Investigating the Impact of a Ranking & Reward System in Promoting Prosocial Behavior in Online Multiplayer Games

Panagiotis Vrettis
p.vrettis@students.uu.nl
Utrecht University
Utrecht, The Netherlands

Abstract

The video game industry has consistently grown to a giant entertainment medium, rivaling the movie, literature, and music industries. In online multiplayer video games, disruptive behavior like toxicity and other harmful conduct have plagued player experience since the introduction of such games. Unfortunately, the few proposed solutions have failed to mitigate, let alone eliminate, the problem. In this master's thesis, a novel approach is proposed to alleviate the issue. A custom system is designed and developed that tracks the player's in-game prosocial behavior and ranks them based on points they gain through their prosocial actions. A user study with 16 participants was carried out to evaluate the effects of this system, where participants play a game of their choice out of three available options: Helldivers II, Overwatch 2, and Rainbow Six: Siege, over two gameplay sessions. Both sessions are recorded and analyzed to extract their prosocial actions. The data from the first session is imported into the system and given to participants during the second session, providing a detailed breakdown of the actions performed, a description of each available action per game, and the total social points of the participant. These components are leveraged to help participants self-reflect on their prosocial performance from the first session, and support them in discovering what other prosocial actions are available in the game of their choice, as well as recognize where they can improve to be more prosocial. The data from the second gameplay session is used for comparison, to investigate whether an observable increase in prosocial behavior occurred. Results indicate that, compared to the first session, there was a significant increase in prosocial behavior among all participants and all three games, as well as a notable rise in the variety of prosocial actions that were performed. The proposed system was also rated very highly for its usability (SUS grade of *A*). Overall, this work strongly supports the potential impact of such a system and its benefits in the context of online multiplayer video game behavior.

CCS Concepts

• **Human-centered computing** → **Collaborative and social computing systems and tools**; **Empirical studies in collaborative and social computing**; *Human computer interaction (HCI)*; *Visualization systems and tools.*

Keywords

Toxicity, Online Games, Prosocial, Proactive Design, Self-Reflection, Self-Regulation

1 Introduction

Multiplayer online video games have been prevalent in the games industry since their invention at the end of the twentieth century. Since then, player counts have increased exponentially, and so have the multiplayer experiences available. Along with this voluminous flood of types of available content, players of every stripe emerged, filling the available servers and building communities. In this way, many games grew outside their original scope and scale, often leading to developers struggling to keep up and to retain control. Sadly, the already widespread on the internet toxicity quickly infiltrated multiplayer games and their communities. This sparked the ongoing fight of game developers (and players) against toxic actors, and plenty of research along with it.

Existing academic literature and game-specific approaches have predominantly attempted to combat this issue by targeting the toxic behavior and its perpetrator. This usually entails a series of warnings against exhibiting what constitutes toxic behavior, providing players with a "report functionality", and the eventual punishment of toxic individuals. This strategy, however, is very limited in its scope. For instance, toxic behavior is a term that is difficult to concretely define [32], which leads to considerable inconsistencies between games and their interpretations of actions of the type inside the game's context [77]. Using terms like "disruptive behaviors" [43] instead of "toxicity" can help to some degree, but this adaptation is slow and inadequate on its own. Furthermore, every game implements its own unique reporting and punishment system. Thus, players have to familiarize themselves with different designs, often with no assistance or guidance by the game itself. Those facts, along with poor design decisions, allow for perpetrators to escape punishment or receive one that is disproportional to the crime, whether harsher or more lenient. Worse still, the reporting system can itself become an instrument in the perpetrators' toolbox. Most such systems can be readily abused by players, who report an innocent player until they are punished by the system due to its simple automated design nature. Finally, all the existing systems follow a reactive design, meaning that someone is going to be targeted by and experience toxic behavior, before the toxic actor is reported and -potentially- punished.

This master's thesis differentiates itself by approaching the problem from a different angle. Namely, why ask people to not be toxic (lest there be consequences), instead of promoting and rewarding prosocial behavior in the game? Multiplayer games should not connote toxicity. Rather, they should signify community and prosocial interaction. The goal was to design and develop a system that is characterized by modularity, allowing it to be used in most (if not all) multiplayer team player-versus-player (PvP) and player-versus-environment (PvE) games. The system is a prototype with

mostly actually implemented functionality, but with some simulated actions that would in a potential final version be completely automated. This detail, however, does not impact the outcome of this thesis. The novel system is based on a points-ranking scheme, where points are added to assign ranks to players based on their in-game prosocial behavior. Players would receive specific in-game rewards exclusive to this system (simulated as part of this study). This basis of design enables it to both appeal to players, as it satisfies both "Social" and "Achievement" in the Gamer Motivation Model [80], as well as be familiar to them, since it mimics a classic ranking system that they are likely to be acquainted with. The system provides a detailed analysis of all prosocial actions available in each game, along with statistics on the ones that the player has performed. The goal of these two characteristics is to help players self-reflect on their prosocial performance, as well as discover new ways to be prosocial in their favorite games.

To evaluate the effectiveness of the aforementioned system, a user study with 16 participants gathered through convenience sampling was conducted. Participants played one of three available games that have been chosen based on a number of factors: Helldivers II, Overwatch 2, and Rainbow Six: Siege. Their gameplay session was recorded, and the data related to the prosocial actions performed was extracted from it. In a second session, the data was fed into the system, which in turn was provided to the participants to explore, before they were asked to play the same game again. The recorded second session was used to investigate whether the self-reflection provided by the system and its various components had an effect on in-game prosocial behavior between the two gameplay sessions. More specifically, the following research questions are addressed:

- RQ1: How does the system designed affect the player's self-reflection when it comes to their in-game prosocial behavior?
- RQ2: How does the system designed affect the amount of prosocial behavior they exhibit in-game?

The system was also evaluated on its design and ease-of-use by the participants, to examine whether usability was -or was not- a factor in the overall assessment of the system.

