position and velocity) to find the optimal action (accelerate or "deaccelerate"), which translates into rolling left or right depending on the movement direction. Communication mechanisms between the game characters (e.g., jumping up and down for the target diamond, following the rectangle movement, detecting when the rectangle stops its movement) are handled using a rule-based approach.

[4]See first place of https://geometryfriends.gaips.inesc-id.pt/results?competition=GF-CoG+2019+-+Cooperation+Track

3.2.4 *Study Levels.*

Concerning the game levels, there were a total of 15 collaborative levels, organised into three sets of 5 versions to be played with each AI partner. These sets were comparable in terms of complexity and can be seen in Figure 2. Each version included four diamonds, three of which required the collaboration between the circle and the rectangle, and one diamond that could only be caught individually by the rectangle (i.e., one individual diamond). Given that the rectangle was controlled by the human, the individual diamond was consistently positioned on the opposite side of the first diamond in the agent's plan when in leader mode. This placement allowed **Agent S** to shift initiative, meaning that if the human player collected the individual diamond first and did not immediately follow the agent's lead, the agent was more likely to shift to follower mode. A set of 10 initial training single-player levels were also defined, in which participants played as the rectangle character before the study, to help them become familiar with the goal, controls and mechanics of the game before interacting with the agents in the final study.

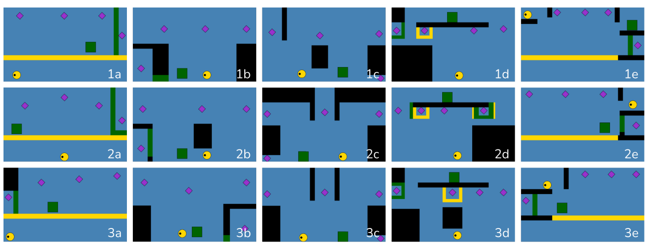

Figure 2: Three sets of *Geometry Friends* levels: Each level (number) consists of five versions (letter). Sets share similar complexity, presenting three collaborative diamonds and one individual diamond for the rectangle in each level.

Figure 3 illustrates the differences between the leader and follower modes of the circle agent in Level *1a*. In leader mode (upper row), the agent executes its own predefined plan, initially going to the diamond on the left. It shows its intentions by jumping up and down to indicate to the human partner, controlling the rectangle, that it should follow and create a platform for the circle to reach the diamond. After obtaining the first diamond, the circle moves to the next target on its plan, the diamond on the right, without waiting for the rectangle to carry it to this target. In follower mode (lower row), the agent aligns with the human player's plan, who decides to go to the diamond on the right first, waiting there for the circle. Then, the circle agent jumps on top of the rectangle to collect this diamond. Afterwards, the human player may use the rectangle to stabilise the circle on its surface and transport it to their next desired target, the diamond in the middle, with the circle synchronising its movements with the rectangle.

4 STUDY METHODOLOGY

Our study aims to explore perceptions of the self, agent and team in a collaborative game with three within-subjects conditions based on two variables: the agent's initiative level (follower or leader) and its willingness to shift initiative. We counterbalanced the conditions to minimise order effects.

4.1 Setup

The study was conducted in-person at three universities in Portugal, Sweden, and France. Participants were placed in isolated rooms to minimise distractions. Main tools included a computer to run *Geometry Friends*, an additional screen/computer for the questionnaire (JotForm[5]), a game controller, and a stopwatch to time each experimental condition of the study.

All subjects participated voluntarily upon request of the experimenter (Portugal and France), or by registering through social media (Sweden). Participants were offered participation prizes, such as movie tickets in Portugal and vouchers in Sweden.

On-site researchers received instructions on conducting the study, welcoming participants, assigning them to experimental conditions, and addressing any concerns they had. The game software and agents' implementation are publicly available on GitHub [6], while the questionnaires can be found in the Supplementary Material section.

4.2 Procedure

4.2.1 *Protocol.*

Participants took around 45 minutes to complete the study. Each participant started by reading and accepting the study's consent form. Then, they filled a questionnaire with demographic information, including age, gender, education, gaming habits/preferences and prior experience with artificial intelligence. Afterwards, participants began the training phase, playing up to 10 single-player levels of the game as the rectangle, until they felt comfortable with the game controls and mechanics. During the main session, subjects played up to five levels of the game with each of the three agents. Participants played approximately 5 minutes per agent, timed by the researcher on-site. This constraint aimed to help participants understand the differences in agent behaviours and their corresponding perceptions, while avoiding an extensive interaction that could lead to loss of focus or forgetting previous interactions, as this was a within-subjects study. Participants were given the option to complete the current level if time expired. For each agent, participants completed a post-game questionnaire including questions on perceptions of the self, the agent and the team, and collaboration strategies used. Following the main session, participants completed the final part of the questionnaire, selecting the agent they preferred and justifying their choice. In the end, a short debriefing section clarifying the study goals and conditions was presented to participants. Participants received similar rewards in Sweden and Portugal. However, due to French legislation, participants in France were unable to receive compensation for their participation. Nevertheless, the study setup remained consistent across locations.

[5] https://eu.jotform.com/myforms/
[6] https://github.com/1000obo/geometry-friends-study

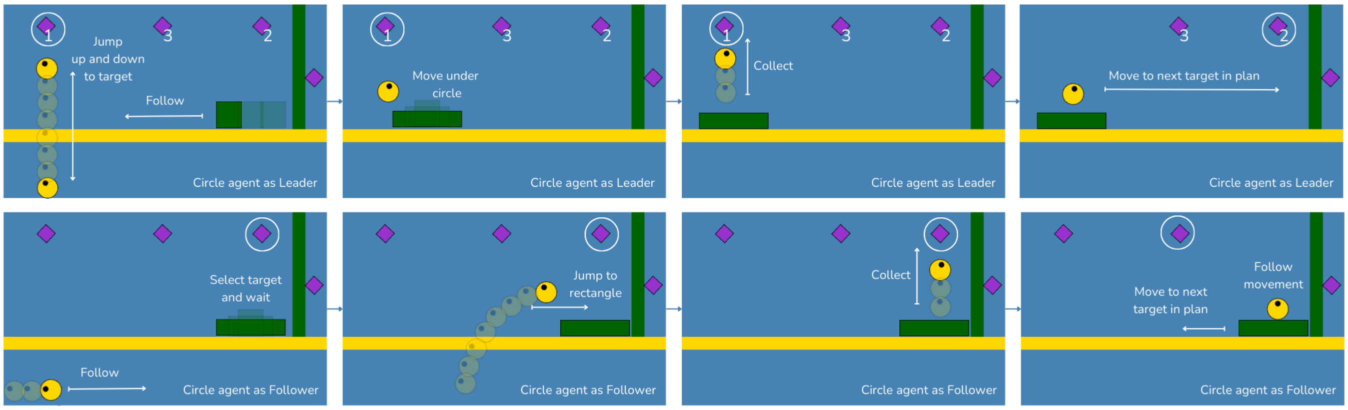

Figure 3: Leader and Follower mode examples for Level *1a* (left to right). Upper row - Agent leads (L): the circle jumps up and down to signal its desired targets to the rectangle. Lower row - Agent follows (F): the circle follows the rectangle's trajectory to the desired targets and the rectangle is also able to carry the circle on its surface.

4.2.2 Conditions. During each session, participants experienced three conditions, playing with each of the three agents (**L**, **F**, **S**). Participants were assigned to one of six different orders which was consistently maintained across all study locations. In the post-game questionnaire, we also assessed whether participants perceived each agent as expected using two manipulation checks. The first one asked about the agent's level of initiative on a scale of 1 (no initiative) to 5 (a great deal of initiative). Initiative was defined as *the extent to which the agent followed their own plan and made decisions independently*, rather than simply following the participant's lead. The second one focused on the agent's initiative shift, assessing whether it changed from showing a lot of initiative to none, or vice versa, within the same level, rated on a scale of 1 (no shifts) to 5 (significant shifts). Additionally, we included a manipulation check to verify if the challenge of the levels remained consistent across all conditions. Participants rated this aspect on a scale of 1 (very easy levels) to 5 (very difficult levels).

4.2.3 Questionnaires. Concerning the questionnaire, it involved quantitative and qualitative measures, as follows. The perceived warmth and competence of the agent was measured using the Warmth and Competence Scales [7]. Participants rated the appropriateness of specific words related to warmth (pleasant, sensitive, friendly, helpful, affable, likeable, approachable, and sociable) and competence (intelligent, organised, expert, thorough) on a Likert Scale from 1 to 7. The consistency of these measures was measured using Cronbach's alpha (Warmth: α = 0.946, Competence: α = 0.899). We used a 7-point Likert scale to measure the satisfaction of the participant's own performance, agent performance, and team performance [33, 41]. We used the Social Identification Scale to measure group satisfaction [1]. Participants rated statements on a Likert Scale from 1 to 5, assessing team cohesion and member satisfaction. Cronbach's alpha was used to assess consistency (α = 0.934).

4.3 Participants

We conducted the study with a total of 62 participants. Two were excluded, one from Sweden due to significant lag in game, and another from France, who failed to submit the questionnaire. Consequently, our final dataset comprised 60 participants, with 20 from Portugal, 20 from Sweden, and 20 from France (more details about the participants in the Appendix). The study received approval from the Ethics Committee of *Instituto Superior Técnico, in Lisbon, Portugal*.

Participants' ages varied from 18 to 70 years old. The majority, 56.57%, fell within the 25-34 age range, followed by 31.67% aged 18-24. Participants aged 35 or older constituted 11.67%. Concerning gender, there were 40 male and 20 female participants. The majority (45.67%) of participants held master's degrees, followed by bachelor's degrees (23.33%), doctorates (21.67%), and high school certificates (8.33%).

Regarding gaming frequency, 48.33% were frequent players of video games, 20.00% were occasional players, and 31.67% were infrequent players. Concerning gaming devices, 44 participants primarily played on computers. 86.67% of participants were familiar with puzzle video games, while 73.33% were familiar with platformers. Regarding their preferences, 81.67% enjoyed single-player games, 61.67% cooperative gameplay, and 60.00% competitive gameplay.