2 Related Work

Toxic behavior is present and has been studied in multiple aspects of modern everyday life. In organizational contexts, leadership toxicity fosters divisive cultures, affects people's self-esteem, threatens occupational security, and physically and mentally intimidates them [65]. In the last two decades, social media has progressively tightened its grip on modern culture and life. Despite its clear benefits, the spread of socially toxic material like toxic masculinity [64], disinformation, conspiracies, extremism, harassment (i.e., cyberbullying), and violence has been exacerbated by social media platforms [70]. Counteracting this effect has been predominantly unsuccessful, since all forms of moderation (e.g., automated, manual, and human-in-the-loop) have generally failed to prevent the publishing of such unwanted content [82]. The primary focus of research, however, is in the domain of video games, and more notably in online multiplayer games, where toxic behavior appears to be an unbridled issue, inducing player churn (i.e., attrition) and revenue loss [51]. Thus, the literature review that follows concentrates on online competitive multiplayer games. Similarly, previous research on prosocial behavior centers on the spectrum of video games.

2.1 Toxicity and Disruptive Behavior

Along with the evolution of all video games in terms of quality, genre variety, revenue, and player count, online multiplayer video games in particular have grown on their own expeditious pace, with players swarming the servers and forming social communities of their own. A major factor in this progress is the competitive aspect of multiplayer games. Player enjoyment and player motivation increases when real time competition is part of the design of a game [49, 81]. More specifically, players competing with other players of similar skill levels, tend to put more effort into the game (e.g., more games played and longer gameplay sessions). On the other hand, when players compete with other players of lower skill levels, they report higher levels of enjoyment and lower levels of arousal [58].

Despite its congenial effect, the element of competition nevertheless constitutes favorable conditions for negative behavior to take place. The most prominent negative phenomenon in video games (and online in general) is toxic behavior, which is a hypernym that encompasses a large variety of negative behaviors in the context of a game (or another online platform), such as harassment, offensive language and threats, exploiting (i.e., taking advantage of bugs and glitches), griefing, flaming, trolling, and cheating [18, 31]. Neto et al. found that there are different in-game text chat communication patterns that are directly related to player performance and level of toxic behavior [62], while Monge et al. showed that toxic behavior significantly worsened team and individual player performance [60]. A notable factor in toxic behavior in all online contexts is the toxic disinhibition component of the online disinhibition effect. This component outlines the perceived lack of restraint an individual feels when communicating online compared to communicating in-person due to decreased behavioral inhibitions [71], meaning that anonymity allows some actors to more easily engage in negative behavior. It is noteworthy that victims often hesitate to report such behavior, unless prompted to do so [54]. It is common for harmful conduct to be normalized in video games and to be seen as part of the experience rather than a problem that needs to be solved [25]. Worse still, players not only accept toxic behavior as par for the course, they also ignore positive behavior because they tend to consider it to be not genuine, and dismiss it as sarcasm [66].

Toxic behavior is also particularly hard to define because it is too broad a term, too subjective, and symptomatic, making it hard to punish actual harmful actors and pardon innocent victims of a reporting system [57]. Different games provide different opportunities for multiple types of toxicity, and investigating them helps understand the player's perspective of toxicity better [52]. This lack of a concrete definition and description has continued being an issue, regardless of attempts to empirically investigate the "why's" of toxic behavior [54]. Weszt H. recommends the introduction of the terms "disruptive behaviors" or "harmful conduct" (i.e., actions people do not want to see in digital spaces) [43] to combat this issue. Many types of behavior are wrapped by those two terms, like competitive or direct harm, disruptive communication or play, fraudulent activity, and inappropriate sharing. Weszt H. also suggests that developers take proactive steps to counter disruptive

behaviors, instead of the more common reactive ones. Some important suggestions are to "promote prosocial behavior through design, narrative, communication, and more", and "set clear expectations for what good behavior looks like", which this work uses to motivate its design.

2.2 Prosocial Behavior

Wittek et al. define prosocial behavior as any social behavior resulting in benefits for others, often involving costs for the self [78]. Such behavior does not have to be significantly personally expensive. In fact, prosocial acts in societies tend to decrease when these acts are perceived to be (or are) personally costly, while when such acts do not involve personal sacrifice, prosocial acts flourish [47]. Some examples are helping others, sharing goods, donating to charity, cooperating towards a common (or not so common) goal, and volunteering. Generally, when there are net costs for the prosocial actor, this is referred to as altruism [28]. Since human societies are strongly dependent on cooperation to thrive, it is clear that any individual act that helps this construct foster is imperative to the society's evolution. Prosocial behavior not only has a positive effect on cooperation, but it also stands as one of its cornerstones [46]. More importantly, because today's modern society is characterized by globalization, prosocial actions outside of just closed groups are imperative for the proper functioning of our diverse heterogeneous societies [23].

A great benefit of giving to others, is that it is emotionally rewarding. A lot of research indicates that happiness is closely tied with prosocial acts. More specifically, both adults and children report a higher amount of positive emotion after exhibiting prosocial acts compared to similar groups that engaged in self-centered acts instead [19]. According to Aknin et al., acting in a prosocial manner is more likely to increase individual well-being when the actor feels that they have the freedom of choosing whether to act this way, they feel some form of social connection to the person receiving the "help", and they can witness how their act makes a difference [21]. Furthermore, Erreygers et al. focus on an online context, where happiness leads adolescents and parents to take their positive emotional states from school and work, respectively, and act on their happiness by exhibiting prosocial behavior online [38].

2.3 Rewarding Prosocial Behavior

Given the societal and personal significance of positive behavior, it is worthwhile to investigate the effects of rewarding such behavior. For the task of animal training, it has been found that rewarding the positive behavior of the animal benefits the process by further reinforcing the wanted behavior [72]. In child development, to achieve the optimal, many factors like positive reinforcement come into play [59]. In this context, rewarding the desired positive behavior is more effective than punishing unwanted negative behavior [24]. Similarly, rewards have a positive effect on general behavior change and decision-making [40], as well as promoting hygienic habits in areas where sanitary conditions are crucial (i.e., health care workers) [22]. On social media platforms, Choi et al. found that using the "creator heart" feature of YouTube (i.e., the video creator has liked a comment on that video) boosted positive engagement received by content creators and increased the visibility of comments, while also encouraging viewers to comment further [34]. Furthermore, Lambert et al. investigated how Reddit users are affected by positive feedback they receive, and found that those who received positive feedback (i.e., received gold on their post and were highly upvoted by others) tend to make more frequent posts per day and of higher quality [55].