Participants demonstrated a higher mean value (M = 3.767, SD = 0.551) for their positive perceptions of Artificial Intelligence compared to their negative perceptions (M = 2.901, SD = 0.682), within the 1 to 5 scale, suggesting that, on average, they held more positive views of AI, using the General Attitudes towards Artificial Intelligence Scale [35].

4.4 Data Collection

In terms of data collection, our study employed a mixed-methods research design. The gameplay during the experiment was recorded through video capture of the computer screen, along with the recording of in-game logs for each level in *CSV* format. These in-game logs included actions and positions of both game characters, agent's role (leader or follower), and current diamonds present in the level at each timestamp. Considering these, we measured for each agent the number of completed levels, their completion time, and the participant's idle time. For Agent S, we assessed the time played with the agent in leader/follower modes.

After playing with the three agents, participants selected their preferred one from a drop-down list. Qualitative questions were in regards of the specific collaboration strategies used with each agent and aspects that justified the preferred agent choice, which were collected through questionnaires.

4.5 Data Analysis

The main objective of this analysis was to determine how different levels of agent initiative are perceived by users, and how these initiatives affect the perceived collaborative qualities. We engaged in a mixed top-down (deductive) and bottom-up (inductive) thematic analysis approach [8] for the open-ended responses, which allowed us to identify patterns and strategies within the data. We support these insights with the statistical analysis of our quantitative questionnaire data using JASP[7]. Our data was first divided by agent initiative. We examined data for influencing factors on AI partner perception (RQ1), collaboration quality (RQ2), and human self-perception (RQ3) using a deductive approach guided by our research questions. Codes were assigned to each quote, and organised into sub-themes contributing to these themes. Second, we used an inductive approach to expand, divide, combine and contrast sub-themes based on the underlying data in order to enhance the understanding of patterns across all initiative types. This process resulted in four main themes, each with several related sub-themes: Perception of AI partner (5.1), perception of overall collaboration (5.2), perception of team interaction (5.4), and (human) self-perception (5.3) when interacting with AI agents. For either step, one author was mainly responsible for the coding while co-authors sampled quotes, coded and discussed them (among all authors) along the process. Representative quotes and counts for impact scale are displayed to highlight shared insights between the participants.

5 RESULTS

The results section is divided into four main sections, each aligned with the research questions outlined in the introduction: perceptions of the agents, the collaboration, the self and the team interaction, given the agents' level of initiative. We added a posterior analysis of the effects of demographic variables on agent preference. We conclude with manipulation checks and order effects.

When discussing Agent S, it should be noted that participants ended up playing in follower mode more frequently. The percentage of time played with the follower ($M = 0.636, SD = 0.382$) was higher compared to the leader ($M = 0.364, SD = 0.382$).

5.1 Perceptions of AI Partners

This section reports quantitative metrics regarding participants' perceptions of the AI partner, such as warmth, competence, and social identification with the team. We identify variables in the qualitative data (from the question on collaboration strategies) that influence these perceptions towards the agents and the team during collaboration. Figure 7 shows a summary of these results.

[7]https://jasp-stats.org/

5.1.1 AI Partner Warmth and Competence. Agent F was perceived as significantly higher in warmth ($M = 4.558, SD = 1.468$) compared to Agent L ($M = 3.794, SD = 1.469$), $t(59) = 3.810, p < 0.001$, and Agent S ($M = 4.127, SD = 1.470$), $t(59) = 2.457, p = 0.017 < 0.05$, as seen in Figure 4. The difference in warmth between Agents S and L was not significant. Concerning competence, Agent F also had the highest scores ($M = 4.388, SD = 1.595$), followed by Agent S ($M = 4.304, SD = 1.523$), and Agent L ($M = 4.188, SD = 1.473$), but these differences were not significant, as seen in Figure 5.

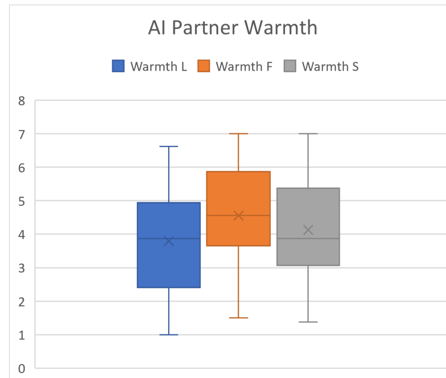

Figure 4: Perceived AI Partner Warmth for each condition (L, F, S). Agent F scored significantly higher in warmth compared to the other agents.

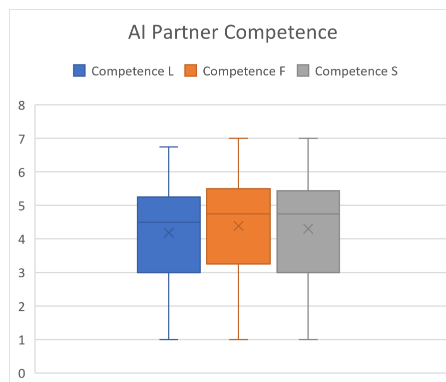

Figure 5: Perceived AI Partner Competence for each condition (L, F, S). Agent F had the highest score on competence, but no significance was found.

5.1.2 Social Identification with the Team. Social identification with the team was higher with Agent F ($M = 2.967, SD = 0.902$) compared to Agent L ($M = 2.747, SD = 0.926$), although not significant. Social identification with Agent S ($M = 2.697, SD = 0.919$) was lower than other conditions. There was a significant difference between social identification with Agent F ($Mdn = 2.800, M = 2.967, SD = 0.902$) and Agent S ($Mdn = 2.600, M = 2.697, SD = 0.919$), $W = 828.500, p = 0.005 < 0.05$. These results can be seen in Figure 6.

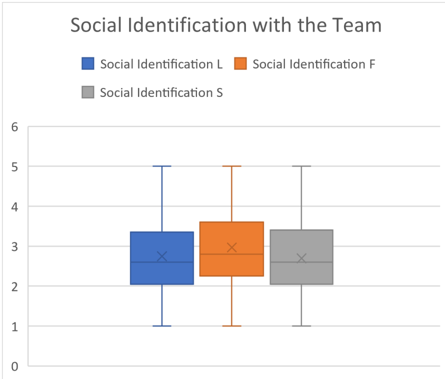

Figure 6: Social Identification with the Team for each condition (L, F, S) with Agent F being significantly higher than Agent S.

5.1.3 AI Leader Perceived as Eager but Stubborn Lacking Attentiveness to Human Partner. We identified in our qualitative analysis that agent L was recognised for its determination and initiative, yet perceived as inattentive to its human partner. It consistently followed its own plan, resulting in a few participants feeling lack of control.

Initiative Some participants (14/60) highlighted Agent L's initiative, describing it as "independent" (F10) and "eager" (S3), having great "determination" (S10). They appreciated its ability to communicate its goals clearly, so participants did not have "to guess" (P12) its plan. Nevertheless, it was perceived as "stubborn" (P18, S20) and unwilling to change plans (F15), feeling "stupid" (S2) and making mistakes (S6).

Control A few participants (8/60) perceived themselves as an "helper" (S9) or a platform to boost Agent L (S18, S11), seeing its communication strategy (jumping in desired position) as a means to become "more helpful" (P9). However, a few (5/60) were dissatisfied with this lack of control, feeling no "influence in the level" (S17). It was unpleasant to "suppress my [participant] will and let him [Agent L] lead" (F15). A few participants (3/60) adjusted their leadership role after trying to act as a leader in the beginning of the level (F6, F8).

Attentiveness A few participants (7/60) were frustrated with Agent L's lack of attentiveness, feeling ignored (S10, P16). They perceived the agent as "less sociable" (P16), and felt like it did not recognise their capabilities (S15). Moreover, one participant felt like "the failure or success [of the team] seems depend really on me [participant] alone" (F17), as the agent had repetitive and non-adaptive actions.

5.1.4 AI Follower Perceived as Slower But Warmer Showing Attentiveness to Human Partner. Agent F was seen as warmer and more attentive to the human partner. A subset of participants also enjoyed the level of control they had with the agent, recognising that it followed their plan. However, a few participants felt it was slower and less intelligent.

Initiative Some participants (16/60) understood that Agent F followed their plan, perceiving it as "friendlier" (P5) and "more cooperative" (P5, P14, P19). Mutual help allowed for "great teamwork" (S1) and their interaction felt "natural" (F13). Nevertheless, Agent F was described as "unpredictable" (S12), less intelligent (S15), and slow, taking "a lot of time" to move (F14). A few participants (6/60) attributed initiative to this agent (e.g., "the ball [Agent F] had the initiative to jump when the goal was above it" (P2)).

Control A few participants (4/60) enjoyed having control when playing with Agent F, preferring to "solve the level in the way I [participant] wanted to" (S17) rather than "following its pattern" (F2). One participant also mentioned that Agent F "was more willing to centralise with" them (P14). Nonetheless, another felt like collaboration was challenging and "out of my [participant] control" (F14).

Attentiveness A few participants (4/60) perceived Agent F as being attentive to their movements, understanding "that when I [participant] turned into a "long" rectangle it [Agent F] jumped instantly to get the diamond" (S13). Moreover, it appeared aware of its limitations, as it "stopped moving when there were no diamond they [Agent F] could reach anymore", letting the participant finish the task (F6).

5.1.5 Shifting AI Partner Perceived with Mixed Perceptions and Unpredictable Control of the Game. Feedback on Agent S varied, with a few participants appreciating its adaptability, and others finding it unhelpful and inconsiderate of the human partner. A few participants felt they had to fight for control of the plan with this agent.

Initiative Agent S was attributed more initiative by a few participants (6/60). One compared it to a "cat", because the agent was "the one telling me [participant] what to do" (P4). Some participants (11/60) noticed its initiative shift. Participants found Agent S "organized" (P2), "smarter" (S10) than Agent F, and "reactive and adaptive to my [participant] movement" (F2). However, a few (4/60) found it "frustrating to work with" (P5), difficult to change its plan (S8, F15) and wanted to "skip playing with this AI partner" (S1).

Control A few participants (4/60) felt like they were "not affecting the outcome" (S2) or had to dispute for control with Agent S (S8, F15, S13). Initially, they felt like their plans aligned, feeling "a good connection" (S8) and balanced decisions (F15). However, the agent sometimes changed plans (S13), "wasn't as easy to coerce" (S8), and if the participant "wanted to do something else, I [participant] always had to lose the argument" (F15) - being described as an "autocrat" (F15).

Attentiveness A few participants (6/60) perceived Agent S as "more oriented to listening" (F19), even when attributed initiative (P9). It knew "what strategy it wanted to use, in a way that involved me [participant]" (S15), and recognised that it should not jump "when there is no diamond" (F13). Only one participant felt like the agent worked "independently" (S19) of their goals.

5.2 Perceptions of Overall Collaboration

This section presents quantitative metrics assessing the effectiveness of the collaboration, such as performance satisfaction of the team and AI partner and objective performance. We also describe the effect of different levels of agent initiative on performance perceptions in the qualitative data. Figure 11 shows a summary of these results.

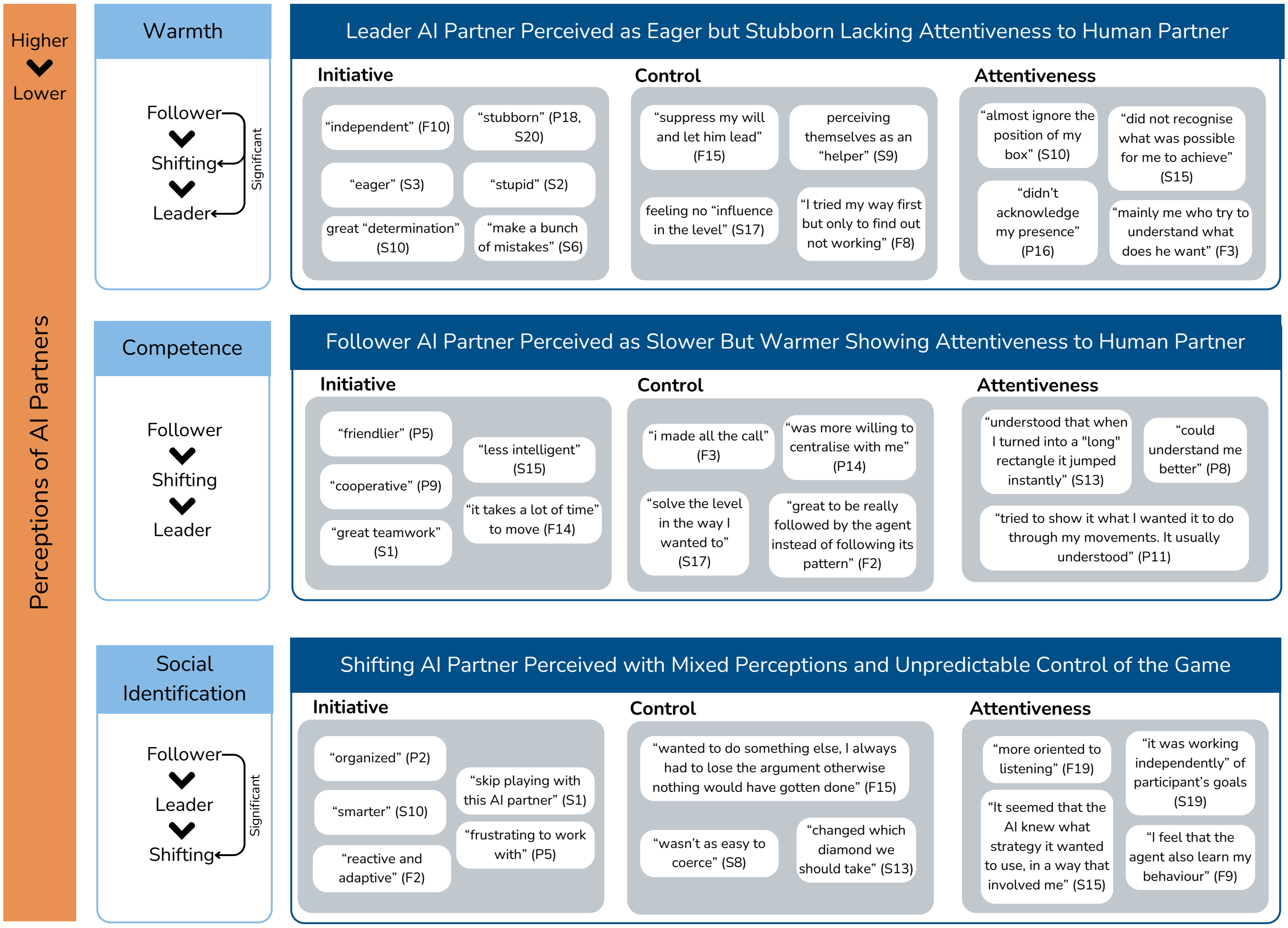

Figure 7: Summary of main quantitative (left column) and qualitative (right column) results on AI partner perceptions. Quantitative results include perceived AI partner warmth and competence, as well as social identification with the team. Qualitative results show representative quotes on perceptions of the leader, follower, and shifting agents, organised by initiative, control, and attentiveness.

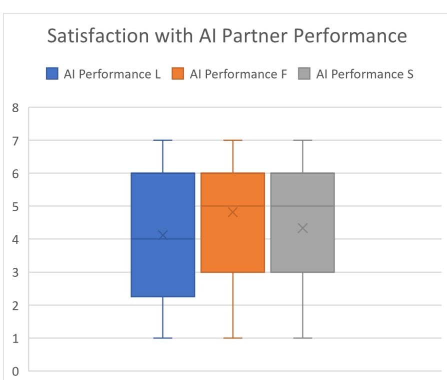

Figure 8: Satisfaction with AI Partner Performance for each condition. Satisfaction with Agent F performance was highest compared to other agents, and significant to Agent L.

5.2.1 Satisfaction with Team and AI Partner Performance. Satisfaction with AI partner performance was significantly higher for Agent F ($M = 4.817, SD = 1.751$) compared to Agent L ($M = 4.117, SD = 1.851$), $t(59) = 2.284, p = 0.026 < 0.05$. While satisfaction with Agent F was higher than Agent S ($M = 4.333, SD = 1.772$), it was not significantly different. These results can be seen in Figure 8. Regarding team performance, satisfaction with Agent F ($Mdn = 5.000, M = 4.783, SD = 1.574$) was significantly higher compared to Agent S ($Mdn = 5.000, M = 4.217, SD = 1.795$), $W = 690.000, p = 0.049 < 0.05$. Similarly, satisfaction with team performance with Agent F was higher than Agent L ($Mdn = 5.000, M = 4.483, SD = 1.652$), though not significantly different. These results can be seen in Figure 9.

5.2.2 Objective Performance. Participants completed more levels in condition L ($M = 2.967, SD = 1.540$) compared to conditions F ($M = 2.750, SD = 1.445$) and S ($M = 2.700, SD = 1.430$), but these differences were not significant, as seen in Figure 10. We did not find any significant differences across conditions in the time spent to complete the levels.

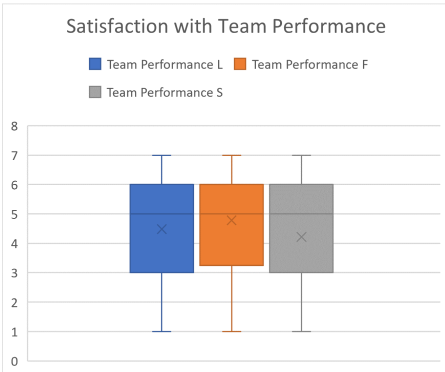

Figure 9: Satisfaction with Team Performance for each condition. Satisfaction with team performance with Agent F was highest and significant to Agent S.

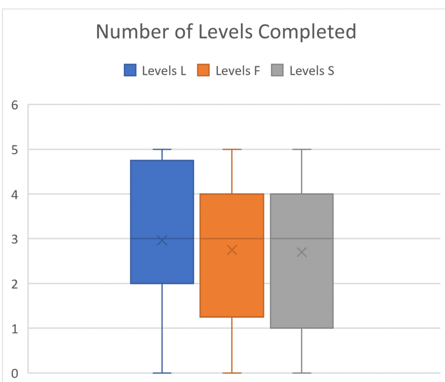

Figure 10: Levels Completed for each condition by participant. More levels were completed with Agent L, followed by Agent F, and then Agent S with no significance.

5.2.3 Impact of AI Partners' Initiative Level on Performance Perceptions in Collaboration. Based on our qualitative analysis, we found that some participants (11/60) perceived Agent F as more stable, faster and skilful, making a participant feel like they "played better" (P8) as well. A few (5/60) shared similar perceptions about Agent S, which felt "better at getting onto of the rectangle" (S7). While Agent L was described as effective by a few participants (5/60), despite not changing strategy (F7), others (11/60) felt like it "lacked efficiency" (P5). It was trying to "finish the game as quickly as possible which was not a good strategy" (S6), making more mistakes. One participant felt like Agent L's lack of trust in them affected team performance (F11). Participants also criticised the performance of Agents F (9/60) and S (5/60) concerning their lack of stability on the rectangle.