Analogous results were found in rewarding prosocial behavior, specifically. It is worth noting that positive feelings themselves can constitute a form of reward for prosocial behavior. Aknin et al. reported that positive feelings not only reinforce prosocial behavior, they also serve as a good predictor of it. At the same time, prosocial behavior (like other types of positive behavior) lead to positive feelings. However, only tentative evidence with adults supports the existence of a "positive feedback loop" [20] between behavior and feelings. Wu et al. argue that rewards are better at promoting prosocial behavior than punishments as, while both are costly, reward is cheaper when such behavior is less uncommon [79].

One unique quality of prosocial behavior is how advertising it (e.g., bragging about it) influences perceived generosity. Interestingly, advertising signals a selfish motivation that erodes the attribution of generosity when the behavior is known, while it has a positive effect when it is unknown [26]. Moreover, the nature of reward as an extrinsic motivator can negatively impact follow-up prosocial behavior. Self-determination theory (SDT) [36] makes distinctions between the types of motivation and the outcomes of each type. In general, it is preferable to have people exhibit positive behaviors because they are self-determined to do so (i.e., intrinsically motivated). To achieve that, designers should aim for an internal type of extrinsic motivation (in terms of perceived causality), to be as close to intrinsic as possible [37]. Thus, we have to be cautious when we are designing the potential rewards and when those are given [48].

Video game companies have previously attempted to counteract toxicity with strategies that reward prosocial acts and the lack of toxic behavior. Riot Games developed a tool called "Honor System" to encourage positive social behavior, which allows players to theoretically acknowledge positive teammates using a voting system, and earn rewards through it [33]. Blizzard Entertainment developed a similar system for Overwatch and Overwatch 2, named "Endorsement System". This is particularly effective as a form of discipline because it includes players as part of the process, by influencing players to fit a pre-defined behavioral mold (or norm). The "Endorsement System" fits Foucault's discipline theory [73], which indicates that disciplinary power involves surveillance and control. Unfortunately, such systems have failed to achieve their design goal. Riot Games themselves have reported that players tend to vote for teammates based on game performance rather than social behavior. The quote is, "We know players sometimes use honor to reward good play rather than good behavior. We were hoping that with a clearer link between the system and the reward, players would be encouraged to use it as a way to show their appreciation of good behavior" [53]. According to Blizzard Entertainment, their "Endorsement System" has decreased toxic behavior successfully. 50% to 70% of players have given endorsements, 26.4% fewer matches in the Americas (16.4% in Korea) contained abusive chat, while the daily abusive players decreased by 28.8% in the Americas (21.6% in Korea) [73]. However, some unintended effects have

been observed. For example, friends cannot endorse each other, meaning that because players tend to endorse teammates (rather than the opposing team), playing with friends in your team can lead to reputation decay. Reports (regardless of the outcome of the subsequent investigation) negatively impact the ranking, and thus the system can be exploited with fraudulent reports. Overwatch used to ban players automatically after enough reports, without any investigation taking place. Pretending to be nice can also be considered a problem, as it may mean that positive behavior is only exhibited under specific conditions (e.g., under surveillance), but it's better than toxicity, and it may lead to a habit and result in long term positive change. Finally, players are more likely to endorse users of specific in-game roles, and avoid endorsing others because of the roles not being played as expected. It is generally unknown why methods of this type of intervention (promoting positive social behavior) fail in the context of multiplayer games.

2.4 Self-Reflection & Self-Regulation

Two final components worth considering are self-reflection and self-regulation. According to Boud et al., self-reflection can be defined as "those intellectual and affective activities that individuals engage into to explore their experience, which leads to new understanding and appreciations" [27], while Moon J. A. describes it as "a form of mental processing with a purpose and/or anticipated outcome that is applied to relatively complex or unstructured ideas for which there is not an obvious solution" [61]. On the other hand, self-regulation in the context of learning refers to one's ability to understand and control one's learning environment [63], and is key to achieving a solid understanding of development and psychopathology [67]. If we want people to learn and adopt a new behavior, self-regulation and self-reflection can assist with that task. Previous research has strongly suggested that self-reflection had a positive effect on academic performance [56]. Modern self-regulation approaches incorporate aspects of both metacognition and self-regulation, focusing on self-monitoring [63]. Kleinman et al. demonstrated that how self-reflection and self-regulation occur can significantly affect performance improvement of esports players in online multiplayer video games [50].

2.5 Behavior Ranking Systems in Video Games

Over the last decade and a half, video game studios developing online multiplayer titles have started adding ranking and reward systems to their games, that focus on in-game behavior. Because these systems have been given a multitude of titles, this thesis takes the liberty to baptize them Behavior Ranking Systems, for the purposes of clarity. The principal reason for this development is the rising levels of disruptive behavior in their games, as well as the effects such a behavior can have on the individual and on society in general. The goals of each system -as described by the developers themselves- are often quite different. They all do, however, seem to aim to reduce disruptive behavior, lightly encourage team play and positive behavior, and in general increase player experience and engagement. The idea stems from competitive ranking systems, where players are ranked based on their in-game performance, like wins, losses, etc. The players also earn rewards based on their ranking, often on top of the usual rewards the game provides them with from regular gameplay. These systems are indeed quite successful at attracting players, to a point where competitive (or ranked) modes are replacing "casual" modes as the playing norm among the majority of players.

Their success is likely owed to their dependence on the feeling of competition they provide, along with the bragging rights that come from the resulting rank. This competitive aspect fits perfectly under the "Social" category of the Gamer Motivation Model, along with its heavy reliance on effective team play and communication, which falls under the same category [80]. The levels and ranks earned through competitive play motivate the player to continue playing and improving, as described by the "Achievement" category in the Gamer Motivation Model. Furthermore, the rewards that are made available exclusively through competitive modes constitute an additional incentive to partake in ranked games. The latter also falls under the "Achievement" category of the Gamer Motivation Model.

To better understand these systems, a detailed description is provided in the subsections that follow. In general, the core similarity between the systems is their reliance on a player-voting approach. Namely, points are awarded to players based on whether they voted for someone else after a match, as well as whether they were voted on by other players. This characteristic seems to be a limiting factor of the systems, as it is discussed in the paragraphs that follow.