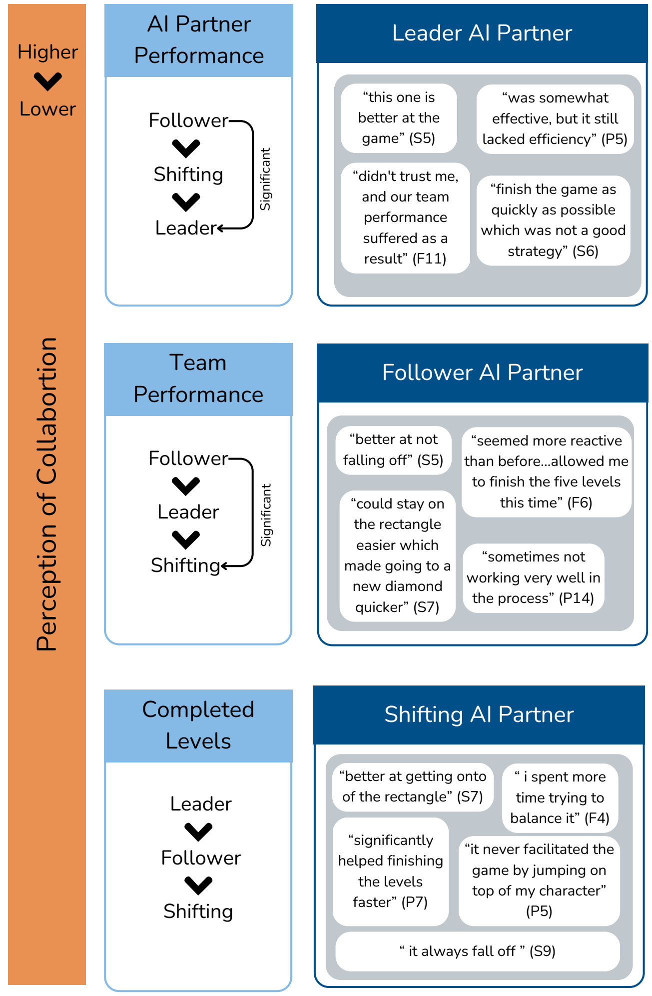

Figure 11: Summary of main quantitative (left column) and qualitative (right column) results on collaboration. Quantitative results include satisfaction with AI partner and team performance, and completed game levels. Qualitative results show representative quotes on perceptions of the collaboration with the leader, follower, and shifting agents, in particular related to their performance.

5.3 Self-Perceptions

This section presents metrics tied to participants' self-perceptions during collaboration. We analyse satisfaction with self-performance and participant idle time, which might be linked to the roles participants perceive themselves to be in collaboration. We include some qualitative references on self-perceptions from the previous results sections. Figure 12 presents a summary of these results.

5.3.1 Satisfaction with Self-Performance. Satisfaction with self-performance was higher with Agent F ($M = 4.700, SD = 1.576$), followed by Agent L ($M = 4.683, SD = 1.396$), and then Agent S ($M = 4.367, SD = 1.636$). These differences were not significant, as seen in Figure 13. In condition L, satisfaction with self-performance ($M = 4.683, SD = 1.396$) was significantly higher than satisfaction with AI partner performance ($M = 4.117, SD = 1.851$), $t(59) = -2.508, p = 0.015 < 0.05$, as seen in Figure 14. For the other conditions, no significant differences were found, even when satisfaction with team performance was lower (minus or equal to 3).

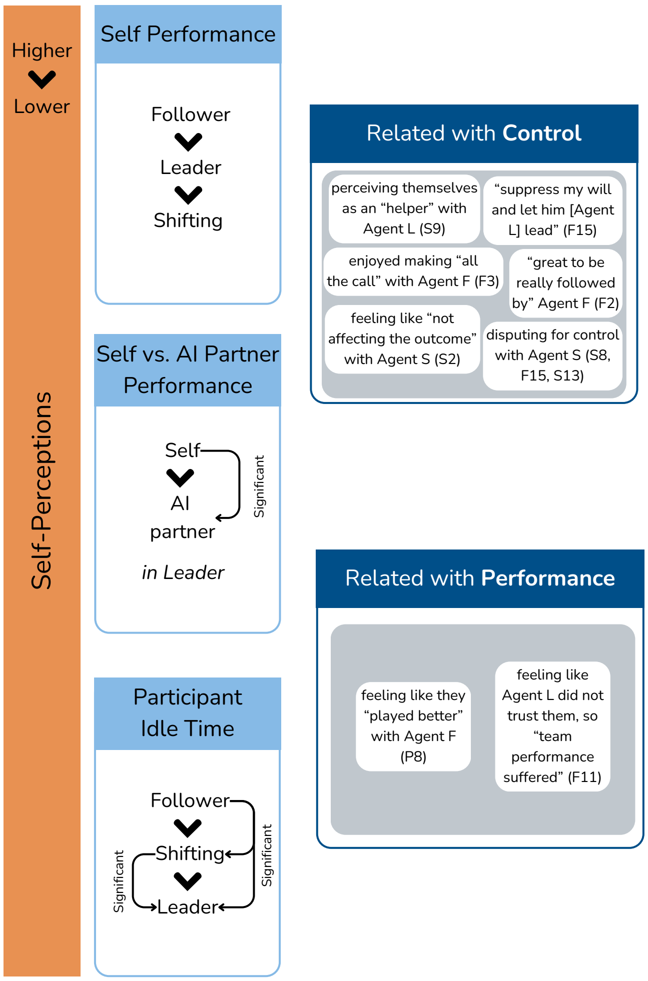

Figure 12: Summary of main quantitative (left column) and qualitative (right column) results on self-perceptions. Quantitative results include satisfaction with self-performance (and its comparison with AI partner performance satisfaction in the leader condition), and participant idle time. Qualitative results show representative quotes on self-perceptions, related to the topics of control and performance, highlighted in the previous results sections.

5.3.2 Participant Idle Time. The percentage of participant idle time in the levels completed was higher with Agent F ($M = 0.320, SD = 0.105$), followed by Agent S ($M = 0.283, SD = 0.116$), and then Agent L ($M = 0.230, SD = 0.104$). There was a significant effect of the agent type on this measure, $F(2, 502) = 30.189, p < 0.001$. These results can be seen in Figure 15.

Figure 13: Satisfaction with Self-Performance for each condition. Satisfaction with self-performance was higher with Agent F, but no significant differences between conditions.

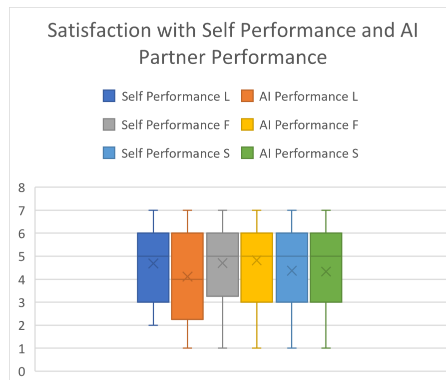

Figure 14: Satisfaction with Self and AI Partner Performance for each condition side by side. In the leader condition, satisfaction with self-performance is significantly higher than satisfaction with the AI partner performance.

5.3.3 Impact of AI Partners' Initiative Level on Participants' Self-Perceptions of Control and Performance in Collaboration. In the first results section, we observe participants' self-perceptions on control in collaboration. A few participants (8/60) felt like they had an auxiliary role with Agent L, while with Agent F they (4/60) enjoyed leading and having control. With Agent S, participants (4/60) felt a lack of influence in the game or had to fight for control. In the second results section, self-perceptions seem to be linked with performance perceptions as well. One participant felt like their performance improved with Agent F, since this agent played better. Conversely, with Agent L, one participant perceived a lower level of trust from the agent, which led to lower team performance.

5.4 Perceptions of Team Interaction

This section analyses participants' preferred AI partners and the reasons for these preferences in a team, which are related to perceptions of the AI partner and collaboration. A summary of these results can be found in Figure 16.

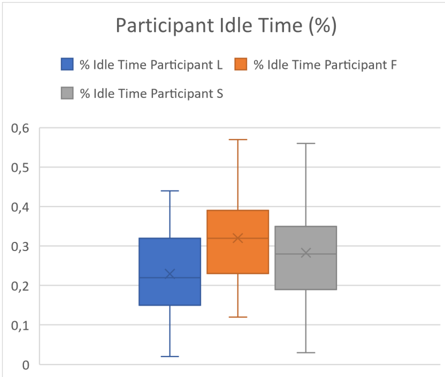

Figure 15: Percentage of participant idle time for each condition. Participant idle time was higher when playing with Agent F, followed by Agent S, and then Agent L, and these differences were significant.

5.4.1 AI Leader Performance and Predictability Preferred Over Friendliness in Collaboration. 30% (18/60) of participants preferred Agent L for being proactive and quicker in executing its plans, despite not following their plan and being perceived as less friendly. This section analyses qualitative responses from participants who preferred Agent L.

Perceptions of AI Partner Some participants (6/18) perceived Agent L as having more initiative, stating that "it did not just follow me around and waited for me [participant] to decide for it" (P17), unlike Agent F. A few participants (2/60) felt like assistants to Agent L (S10, F17), which was compared to a "human friend" by one participant, making them "chuckle a few times" for its occasional mistakes (P4). Its clear and understandable plan contrasted with Agent F's (S10), feeling "less frustrating" not having "to guess" (P12). Some participants (6/18) also felt " more part of a team" (P13) and better "chemistry" (S11) with Agent L.

Perceptions of Collaboration Most participants (11/18) felt like performance was better and collaboration easier with Agent L, for its initiative (P10, P16, F17), predictability/mechanical behaviour (S16, F6), and precision (S12, F5). Agent L having "its own goals and that I [participant] was just an help for it to success [sic] was easier – in terms of task completion" (F17). A few participants (3/18) appreciated the "slightly faster gameplay" (S3) and decision-making (S3) compared to other less proactive agents (P13).

Preference for Consistency and Efficiency Over Adaptability A few participants (3/18) that favoured Agent L had negative perceptions, calling it "the dumbest" (F6), "cold, and unfriendly" (S10), and the one they liked less (P16). However, they prioritised its predictability (F6) and efficiency in completing levels (S10, P16) over adaptability and warmth, finding "intelligent and adaptive" agents difficult to synchronise (F6) and less efficient (S10).

5.4.2 AI Followers Agreeableness and Control Preferred Over Performance in Collaboration. Approximately 42% (25/60) preferred Agent F, mentioning its ability to understand them, adapt to their actions, and its perceived warmth. However, a few participants felt it might not be the most efficient. This section analyses qualitative responses from participants who favoured Agent F.

Perceptions of AI Partner Some participants (8/25) enjoyed Agent F for following their initiative, contrasting with Agent L, which showed no "consideration on my [participant] side" (S18). Some participants (9/25) noted better collaboration and teamwork with Agent F, which "adapted accordingly" (P11) to their goals, compared to other agents (P5). It was described as "likeable" (S1), "balanced" (S2), "reliable" (S5) and "responsive" (S19). They felt that Agent F respected (F11) and trusted them (F11) as a teammate, understanding "what I [participant] tried to communicate" (P11). Additionally, some participants (7/25) felt in control with Agent F (2S, S17), deciding "which diamond to aim for" (S13) in the fastest way (P6). In contrast, one participant perceived Agent L as having more control, being "the real boss" (S1), while with Agent S they had to "show the other [Agent S] who is [sic] boss" (S1).