2.5.1 League of Legends - Honor System. Riot Games' League of Legends was the first online multiplayer game to establish an in-game behavior ranking system. Dubbed the Honor System, it was introduced on October 1, 2012 to help incentivize positive behavior among the League of Legends community, primarily by identifying and rewarding players who positively influence the in-game experience of others during gameplay. The idea is that the extra impetus provided by this system would be enough to decrease disruptive behavior in the game and lead to a positive player experience. Given League of Legends' long history (released in 2009), the game has gone through multiple versions over the years, and along with it, the Honor System has undergone multiple iterations. These major versions are worth revisiting to better understand the evolution of the system, as well as to identify the reasons that led to the relevant changes.

In its original rendition, the Honor System used to be a four-tier system that lacked actual rewards altogether [13]. Essentially, players were ranked on four different categories (Figure 1), separately:

- **Friendly**: for players that have a positive impact on the game and make the match enjoyable, win or lose
- **Helpful**: for players that share their know-how and actively help other players improve their gameplay
- **Honorable Opponent**: for players from the opposing team that remain humble in victory or graceful in defeat and/or behave in a positive manner throughout the game
- **Teamwork**: for players that put the team ahead of themselves. This includes anything from forming great plans, helping struggling lanes recuperate, and more

Players can vote for others to increase their honor. This would occur on the result screen at the end of each player-versus-player match. Each match provided players with a limited number of votes to distribute, so that spamming/farming of said votes did

not occur. All votes were anonymous, with the receiving player not been informed who honored them. The "Friendly", "Helpful", and "Teamwork" categories were available for teammates, while the "Honorable Opponent" category was exclusively available for players of the opposing team. Players can also vote for friends or regular teammates, but this will provide fewer points towards the chosen category compared to being honored by a stranger. The player can review their Honor points in each category in their profile, displayed as four separate integer values (their exact meaning is unclear). This information is also available to all players who wish to view someone else's profile.

After a player has consistently received honor in the appropriate categories, they will be presented with an animation on their player profile, and will be awarded a crest. There are four crests, but they do not all directly correlate with a single Honor category. For the "Honorable Opponent" category, this is achieved by consistently being honored in this category, while for the other categories, a combination of the three must be attained. Regardless of how many crests a player has obtained, only the rarest one will be displayed on their profile and during matches. Although Riot Games have kept the actual details of the inner workings of the system under wraps, it is known that a combination of "Helpful" and "Friendly" points will provide the players with the "Great Mentor" crest. The "Great Teammate" and "Great Leader" crests are a bit more unclear. Crests are temporary rewards designed to encourage players to remain honorable even after receiving the recognition that comes with acquiring a crest. Thus, they can be lost if players stop exhibiting positive in-game behavior and no longer gain any honor. Receiving punishments through the game's reporting system will also remove all crests from a player, similarly to honor points.

Figure 1: The four categories of the initial version of the Honor System, as seen in the game.

In 2017, Riot Games performed the first major update to the Honor System. The change was that voting was now done before entering the post-game lobby, and the categories were reduced to just three:

- **Stayed Cool**: for a teammate who stayed calm and was a provider of a positive playing experience
- **Great Shotcalling**: for a teammate who had good leadership and was able to lead your team to a triumphant win or a gracious defeat
- **GG <3 (sic)**: for a teammate who was Honor-worthy but didn't fit into a neat and tidy box of what they did to earn it

It is immediately clear that those new titles are far more vague than before. In this update, a player who received three votes had this achievement of theirs distinguished towards the whole post-game lobby. It also encouraged the whole team to vote, since that would lead to an overall slight increase in Honor for all teammates. Notice that players of the opposing team can no longer be voted for, as the official descriptions given for each of the three categories exclusively refer to "teammates".

Finally, in 2024, a major overhaul of the Honor System took place, leading to the one that exists in-game today. The titled categories were removed altogether and were replaced by levels of numerical representation [12, 33]. There are six Honor levels (numbered 0 to 5), with the levels above and including 2 indicating a positive behavior standing. Each Honor level has three checkpoints, except Honor level 5 which does not have any, as it is the highest. Players get promoted to higher Honor levels by receiving honors and playing matchmade (including player-versus-AI) games. Similarly to previous versions, progression is not explicitly (at least clearly) tied to the absolute number of honors received. At the start of each year, players at Honor level 3 and above have their level reset to 2 in order to progress through the ladder again, though the game rewards players for an additional checkpoint per level above level 2 upon this reset. Each checkpoint the player reaches and every level-up grants unique rewards. Honor levels 1 and 0 are restricted Honor levels obtainable only through penalties given by the reporting system of the game, as they signify a negative behavior standing, and progressing through them carries far fewer rewards than the positive behavior standing levels. Level 0 players are also labeled "Dishonorable". Before entering the post-game lobby, each player is given 40 seconds to commend up to four teammates or members of the enemy team (one time per player). Players receive by default one vote to use per match they finished, but can get extra votes if they were honored in the previous match and if they are currently at Honor level 5. Votes available do carry over between matches, but are limited to a maximum of four votes. Like the previous version, receiving three votes after a match will be indicated to the whole lobby. In terms of rewards, there are now three different ones: "Mastery Chests", "Key fragments", and "Loading screen flairs". While explaining these rewards is far beyond the scope of the subject matter of this thesis, it becomes amply clear that this latest version of the Honor System relies heavily on the extrinsic motivation of regular rewards (Figure 2). In terms of punishment, low levels are limited or prevented altogether from using the chat system, the ping system, the emote system, and the targeted ally system [33].

It is important to note here that the system is now updated more regularly, so it is possible some of the above information is slightly outdated at the time of reading.