Perceptions of Collaboration Some participants (7/25) described Agent F as "more reactive and with better skills" (F2). Moreover, a few participants (3/25) enjoyed Agent F's patient approach, waiting for them to adjust the rectangle for its jump (P14). They noted that the "game reward us for being careful" (F3), so being "hasty" (F3) like Agent L "make [sic] us waste time" (F3) and disrupts the "organization and stability of the team" (S6).

Preference for Agreeableness and Followership Over Efficiency While some participants were satisfied with Agent F, a few (3/25) felt it made "silly mistakes" (S8), being less efficient and "clumsier so the goal [sic] were harder to achieve" (F15). Nevertheless, they preferred having control with an agreeable agent over one that does "its own path" (F2) and makes participants feel less involved and "less proud to be part of the team" (S18). Overall, the other agents felt "really stubborn and it resulted in a more tensed and frustrating experience, even if we achieved the goals more effectively" (F15).

5.4.3 Preference in Being Followed by AI Partner with Shifting Initiative in Collaboration. Around 28% (17/60) of participants preferred Agent S. While a few participants enjoyed its initiative balance, many of those who selected it as the most enjoyable found it similar to Agent F. This section analyses qualitative responses of participants who preferred Agent S.

Perceptions of AI Partner A few participants (2/17) highlighted Agent S's initiative shift, feeling like "it had the better balance of initiatives" (F4), and that "the decision to the order of the diamonds was split" (P9) between them and the agent. Around half (8/17) found it more collaborative and adaptive to the participant's actions, feeling "part of a team" (F9), great teamwork (S14) and an enjoyable "partnership" (F4). Moreover, they perceived Agent S as more aware of the game environment (S15, P18) and "less stubborn" (P7) than other agents. It was also perceived as communicating more (S15, P7), although not always effectively (S15).

Perceptions of Collaboration A few participants (3/17) expressed greater satisfaction with Agent S's performance, noting it made less "obvious mistakes like jumping to [sic] early and getting in the way" (S7) and it was more "speedy" (14F). Nonetheless, one participant found Agent L more efficient and easier to adapt to than Agent S (P9). Despite this, due to Agent S's initiative balance, they preferred this agent, as it contributed to "a less mechanical game, but a more natural one" (P9).

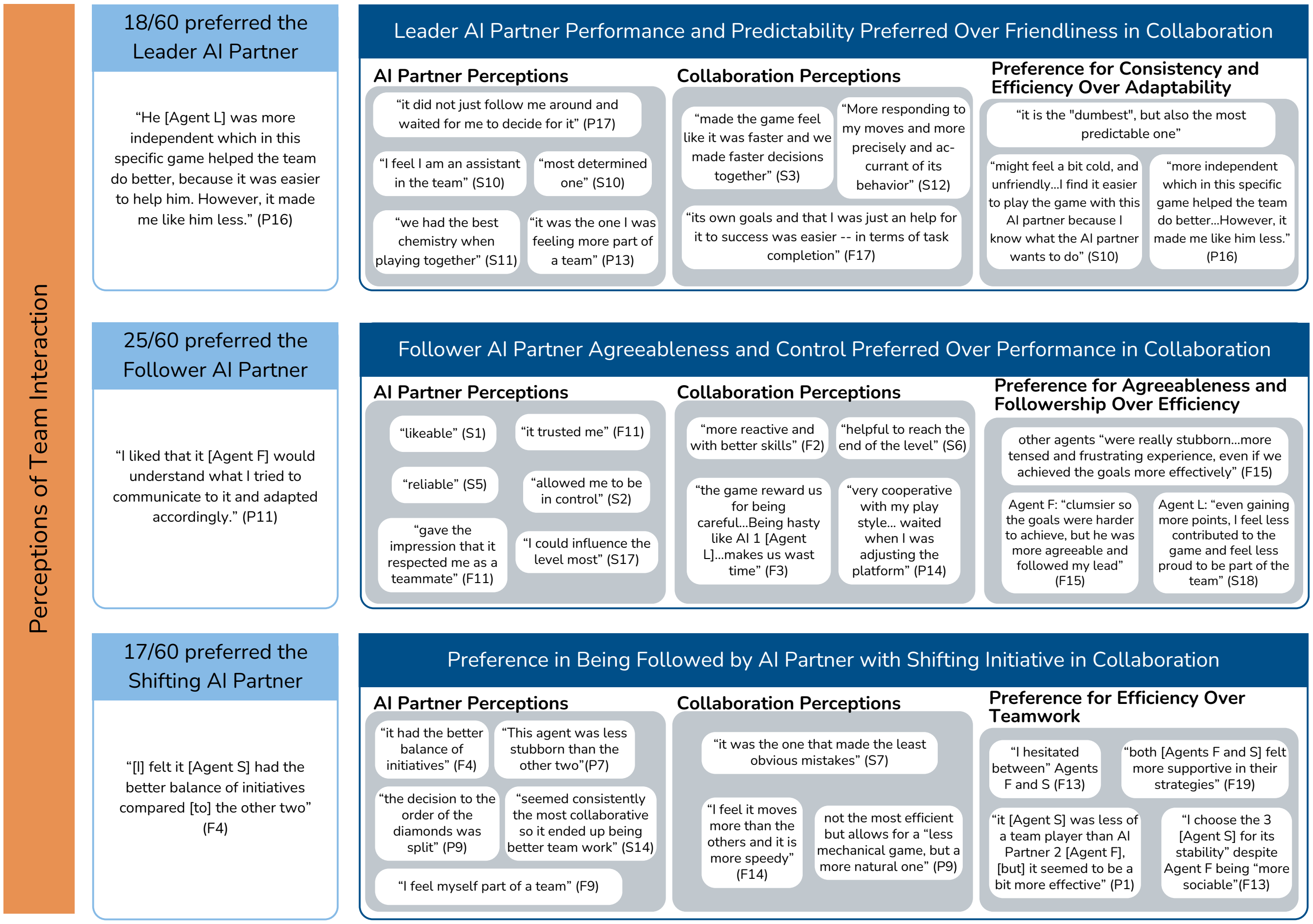

Figure 16: Summary of main quantitative (left column) and qualitative (left and right columns) results on team interaction perceptions, in particular related to AI partner preference, as shown in the quantitative results. Qualitative results show representative quotes for each agent (leader, follower, shifting) concerning this preference. These are organised mainly into perceptions of the AI partner and collaboration.

Preference for Efficiency Over Teamwork Paritcipants hesitated choosing Agent F or Agent S as the most enjoyable (F13), since they were perceived as similar, with one participant stating they "didn't notice the difference" (F16). Both agents felt "more supportive in their strategies" (F19), but some differences were found. A few participants (2/17) felt Agent S "was less of a team player" (P1) and less sociable (F13) than Agent F. Nevertheless, they still perceived Agent S as "a bit more effective in getting to the goals" (P1) and having more "stability" (F13), leading to a preference for it.

5.5 Demographics Influence on AI Partner Preference

In a posterior analysis, we assessed the influence of demographic data on AI partner preference.

Gender Significant gender-based differences were found in agent preference, $F(2, 58) = 7.558, p = 0.008 < 0.05$ (see Fig. 17). Male participants favoured Agents F and L ($M = 1.800, SD = 0.723$), while female participants preferred Agent S ($M = 2.350, SD = 0.745$).

Female participants tended to enjoy more collaborative and attentive behaviours, preferring Agent S (10/17), followed by Agent F (7/25), and then Agent L (3/18). A few (2/17) mentioned training effects for Agent S, which might have made the difference between

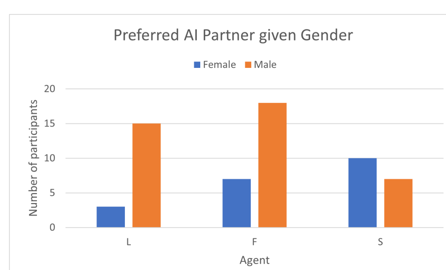

Figure 17: Distribution of AI Partner Preference by Gender with significant differences between female and male.

choosing Agent F or S. Male participants favoured Agent L (15/18) for its predictable and quick behaviour over being adaptable, but many also enjoyed Agent F (17/25). There is, however, a tendency for female participants to choose Agent L less compared to other agents that feel "more supportive in their strategies" (F19).

Experience and Preferences in Video Games and Prior Perceptions of AI Regarding experience with video games, no significant differences were found between frequent and non-frequent players. However, frequent or casual players tended to prefer Agents F and L more ($M = 1.854, SD = 0.727$) than non-frequent players ($M = 2.263, SD = 0.806$), who favoured Agent S. A trend was noted in gameplay preferences, with 77.78% of participants who chose Agent L liking competitive games, followed by 64.00% of those who chose Agent F and 35.29% of those who chose Agent S, but this difference was not significant. No relevant differences were identified for the AI partner preference given prior perceptions of AI.

Location There were no significant differences in preference across all study locations. Nevertheless, Sweden and Portugal had a similar pattern, with participants favouring Agent F, followed by Agent L and then Agent S. In contrast, participants from France preferred Agent S, followed by the other two agents.

5.6 Manipulation Checks

By analysing the manipulation checks in the post-game questionnaire, we found that Agent L was perceived with significantly higher initiative ($Mdn = 4.000, M = 3.617, SD = 1.151$) than Agent F ($Mdn = 3.000, M = 2.883, SD = 1.121$), $W = 984.000, p = 0.002 < 0.05$, and Agent S ($Mdn = 3.000, M = 3.117, SD = 1.010$), $W = 726.000, p = 0.006 < 0.05$. These results can be seen in Figure 18. However, as seen in Figure 19, we did not find significant differences in the initiative shift score across conditions, even though this score was higher for Agent S ($Mdn = 2.000, M = 2.267, SD = 1.103$), compared to Agents F ($Mdn = 2.000, M = 2.033, SD = 1.089$) and L ($Mdn = 1.000, M = 1.983, SD = 1.282$). There were no significant differences across conditions on the perceived challenge of the levels played.

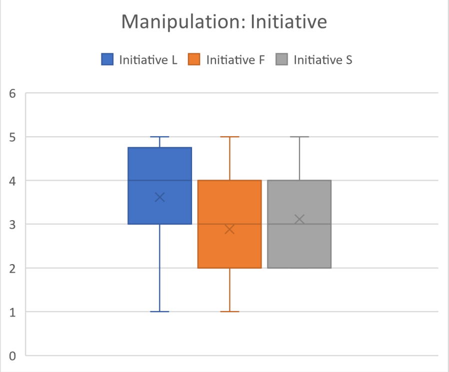

Figure 18: Initiative manipulation check for each condition with significantly higher initiative of Agent L.

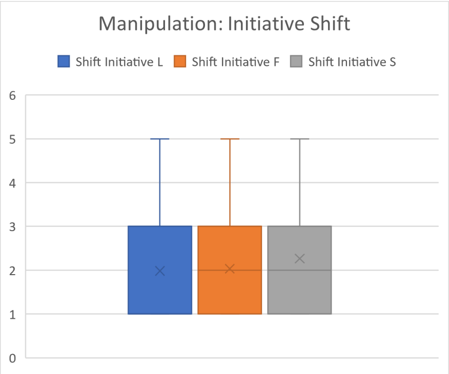

Figure 19: Initiative Shift manipulation check for each condition showing higher measures for Agent S (not significant).

5.7 Order Effects

In this section, we report on order effects we found for some of the measures presented before.

Competence scores of Agent S ($F(2, 57) = 4.934, p = 0.011 < 0.05$) and Agent F ($F(2, 57) = 2.341, p = 0.034 < 0.05$) significantly increased with the presentation order of condition S, being higher when this condition was presented last.

Satisfaction with Agent S performance significantly increased with the presentation order of condition S ($F(2, 57) = 2.829, p = 0.019 < 0.05$). Satisfaction with Agent L performance ($F(2, 57) = 3.157, p = 0.036 < 0.05$) and its team performance ($F(2, 57) = 2.464, p = 0.021 < 0.05$) significantly decreased with the presentation order of condition F, being lower when this condition was presented last.

Participants' idle time increased with the order of presentation of the conditions ($F(4, 496) = 2.638, p = 0.033 < 0.05$).

Order effects were found for AI Partner preference based on the presentation order of conditions L ($F(2, 57) = 3.938, p = 0.025 < 0.05$) and S ($F(2, 57) = 5.270, p = 0.008 < 0.05$), indicating participants' tendency to prefer the agent they have seen last in the conditions, if this was L or S.

6 DISCUSSION

We conducted a study with 60 participants using the collaborative game *Geometry Friends*. Participants played with three agents showing different levels of initiative: leader, follower and shifting. We evaluated how initiative affected participants' perceptions of the agents, collaboration, self and team interaction.

In this section, we discuss the study results for each research question and relate them with prior human-AI collaboration research. We identify factors that should be considered when designing initiative in collaborative agents, such as situational context and individual preferences. We conclude with study limitations and ideas for future work.

6.1 Research Questions

RQ1: How does an agent's initiative influence the perception of AI partners?

Participants perceived the follower agent as warmer and more competent than the other agents. Research in human-robot interaction shows that warmth and competence are key factors in building trust [13]. Social identification with the team was also higher with the follower. While the leader's determination was appreciated, participants found it stubborn and inattentive, with some reporting a lack of control in the game. The follower was perceived friendlier and more attentive, though some found it slower, they enjoyed deciding the game plan. Opinions on the shifting agent varied, with some enjoying collaborating with it and others feeling they had to fight for control.

AI partner perceptions influence trust levels and are mediated by feelings of control: If an agent insists on its own plan without establishing trust with the human, like the leader agent, it may seem unfriendly. In contrast, an agent that follows the human's plan may appear warmer and gain more trust. Nonetheless, it may seem slower as it attempts to accompany the human and understand its goal through one-way adaptation. Research shows that agents who adapt to human preference, while following the optimal plan, are perceived as more trustworthy, likeable and intelligent than those with fixed strategies [26]. However, the shifting agent was not perceived as expected. After the agent shifted to follower mode, some players might have taken too long to act, prompting the agent to revert to leader mode. Losing control of the game plan after having it could have made the agent appear inconsistent. Control is therefore a feature that influences trust perceptions towards agents. Prior research shows that high decision control increases trust in recommendation systems [43]. Humans may trust the follower agent more since they have complete control over the game plan. In addition, findings suggest that AI teammates are usually seen as having to be obedient, just like the follower agent, when compared to a human teammate [42]. Then, losing control with the shifting agent without understanding why may lead to more negative perceptions.

RQ2: How does an agent's initiative impact the perceived quality of collaboration?

Satisfaction with the follower agent's performance was higher compared to the other agents, as was the team performance satisfaction with this agent. Participants completed more levels with the leader agent, but its perceived quickness in playing the game gave the impression it made mistakes and lacked efficiency, despite being effective. Other agents appeared to have better skills but were criticised for not stabilising properly.

Perceptions of collaboration quality related to optimal plan: The leader agent did not follow the optimal plan, having a predefined order for diamond collection designed not to align with the expected human's plan (i.e., the optimal plan). This likely contributed to a lower performance satisfaction, which contrasts with the team's objective performance. Moreover, it was highlighted by some participants that this agent rushed so it made mistakes, being described as effective, but not efficient.

Implicit communication requires time: The perceived performance of the following agent was higher, likely due to its understanding of the human's presence and plan, even though its objective performance was lower. The lack of an explicit human-AI communication mechanism required humans to learn how to communicate with the agent, potentially resulting in more time spent in levels and a perception of this agent as slower, as seen before.

RQ3: How does an agent's initiative impact the humans' self-perception?

Participants were more satisfied with their performance when playing with the follower agent compared to other agents, despite being more idle with it. Satisfaction with the leader's performance was significantly lower than self-performance satisfaction, and participants felt they played an auxiliary role with this agent.

Self perceptions related to sense of responsibility in task: Research indicates that when an agent plays a leadership role, people feel less responsible for task performance [23]. With the leader agent handling all game planning, participants felt they had minimal influence, resulting in a reduced sense of accountability for the outcome, and possibly making them feel like they performed better than the agent. Participants' idle time may be related to the role humans perceive themselves in collaboration. Working more as a helper may lead participants to move more to accompany the leader agent. Conversely, when interacting with the follower agent, communicating involves patience (e.g., staying in place to indicate the target goal to the agent). Feeling like an helper and making no decisions with the leader agent is closely related to the previously mentioned lack of control, which may create distrust towards the agent and a sensation that the agent does not trust human abilities.

RQ4: How does agent's initiative affect the overall team perception?

Order effects were present in agent preference, yet participants identified differences between the agents when justifying their choice, especially between the leader and the follower. 18 participants preferred the leader agent for being proactive and faster, despite being less friendly. In contrast, 25 preferred the follower agent for its warmth, despite being perceived as less efficient. 17 participants favoured the shifting agent, but tended to play in follower mode more frequently. Only a few participants recognised its initiative balance.

Initiative preference related to context: Previous preferences were mediated by perceptions of the agent and collaboration. Research suggests that individuals prefer agents that consider their preferences when planning tasks [17], so one would expect that the leader agent would not be chosen as often. Interestingly, people develop leading and following behaviours differently over time [38]. Thus, we assume that this preference depends on individual characteristics, the framing of the situation (context), and the interaction between the two (what the framing entails for the human). For example, in a gaming context, some prioritise "achievement", "fast-paced", "winning", while others value "socializing", "collaboration", "fun". Participants inclined towards the former might not mind playing a support role if they win, whereas those in the latter group may prefer an active role in deciding the game plan. This might explain why many participants preferring the leader agent enjoy competitive games. Possibly, by highlighting collaboration and team decision-making, the leader would be chosen less often. Alternatively, if prizes were awarded based on game scores, the follower might be selected less.

Initiative preference related to personal characteristics: Female participants chose the leader agent less frequently than male participants. Female preference might lean towards patient, supportive agents, whereas some male participants prefer proactive goal-oriented agents. This aligns with gender differences found in leadership styles, as explored in prior research [6]. We found that personal goals can influence this trend, so these groups are not entirely homogeneous. This demonstrates how gamified approaches can provide insights into demographic influences on human-AI teams. Although we did not find significant results with video games experience, exploring whether experts prefer fast-paced goal-oriented games, leading to preferring leader agents, could be interesting. Moreover, the study included many international students, being difficult to draw conclusions, but controlling cultural variables in follow-up studies could reveal their impact on preference.

6.2 Design Factors

We propose a vocabulary of essential factors to be considered in the design of human-AI collaborations, in particular when modelling different levels of initiative in agents. Central to this vocabulary are considerations of the context of the situation and individual human differences, the importance of effective human-AI communication, and the necessity of developing a trustworthy relation between humans and agents through positive perceptions of the AI partner and the team.

Interaction between individual differences and context Individual differences may play a crucial role in AI partner preference. However, the context of the situation and how people perceive it should also be a factor to consider. For instance, fast-paced contexts with defined goals could benefit from agents with more initiative and taking the lead, while slower-paced scenarios could benefit from agents that adapt to human initiative. Moreover, exploring human preferences for task-oriented or people-oriented leadership styles, or a combination of both, could further enhance understanding on this factor.

Trust Establishing trust between the human and the agent in advance is important, specifically for faster-paced scenarios where the agent needs to take the lead. Trust building may rely on the agent showing that it is aware of the human's capabilities and willing to "listen" to them (e.g., their game plan). This becomes relevant in cases where the agent needs to play a leadership role, such as in emergency scenarios where actions need to be quick and faultless. We propose exploring whether prior interactions with the agent could be a strategy to build trust before the agent plays a leader role (with caution against creating "overtrust" [38]). Humans will be more familiarised with the agent's skills and gain more confidence in the human-AI team.

Communication When control is given to the human in a human-AI team, taking this control without explicitly communicating why can lead to confusion and reduced trust. Key aspects to consider include assessing the necessity of shifting between leader and follower roles and the impact of informing humans in advance of which situations require the agent to take full control. Additionally, while implicit communication mechanisms may contribute to understanding the agents and allow a more fluid interaction, they may not be sufficient in environments with adaptive initiative levels. While research shows that explicit communication or complete behaviour descriptions helps achieve better team performance and understanding [9, 34], games like *Geometry Friends* may be ideal to explore various types of communication in human-AI teams and the contexts in which they are necessary.