2.5.2 Overwatch - Endorsements System. Blizzard Entertainment's Overwatch (the original game) became the first major online multiplayer shooter to introduce a behavior ranking system in 2018, two years after it was first released. This Endorsements System is meant to allow players to acknowledge other players' positive in-game behavior [1]. The system uses a five-level ranking approach (numbered 1 to 5). Players level up after receiving enough endorsements within a specific amount of time. Endorsement level 1 is the

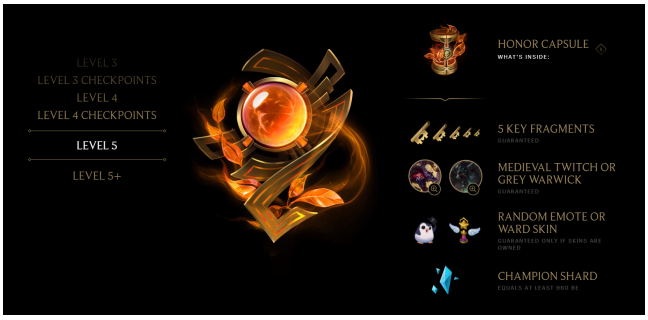

Figure 2: The details of Honor level 5 in the latest version of the Honor System, as shown in the game. This design showcases the shift to a strong focus on rewards.

Figure 3: The three different types of endorsements that could be given to players in the initial version of the Endorsements System, as shown in the game.

exception to this rule, since the only way to be ranked level 1 is to be punished by the game's reporting system, a demotion that occurs regardless of the previous level the player had. After each match, players can endorse up to two other players in total from either their team or the opposing team. Like the Honor System, endorsements are also anonymous, with the receiving player not been informed who endorsed them. Endorsing the same player is limited to once every 12 hours. In contrast with the Honor System's design, votes cannot be kept for use in later matches. To encourage consistency in positive in-game behavior, the player's Endorsement level is slightly diminished after each match, based on how many players are able to endorse the player. This initial version of the Endorsements System allowed the player to provide other players with three different types of endorsement (Figure 3):

- **Sportsmanship:** available for both teammates and players of the opposing team, this endorsement is offered to players that show a positive attitude throughout or just after the match
- **Good Teammate:** only available to teammates, this endorsement is offered to players that have proven to be helpful throughout the match (e.g., healing/reviving teammates) and/or have communicated effectively with teammates
- **Shot Caller:** only available to teammates, this endorsement is offered to players that have led the team throughout the match and/or have provided a strategy for winning the match

Similarly to League of Legends' Honor System, players are rewarded with in-game content based on their Endorsement level. More specifically, a player who maintains their Endorsement level is rewarded with in-game loot boxes, which are opened to unlock a large variety of cosmetics and other in-game items. The number of loot boxes given to the player each time they are rewarded depends on their Endorsement level. Namely, the player receives $Endorsementlevel - 1$ loot boxes each time. Level 1 players are not given any rewards, since those players are not considered to be currently of positive social standing. The current Endorsement level of each player was displayed using a simple icon (i.e., a simple design around the number of the level) on their profile page.

With the official release of Overwatch 2 in 2023 (2022 in early access), the Endorsements System was revamped. It followed the trend set by the Honor System, by removing the three distinct types of endorsement altogether, and opting for a simple "click to endorse" design [8]. The player's current Endorsement level is displayed in the scoreboard for teammates, on the career profile of each player, and in the voice-chat. The indicator for it simply showcases the player's current level with a design around it, similarly to the previous version of the system (Figure 4). Now, all players begin at Endorsement level 1 and can progress to a max level of 5, with every endorsement received impacting that level slightly. Also, while this was subsequently changed a year after in a much requested update, the system did not allow players to endorse their friends anymore. This limitation, along with the relatively small chance that a player will endorse players of the opposing team, made it quite difficult to be endorsed if you played with friends, especially since the game focused on 5vs5 matchmaking instead of the 6vs6 of its predecessor. Furthermore, because the game removed its loot box system in favor of the addition of a more modern battle pass system (loot boxes were added back in a 2025 update), the rewards received by players for their Endorsement level was changed to periodically earning battle pass experience points, based on the level. Players also receive a few battle pass experience points whenever they endorse another player after a match, further encouraging the consistent use of the system. Being punished by the reporting system may now strip players of their Endorsement level and place them at level 0 instead of level 1, which now is considered a level of somewhat positive standing. Finally, creating a publicly accessible custom match to others requires the player to be at Endorsement level 3 at a minimum, while using any of the in-game text-chat or voice-chat functionalities requires an Endorsement level of 2 at least.

It is worth mentioning, here, that the original Overwatch had a separate system that allowed players to vote for those that they believed performed admirably during a match. In the post-match screen, it showcased the most performing players in panels called cards. These cards would show statistics that were also available in the scoreboard, along with an "MVP" statistic and a "Gold Medals Earned" statistic. The first is based on the "On-fire Meter" functionality that the game had. The latter is earned by the player that received four or five gold medals during the match. Players could then vote for any one of those cards (their own included). Cards with five votes would reach the "Epic" status, leading to a sound effect and a hero-specific voice line. Cards with ten votes

reached the "Legendary" status, leading to a different sound effect and a hero-specific voice line. This system did not carry over to Overwatch 2. This may be a significant factor on the issues of the Endorsements System's design, since there are no longer two separate systems where one is based on social standing and the other on pure in-game performance. Specifically, the fact that players predominantly endorse support-class players, and those they found to have performed better, defeats the goal of a social standing system.

Figure 4: The indicators for each Endorsement level in the current version of the Endorsements System, as shown in the game.

2.5.3 Rainbow Six: Siege - Reputation System. In 2020, Ubisoft announced the integration of a morality system in Rainbow Six: Siege [68], called the Reputation System. The system's goal was set to be the discouraging of toxicity within the game's community and the promotion of positive in-game behavior. The design was meant to provide players with the tools they need to help them comprehend when and why an action of theirs (or of other players) is disruptive, primarily based on the game's own code of conduct. The system was slowly introduced into the game in 2021, but mostly in a phase of testing different strategies on what statistics to track, etc. During this testing period, no aspect of the system was visible to the player in the game itself or in their Ubisoft account page [7]. Like the aforementioned systems, the Reputation System awards each player with a score based on their social standing. Low ranked players have sanctions and messaging restrictions imposed upon their account, while high ranked players will receive some rewards for their positive standing.