6.3 Limitations and Future Work

One limitation of the study is reduced ecological validity of using a game as a study tool. While it allows us to isolate and study particular effects, it would be beneficial to confirm our insights in a more realistic context. Further studies are required to generalise our findings, especially considering the goal-oriented nature of the gaming environment. Moreover, the role differences between the rectangle and circle, as the rectangle is unable to jump so it seems like it plays a supportive role, could impact participants' perceptions. We see an opportunity to switch these characters in future research to investigate the impact of agent representation on perception of initiative. Participants' interactions with all three agents in one session resulted in order effects in the results. While longer interaction times, exceeding 5 minutes, could capture more detail, they would have reduced the number of participants in the study limiting the insights we presented.

Based on our presented work, we see that *Geometry Friends* has potential for additional studies on human-AI collaboration. Exploring implicit vs. explicit communication mechanisms (in particular when the agent shifts initiative), investigating the effect of cultural differences in AI partner preferences, and re-framing the game to be more focused on task vs. collaboration, may be interesting avenues for future research.

7 CONCLUSION

This paper aims to explore initiative in human-AI collaboration by gathering insights on human perceptions of AI partners, the quality of collaboration, the self, and the overall team preference to understand when one should follow and when one should lead.

Our paper presents three main contributions, as it 1) provides detailed insights into perceptions of AI partners, collaboration quality, self, and team preferences given different levels of agent initiative, 2) proposes a vocabulary of factors from our findings relevant for the design of human-AI collaborations, highlighting the importance of the interplay between context and individual differences, trust and effective communication in human-AI collaboration, and 3) introduces a controlled game environment as an effective platform to conduct human-AI collaboration studies.

We used a collaborative game, *Geometry Friends*, to conduct a within-subjects study with 60 participants from Portugal, Sweden, and France. Participants played with three agents of varying levels of initiative: leader, follower, and shifting. Their interaction was assessed by evaluating human perceptions of warmth, competence, social identification with the team, satisfaction with performance, as well as preferences.

We found that AI partner perceptions are mediated by feelings of control in the game, and may influence trust levels in human-AI collaboration. Agents that follow human initiative make humans feel more in control, being seen as more friendly, and, possibly, more trustworthy. Initiative shifts of the agent that result in the human losing control in collaboration without understanding why, may lead to negative agent perceptions. Concerning collaboration quality, if an agent does not follow the optimal plan, even if it is objectively faster in reaching the goals, its performance can be perceived as lower. Moreover, implicit communication, in particular for the agent following human initiative, takes time, as it has to understand the human's goals, reducing the objective quality of collaboration. Self-perceptions are related to participants perceiving themselves with an auxiliary role when collaborating with an agent with more initiative, feeling less responsible for overall task performance. Finally, AI Partner preference in the team is influenced by the interplay between context (e.g., a game) and its meaning to the human (e.g., a competitive player may prefer an agent with more initiative, as it sees game contexts as "fast-paced").

Our work provided valuable insights on how to design agents with different levels of initiative, specially the implications and mediating aspects of such a design. As AI systems continue to advance, the question of collaboration between humans and AI becomes increasingly relevant. This study is an important step towards a better understanding of initiative in collaborative systems and provides valuable suggestions on "when to lead or follow".

ACKNOWLEDGMENTS

This work was funded by the European Union Horizon 2020 Research and Innovation Programme as part of the Humane-AI-net project, under grant agreement number 952026. It was supported by national funds through Fundação para a Ciência e a Tecnologia (FCT) with reference UIDB/50021/2020. Inês Lobo acknowledges her FCT grant (2021.04796.BD).

REFERENCES

[1] Kathleen Allen, Richard Bergin, and Kenneth Pickar. 2004. Exploring trust, group satisfaction, and performance in geographically dispersed and co-located university technology commercialization teams. In *VentureWell. Proceedings of Open, the Annual Conference.* National Collegiate Inventors & Innovators Alliance, 201.

[2] Saleema Amershi, Dan Weld, Mihaela Vorvoreanu, Adam Fourney, Besmira Nushi, Penny Collisson, Jina Suh, Shamsi Iqbal, Paul N Bennett, Kori Inkpen, et al. 2019. Guidelines for human-AI interaction. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems.* 1–13.

[3] Zahra Ashktorab, Casey Dugan, James Johnson, Qian Pan, Wei Zhang, Sadhana Kumaravel, and Murray Campbell. 2021. Effects of communication directionality and AI agent differences in human-AI interaction. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems.* 1–15.

[4] Zahra Ashktorab, Q. Vera Liao, Casey Dugan, James Johnson, Qian Pan, Wei Zhang, Sadhana Kumaravel, and Murray Campbell. 2020. Human-AI Collaboration in a Cooperative Game Setting: Measuring Social Perception and Outcomes. *Proc. ACM Hum.-Comput. Interact.* 4, CSCW2, Article 96 (oct 2020), 20 pages. https://doi.org/10.1145/3415167

[5] Michael J Baker. 2015. Collaboration in collaborative learning. *Interaction Studies* 16, 3 (2015), 451–473.

[6] Victoria Bellou. 2011. Do women followers prefer a different leadership style than men? *The International Journal of Human Resource Management* 22, 13 (2011), 2818–2833.

[7] Kirsten Bergmann, Friederike Eyssel, and Stefan Kopp. 2012. A second chance to make a first impression? How appearance and nonverbal behavior affect perceived warmth and competence of virtual agents over time. In *Intelligent Virtual Agents: 12th International Conference, IVA 2012, Santa Cruz, CA, USA, September, 12-14, 2012. Proceedings 12.* Springer, 126–138.

[8] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health* 11, 4 (2019), 589–597.

[9] Ángel Alexander Cabrera, Adam Perer, and Jason I Hong. 2023. Improving human-AI collaboration with descriptions of AI behavior. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1 (2023), 1–21.

[10] Micah Carroll, Rohin Shah, Mark K Ho, Tom Griffiths, Sanjit Seshia, Pieter Abbeel, and Anca Dragan. 2019. On the utility of learning about humans for human-ai coordination. *Advances in neural information processing systems* 32 (2019).

[11] Alejandro Catala, Fernando Garcia-Sanjuan, Patricia Pons, Javier Jaen, and Jose A Mocholi. 2012. Agoras: Towards Collaborative Game-Based Learning Experiences on Surfaces. *International Association for Development of the Information Society* (2012).

[12] Megan Charity, Isha Dave, Ahmed Khalifa, and Julian Togelius. 2022. Baba is Y'all 2.0: Design and Investigation of a Collaborative Mixed-Initiative System. *IEEE Transactions on Games* (2022).

[13] Lara Christoforakos, Alessio Gallucci, Tinatini Surmava-Große, Daniel Ullrich, and Sarah Diefenbach. 2021. Can robots earn our trust the same way humans do? A systematic exploration of competence, warmth, and anthropomorphism as determinants of trust development in HRI. *Frontiers in Robotics and AI* 8 (2021), 640444.

[14] Shi Feng and Jordan Boyd-Graber. 2019. What can ai do for me? evaluating machine learning interpretations in cooperative play. In *Proceedings of the 24th International Conference on Intelligent User Interfaces.* 229–239.

[15] Matthew C Fontaine, Ya-Chuan Hsu, Yulun Zhang, Bryon Tjanaka, and Stefanos Nikolaidis. 2021. On the importance of environments in human-robot coordination. *arXiv preprint arXiv:2106.10853* (2021).

[16] Katy Ilonka Gero, Zahra Ashktorab, Casey Dugan, Qian Pan, James Johnson, Werner Geyer, Maria Ruiz, Sarah Miller, David R Millen, Murray Campbell, et al. 2020. Mental models of AI agents in a cooperative game setting. In *Proceedings of the 2020 chi conference on human factors in computing systems.* 1–12.

[17] Matthew Gombolay, Anna Bair, Cindy Huang, and Julie Shah. 2017. Computational design of mixed-initiative human–robot teaming that considers human factors: situational awareness, workload, and workflow preferences. *The International journal of robotics research* 36, 5-7 (2017), 597–617.

[18] Fritz Heider and Marianne Simmel. 1944. An experimental study of apparent behavior. *The American journal of psychology* 57, 2 (1944), 243–259.

[19] Eric Horvitz. 1999. Uncertainty, action, and interaction: In pursuit of mixed-initiative computing. *IEEE Intelligent Systems* 14, 5 (1999), 17–20.

[20] Matthew Johnson and Alonso Vera. 2019. No AI is an island: the case for teaming intelligence. *AI magazine* 40, 1 (2019), 16–28.

[21] Hyun-Tae Kim, Du-Mim Yoon, and Kyung-Joong Kim. 2014. Solving Geometry Friends using Monte-Carlo tree search with directed graph representation. In *2014 IEEE Conference on Computational Intelligence and Games.* IEEE, 1–2.

[22] Florian Lehmann. 2023. Mixed-Initiative Interaction with Computational Generative Systems. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems.* 1–6.

[23] Xin Lei and Pei-Luen Patrick Rau. 2021. Effect of relative status on responsibility attributions in human–robot collaboration: Mediating role of sense of responsibility and moderating role of power distance orientation. *Computers in Human Behavior* 122 (2021), 106820.

[24] Pedro Faria Lopes, David Jardim, and Isabel Machado Alexandre. 2011. Math4Kids. In *6th Iberian Conference on Information Systems and Technologies (CISTI 2011).* IEEE, 1–6.

[25] Geoff Musick, Thomas A O'Neill, Beau G Schelble, Nathan J McNeese, and Jonn B Henke. 2021. What happens when humans believe their teammate is an AI? An investigation into humans teaming with autonomy. *Computers in Human Behavior* 122 (2021), 106852.

[26] Stefanos Nikolaidis, David Hsu, and Siddhartha Srinivasa. 2017. Human-robot mutual adaptation in collaborative tasks: Models and experiments. *The International Journal of Robotics Research* 36, 5-7 (2017), 618–634.