In its maiden form, the Reputation System ranked players with a score according to the positive or negative actions taken during gameplay. This score was calculated according to two layers: actions and branches [3]. Action score is based on the various actions a player takes in-game, like reporting other players, muting teammates on text or voice chat, friendly fire, friendly gadget destruction, intentional team killing, and more (the developers do not fully list all actions). These actions lead to specific branches, based on how they relate to each other (the developers do not go into detail about how branches were formed). Player actions and their resulting branches are summed up to then influence the player's reputation (Figure 5). Based on the score, the player was assigned a level in a five-level system (numbered from 0 to 4). Players with Reputation levels 0 and 1 are considered to be of a negative standing, players with level 2 a neutral standing, and players with levels 3 and 4 a positive standing. A negative standing was obtained by players who repeatedly exhibited disruptive behavior and violated the game's code of conduct, which led to sanctions (e.g., suspensions from matchmaking, restrictions from specific playlists, and limitations on game currency earned) for a specified amount of time. On the other hand, players of positive standing were rewarded with in-game content like "Alpha Packs" and recognition within the community. This version of the system remained in a very primitive condition, with the developers opting to use it as a test bed.

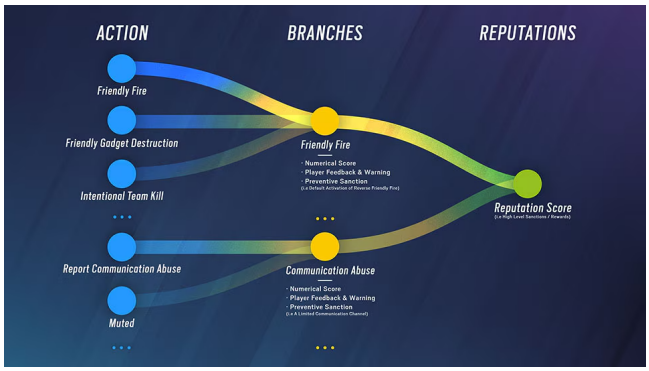

Figure 5: The actions and branches of the initial design of the Reputation System.

In 2022, Ubisoft launched the beta version of the Reputation System, along with its "Ubisoft Fair Play Program Beta" initiative [4]. The new Reputation System tracks a variety of in-game player actions, and assigns standardized measurement units called Units to them. This process allows for a more transparent and comprehensive to the player calculation and comparison, in stark contrast to the two aforementioned behavior ranking systems. Units are then classified into positive and negative Units, on the total of which the player's overall standing is based. Players who accumulate too many negative Units are considered to have a low standing, while those that receive positive Units are regarded as promoters of a healthy collaborative environment and have a high standing. All actions made by the player have immediate and lasting effects on their Reputation, with all actions being tracked for up to 100 matches in the last 90 days. This signifies that actions taken more recently are considered to be more representative of the player's behavior compared to actions taken in the long past, allowing the system to partially factor in player reform over time. Also, in contrast to the other two aforesaid systems, the Reputation System relegates the player's standing into five titled (as opposed to numerically labeled levels) Standings (Figure 6):

- **Dishonorable:** the lowest standing (negative category), Dishonorable players have shown repeated disruptive behavior and affect others negatively. Major improvement in conduct is essential to improve from this standing
- **Disruptive:** the second-lowest standing (negative category), Disruptive players have demonstrated disorderly conduct more often than honorable behavior, leaving room for improvement
- **Respectable:** the most common standing (neutral category), Respectable players are positive members of the community whose behavior has not affected others negatively very often. Players at this standing can still aim for a higher (and positive) one by pushing to do even better

- **Esteemed:** the second-highest standing (positive category), Esteemed players are valuable members of the community and positively collaborate with dedication and a level head
- **Exemplary:** the highest standing (positive category), Exemplary players serve as community cornerstones, positively influencing others while demonstrating strong leadership

Figure 6: The five Standings available in the latest version of the Reputation System, as seen in-game.

In the in-game overview page of the Reputation System, a visual representation of the player's current Standing is provided by the Standing Chart (Figure 7), where the horizontal axis displays the positive Units, the vertical the negative Units, the lines indicate the boundaries of each standing, and the dot marks the present Standing of the player. An emblem associated with the Standing is present next to the chart (all emblems are shown in Figure 6). The Standing also leads to impacts for low standings and bonuses for high standings. Impacts range from significant reduction in experience points (XP) and Renown (in-game currency) gained from each match, getting locked from ranked, standard, and other playlists, to simply preventing ranked mode rewards, depending on the Standing (Dishonorable players have the most major restrictions, Disruptive players have a few, while any other standing imposes no such restrictions). Bonuses are meant to commend players of high standing, who are provided with standing bonus points (more for higher standings) that eventually unlock in-game rewards.

Figure 7: The overview page available in the latest version of the Reputation System, as seen in-game. From left to right, the Standing Chart, Standing Emblem, and impacts and bonus can be seen.

Ubisoft's Reputation System combines a post-match player voting method (as seen in other systems) with automated action detection to calculate Units. This amalgamation allows for a more multifaceted approach to behavior ranking, by taking into account multiple aspects of the actions available to the player. The voting system itself is similar to the earlier versions of those present in League of Legends and Overwatch, with the noteworthy divergence that voting is not anonymous and the receiving player is informed about who commended them along with the type of commendation. Namely, at the end of each match, players vote to commend up to two teammates on one of three commendation types:

- **Valor:** a player who is positive, brings joy, inspires, and keeps cool
- **Dedication:** a player who is engaged and reliable
- **Guidance:** a player with good leadership, is knowledgeable, and helpful

Receiving commendations positively affects the player's Reputation (2.777 positive Units per commendation received), and if a player maintains 10 commendations in the past 15 matches over 7 days, they are considered to be on a commendation streak, which is marked by a flame icon next to their username that is visible to all players in a lobby. A "Worthy Opponent" commendation can be given by the whole team to the opposing team, signifying a team that was fair, fun, and fulfilling to play against (receiving such a commendation grants 3.5714 positive Units). Commendations, like actions, are recorded for up to 100 matches in the last 90 days.