[27] Rui Prada, Phil Lopes, João Catarino, Joao Quitério, and Francisco S Melo. 2015. The geometry friends game AI competition. In *2015 IEEE conference on computational intelligence and games (CIG).* IEEE, 431–438.

[28] João Quitério, Rui Prada, and Francisco S Melo. 2015. A reinforcement learning approach for the circle agent of geometry friends. In *2015 IEEE Conference on Computational Intelligence and Games (CIG).* IEEE, 423–430.

[29] José Bernardo Guimarães Rocha. 2009. *Geometry friends.* Ph. D. Dissertation. Master's thesis, IST, University of Lisbon.

[30] Jeremy Roschelle, Stephanie D Teasley, et al. 1995. The construction of shared knowledge in collaborative problem solving. In *Computer-supported collaborative learning*, Vol. 128. 69–197.

[31] Ana Salta, Rui Prada, and Francisco Melo. 2019. Solving motion and action planning for a cooperative agent problem using geometry friends. In *EPIA Conference on Artificial Intelligence.* Springer, 86–97.

[32] Ana Salta, Rui Prada, and Francisco S Melo. 2020. A Game AI Competition to foster Collaborative AI research and development. *IEEE Transactions on Games* 13, 4 (2020), 398–409.

[33] Vida Scarpello and John P Campbell. 1983. Job satisfaction: Are all the parts there? *Personnel psychology* 36, 3 (1983), 577–600.
[34] Beau G Schelble, Christopher Flathmann, Nathan J McNeese, Guo Freeman, and Rohit Mallick. 2022. Let's think together! Assessing shared mental models, performance, and trust in human-agent teams. *Proceedings of the ACM on Human-Computer Interaction* 6, GROUP (2022), 1–29.
[35] Astrid Schepman and Paul Rodway. 2020. Initial validation of the general attitudes towards Artificial Intelligence Scale. *Computers in human behavior reports* 1 (2020), 100014.
[36] Rui Soares, Francisco Leal, Rui Prada, and Francisco S Melo. 2016. Rapidly-Exploring Random Tree approach for Geometry Friends.. In *DiGRA/FDG*.
[37] Konrad Sowa, Aleksandra Przegalinska, and Leon Ciechanowski. 2021. Cobots in knowledge work: Human–AI collaboration in managerial professions. *Journal of Business Research* 125 (2021), 135–142.
[38] Emma M Van Zoelen, Emilia I Barakova, and Matthias Rauterberg. 2020. Adaptive leader-follower behavior in human-robot collaboration. In *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 1259–1265.
[39] Dakuo Wang, Elizabeth Churchill, Pattie Maes, Xiangmin Fan, Ben Shneiderman, Yuanchun Shi, and Qianying Wang. 2020. From human-human collaboration to Human-AI collaboration: Designing AI systems that can work together with people. In *Extended abstracts of the 2020 CHI conference on human factors in computing systems*. 1–6.
[40] Isaac Wang and Jaime Ruiz. 2021. Examining the use of nonverbal communication in virtual agents. *International Journal of Human–Computer Interaction* 37, 17 (2021), 1648–1673.
[41] John P Wanous, Arnon E Reichers, and Michael J Hudy. 1997. Overall job satisfaction: how good are single-item measures? *Journal of applied Psychology* 82, 2 (1997), 247.
[42] Rina R Wehbe, Edward Lank, and Lennart E Nacke. 2017. Left them 4 dead: Perception of humans versus non-player character teammates in cooperative gameplay. In *Proceedings of the 2017 conference on designing interactive systems*. 403–415.
[43] Monika Westphal, Michael Vössing, Gerhard Satzger, Galit B Yom-Tov, and Anat Rafaeli. 2023. Decision control and explanations in human-AI collaboration: Improving user perceptions and compliance. *Computers in Human Behavior* 144 (2023), 107714.
[44] H James Wilson and Paul R Daugherty. 2018. Collaborative intelligence: Humans and AI are joining forces. *Harvard Business Review* 96, 4 (2018), 114–123.
[45] Rui Zhang, Nathan J McNeese, Guo Freeman, and Geoff Musick. 2021. " An Ideal Human" Expectations of AI Teammates in Human-AI Teaming. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–25.

A DEMOGRAPHIC DETAILS OF PARTICIPANT

Detailed participant demographics by country.

ID	Age	Gender	Highest Degree	Video Games Frequency	Competitive Games Preference	Cooperative Games Preference	Positive AI Perceptions	Negative AI Perceptions
1	18–24	M	Masters	Often	Like	Like	4,00	2,25
2	18–24	F	Bachelors	Always	Strongly Like	Neutral	4,58	3,33
3	25–34	M	Doctorate	Never	Like	Strongly Like	4,42	2,13
4	25–34	M	Bachelors	Always	Like	Strongly Like	4,58	2,88
5	25–34	F	Masters	Rarely	Dislike	Neutral	4,00	4,13
6	25–34	M	Bachelors	Sometimes	Like	Strongly Like	4,25	3,67
7	25–34	F	Masters	Rarely	Neutral	Strongly Like	4,75	2,88
8	18–24	F	Bachelors	Sometimes	Neutral	Neutral	4,75	1,38
9	25–34	F	Masters	Rarely	Dislike	Neutral	3,50	3,13
10	25–34	M	High School	Always	Strongly Like	Strongly Like	3,67	3,75
11	25–34	M	Masters	Rarely	Neutral	Neutral	3,83	2,63
12	18–24	M	Bachelors	Always	Like	Neutral	2,75	3,50
13	18–24	F	Bachelors	Rarely	Dislike	Like	4,08	2,50
14	18–24	M	High School	Always	Like	Neutral	3,75	3,00
15	25–34	F	Masters	Sometimes	Like	Like	4,67	1,33
16	18–24	M	Bachelors	Often	Strongly Like	Strongly Like	4,17	2,50
17	25–34	M	Masters	Always	Like	Like	4,08	2,00
18	25–34	M	Masters	Often	Like	Strongly Like	4,17	3,88
19	18–24	F	Masters	Rarely	Neutral	Like	3,92	2,63
20	25–34	F	Masters	Rarely	Neutral	Neutral	3,25	2,88

Table 1: Demographics of participants in Portugal (denoted as P)

ID	Age	Gender	Highest Degree	Video Games Frequency	Competitive Games Preference	Cooperative Games Preference	Positive AI Perceptions	Negative AI Perceptions
1	18–24	M	Bachelors	Often	Strongly Like	Like	4,25	2,20
2	25–34	M	Doctorate	Always	Strongly Like	Neutral	3,83	2,00
3	25–34	M	Masters	Sometimes	Like	Like	3,92	2,75
4	18–24	M	Bachelors	Often	Like	Neutral	3,42	3,38
5	25–34	M	Masters	Often	Dislike	Strongly Like	3,08	1,88
6	18–24	M	Masters	Often	Strongly Like	Strongly Like	3,67	3,13
7	18–24	M	High School	Often	Like	Strongly Like	4,17	2,63
8	18–24	M	High School	Often	Like	Strongly Like	4,17	2,00
9	56–70	M	Bachelors	Never	Does Not Apply	Does Not Apply	3,42	3,13
10	25–34	M	Doctorate	Always	Like	Like	3,83	2,75
11	18–24	M	Bachelors	Sometimes	Like	Like	3,33	3,5
12	35–45	M	Masters	Sometimes	Strongly Like	Like	3,58	3,00
13	18–24	M	High School	Often	Strongly Like	Like	4,00	2,13
14	25–34	F	Doctorate	Often	Neutral	Dislike	2,55	3,50
15	35–45	F	Doctorate	Never	Does Not Apply	Does Not Apply	3,25	2,67
16	25–34	M	Doctorate	Sometimes	Strongly Like	Like	3,75	2,50
17	25–34	M	Masters	Often	Strongly Like	Strongly Like	4,42	2,63
18	25–34	F	Masters	Sometimes	Strongly Dislike	Strongly Like	4,42	1,38
19	25–34	M	Doctorate	Always	Dislike	Like	3,75	4,00
20	25–34	M	Masters	Always	Strongly Like	Strongly Like	4,67	3,63

Table 2: Demographics of participants in Sweden (denoted as S)

ID	Age	Gender	Highest Degree	Video Games Frequency	Competitive Games Preference	Cooperative Games Preference	Positive AI Perceptions	Negative AI Perceptions
1	18−24	F	Bachelors	Often	Like	Neutral	3,25	2,88
2	25−34	M	Doctorate	Often	Like	Strongly Like	3,50	3,38
3	25−34	M	Masters	Always	Like	Strongly Like	3,50	4,13
4	25−34	M	Bachelors	Rarely	Neutral	Like	3,75	2,88
5	35−45	M	Doctorate	Always	Like	Strongly Like	3,83	3,13
6	25−34	M	Doctorate	Rarely	Like	Neutral	2,50	4,13
7	18−24	F	Bachelors	Rarely	Does Not Apply	Neutral	3,25	2,63
8	25−34	M	Masters	Sometimes	Like	Neutral	3,83	3,88
9	25−34	F	Doctorate	Rarely	Like	Like	3,50	3,00
10	25−34	F	Masters	Sometimes	Strongly Like	Neutral	3,00	2,75
11	18−24	M	Masters	Always	Like	Like	3,56	3,67
12	25−34	F	Masters	Never	Does Not Apply	Does Not Apply	4,33	2,00
13	35−45	M	Doctorate	Never	Does Not Apply	Does Not Apply	3,08	3,38
14	25−34	F	Masters	Sometimes	Strongly Like	Strongly Like	3,58	3,63
15	25−34	M	Masters	Rarely	Strongly Dislike	Neutral	4,33	2,88
16	45−55	M	Masters	Often	Neutral	Strongly Like	3,25	2,88
17	25−34	M	Doctorate	Often	Neutral	Like	2,58	2,75
18	35−45	F	Masters	Never	Does Not Apply	Does Not Apply	3,70	2,75
19	18−24	F	Masters	Sometimes	Dislike	Strongly Like	3,20	3,25
20	25−34	M	Masters	Rarely	Dislike	Dislike	3,83	3,00

Table 3: Demographics of participants in France (denoted as F)