The system also encourages positive communication that expresses thoughts, feelings, and ideas in a way that fosters collaboration in the game. In its present version, it only automatically tracks text-chat messages, and provides positive Units according to how many such messages were sent by the player during the match; 2.7777 for 1 message, 3.1250 for 2 messages, and 4.1666 for 3 messages. The player can view their points as represented on a ruler-like gauge, on the appropriate in-game page (Figure 8). Those are also recorded for up to 100 matches in the last 90 days. While voice-chat tracking is planned for a future update, no date has been set. It is noteworthy, that while the developers denote these Unit values, they do not indicate how they were chosen.

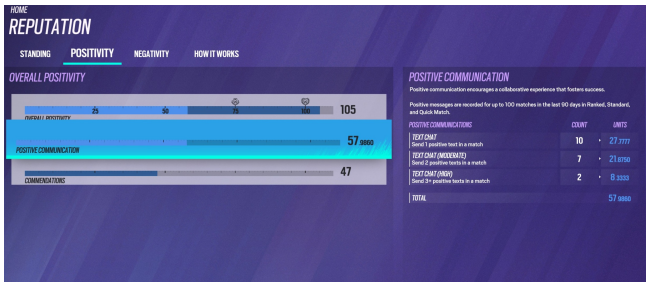

Figure 8: The overall positivity page of the latest version of the Reputation System. Here, the positive communication section is in focus, showcasing its ruler-like gauge representation.

A substantial portion of the Reputation System's design focuses on the negative actions that can be taken by players. Accumulating too many negative Units will cause the player's Reputation to decline. Those actions are also recorded for up to 100 matches in the last 90 days. In the negativity section, the player's negative actions

are outlined in a similar manner to positive actions (Figure 9), and are divided into six categories:

- **Abandons:** abandoning a match (i.e., leaving before it is over) affects all players negatively by removing the potential for a fun and fair match. For each abandon, players receive 6.25000 negative Units
- **Ally damage:** damaging allies is considered especially disruptive, since, by the game's design, every fraction of health matters. Damaging an ally 2 times grants the offending player 3.1250 negative Units, 3 to 4 times grants 5.0000, and 5 times or more grants 8.3333
- **Ally gadget destruction:** destroying devices placed by allies results in poor teamwork, flawed strategy, and lower chances of victory. Destroying 4 to 5 gadgets in a single match results in 4.1666 negative Units added to the offender, 6 to 7 results in 6.25000, and 8 or more to 8.3333 negative Units
- **Friendly fire:** downing, injuring, or killing an ally strongly impacts player experiences and disparages the competitive nature of the match, where even a single team kill can be detrimental. The "Reverse Friendly Fire" (RFF) penalty (i.e., the player trying to damage allies gets the damage dealt to themselves instead) can be applied if a player damages allies too much, and can carry over to multiple subsequent matches if the behavior persists. Intentional team kills result in 5.0000 negative Units added to the offender, while squad RFF leads to 3.5714 negative Units
- **Mousetrap:** console players who play with a mouse and keyboard in console matchmaking pools gain an unfair advantage compared to others who use the console's controllers, since matches are carefully balanced for each input type. If the relevant warnings are ignored, players will be forced to matchmake in cross-play pools only (i.e., PC and console). For each mousetrap detected by the system, the culprit receives 12.5000 negative Units
- **Negative communication:** communicating inappropriately disrupts collaboration and promotes a hostile environment (currently only applies to text-chat). A "Preventative Mute" penalty can be applied for too many negative texts in a single match. If 50 Units of negative communication are reached, the aforementioned penalty is applied to the next 20 matches. Improper texts can either be flagged or blocked altogether by the system, based on their severity. Offenders receive 3.1250 negative Units for 1 blocked message, 4.1666 for 2, and 8.3333 for 3 or more. Players who had messages flagged, receive 3.1250 negative Units for 2 messages, 4.1666 for 3, and 8.333 for 4 or more

2.5.4 Dota 2 - Player Behavior Summary. Valve Corporation added such a system in their Multiplayer Online Battle Arena game Dota 2. The first iteration was a simple "vote to commend" feedback system that was added in 2011 during the beta testing phase of the game, where players could commend others on either team [2, 10]. There are four types of commends: "Friendly" for players that were amiable, "Forgiving" for those that were quite forgiving, "Teaching" for those that were great teachers, and "Leadership" for those that were good leaders. In contrast with the other systems, Dota 2 players are able to accompany their vote with a short custom

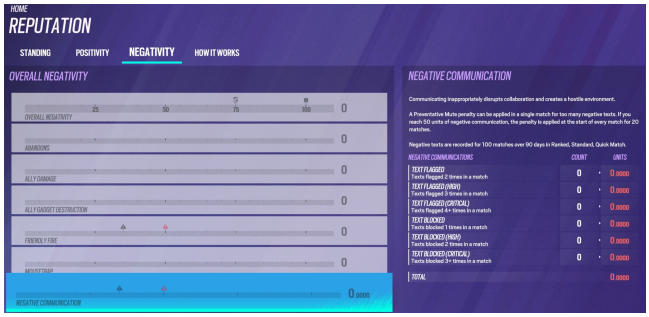

Figure 9: The negativity page of the latest version of the Reputation System. Here, the negative communication section is in focus, showcasing its ruler-like gauge representation.

message to the player they are commending (Figure 10). Naturally, those commends are given on the "Endgame" (post-match) screen after each match. Currently, each player can give out a total of 16 commends per week (although this number has changed over the game's lifetime). A chat notification is sent to the commended player, along with the commending player's name (not anonymous). Commends also affect a player's behavior score (how exactly is not known). Players with a high behavior score get matched with other players with a high score and vice versa [6]. The way that score works behind-the-scenes has not been disclosed by the developers, so this thesis will not delve any further into the topic.

Figure 10: The Commend page of Dota 2, as seen in-game. Here the player can choose a commend type and write a brief message for the chosen player.

In 2016, Dota 2 received the Conduct Summary system. It was an alert page that was automatically sent to the player after every 15 matches, if they receive 3 or more reports within those 15 matches, or if they enter "Low Priority" (i.e., they were penalized for abandoning matches, being reported too many times, or a pattern of bad behavior is detected) [5]. The page displays the number of commends and reports received over the last 15 matches and a "Lifetime Behavior Score" with a maximum value of 13, 000 (Figure 11). The system received a handful of updates, before being replaced

by the current Player Behavior Summary in 2023. It plays the role of a notice that alerts players of their behavior based on feedback from other players, and its design is quite similar to that of the system it replaced [15]. The behavior and communication scores of a player based on the number of reports are displayed, where the former reflects the quality of the player's in-game actions and dictates availability of features such as post-game item drops, game pausing and ranked play, while the latter reflects the quality of a player's in-game chat and speech interactions (Figure 12). If the player's communication score gets too low, they will be text-chat and voice-chat muted, and all other communication will be on a thirty-second cooldown. It is worth noting that those two scores are not dependent on each other, and that there is no longer an overall score. After reaching a certain behavior or communication score, the player unlocks (or loses in case of a lower score) specific gameplay or post-game features like access to ranked mode, post-game rewards, unthrottled messages, text and voice chat, etc. The page also contains a list of reports the player made that led to action being taken. Getting commended counts positively towards the player's score, although it is unclear on which of the two scores, or whether it boosts both.

Figure 11: The Conduct Summary alert page of Dota 2, as seen in-game. Here the player can see an overview of the commends they have received over the last 15 matches, along with the times they abandoned a match, how many times they were reported, and their communication abuse.

2.5.5 Others. Few other titles have attempted the integration of a behavior ranking system. This is perhaps surprising, given that players of recent games like Valorant (also made by Riot Games) [9] and Marvel Rivals by NetEase Games are continuously asking for such systems to be implemented. The reasoning behind these demands is improving matchmaking and reducing exposure to in-game disruptive behavior. In general, though, the vast majority of online multiplayer video games do not have a system for ranking behavior or for providing a way for the player to reflect on their behavior and perhaps reform or improve it. The reasoning behind this is challenging -if not impossible- to pinpoint, but the somewhat high monetary and time costs in relation to the low commercial gains (e.g., income and sales) of implementing such a system is likely a significant factor.

Figure 12: The current Player Behavior Summary alert page of Dota 2, as seen in-game. Here, the player can see their current behavior and communication scores, along with a list of reports the player made that led to action being taken.

2.6 Prosocial Behavior in Modern Games

Scientific papers and other sources about what has been done in the past around promoting prosocial behavior in games are limited, and often one-sided. Most examples are usually of games with an online aspect, rather than fully-fledged multiplayer games (especially with PvP competition in their design). The "soulsborne" (Dark Souls franchise, Bloodborne, and Elden Ring) games do have a messaging system (players are allowed to write messages using predefined words and phrases at any location), but this system is widely abused and filled with useless misinformation and inappropriate innuendos of all kinds. The most interesting system would be the one from Death Stranding (also includes a more primitive message-leaving system), which allows players to use objects that others have placed to traverse the environment (e.g., ladders, bridges, ropes, etc.). The problem here is that this is an implicit system, since players are in reality leaving those objects behind for themselves to reuse, rather than to explicitly help other players. All in all, both of those game examples do not need to solve any toxicity issue in them, since such a problem does not really exist to an observable extent.

H. Weszt in the Digital Thriving Playbook has multiple informative articles about prosocial actions in modern games and their design. He provides a list of 28 action archetypes (e.g., acknowledging, appreciating, collaborating, comforting, etc.) that can help proactively design a game for prosocial behavior [44]. He further showcases the existence of intended-by-design and inadvertent prosocial actions in games, by contributing a set of 25 examples from popular games [41]. Furthermore, to better facilitate prosocial design in games, H. Weszt demonstrates a three part design method [42]. First, aspects like setup screens, game lobbies, loading screens, and tutorials set up the stage for the behavior by establishing to the player what kind of prosocial activities are possible and expected during gameplay. Second, meaningful prosocial actions take place during gameplay sessions, which in turn build satisfying social experiences and happiness-generating habits. This can lead to the player valuing prosocial actions and thus encourages them.

Third, rewarding prosocial behavior helps bring prosocial gameplay to a satisfying conclusion and compels the player to keep acting prosocial over time and throughout multiple sessions. H. Weszt also references Requirements, Partners, and Moments (RPM), which is a design method for generating social satisfaction in multiplayer games, and aims to bridge the gap between system design and social design [45]. Its strategy is to make players feel good about themselves and attribute that feeling to other people. RPM leverages intrinsic motivation, trying to avoid extrinsic encouragements.

2.7 Identified Research Gap

In this literature review, it becomes clear that efforts to mitigate disruptive behavior in online video games have been only partially fruitful. At the same time, it has been found that prosocial behavior and rewarding prosocial actors can have a positive effect in a multitude of contexts, including video games. In fact, actions that can be considered prosocial are already available in many multiplayer online video games. However, research and industry attempts at explicitly increasing in-game prosocial actions have been few and limited to specific games (e.g., League of Legends and Overwatch), as well as mostly unsuccessful. The majority of games that incorporate some form of Behavior Ranking System heavily focus on the tracking of negative actions, while anything positive and prosocial takes a back seat. Because self-reflection has been found to be an important factor in learning and adopting new behaviors, as well as enhancing existing ones, this thesis aims to leverage this potential to boost prosocial behavior, by creating a novel tool that can be integrated with any online multiplayer video game that supports prosocial actions in the gameplay loop. Rainbow Six: Siege's implementation of the Reputation System is used as a small inspiration, as it tracks positive in-game text-chat messages and awards some points for them. However, its visualization of the player's scores is difficult to comprehend without reading over multiple pages and paragraphs of information on the system's implementation details, and the focus remains primarily on negative action tracking. In contract, the proposed system in this thesis tracks all prosocial actions and uses a more user-friendly visualization to help players self-reflect, self-regulate, and to encourage them to be more prosocial in a variety of ways, and focuses exclusively on prosocial actions.

3 Prosocial System Design

This section provides a detailed description of the design and functionality of the system of this thesis. It describes both the user interface (UI) and how the system itself calculates and displays its information to the user. It also clarifies the distinction between operability that is actually implemented in the system and the facets that are manually inputted into the system by the researcher.

3.1 Goal

The proposed novel system builds upon the Behavior Ranking Systems found in modern online multiplayer games, by providing users with a "prosocial rank", which is adjusted based on the number of "social points" the user gains from a match in any game. In a real-world use case, such a system would consist of an external application (e.g., similar to Valve Corporation's Steam) that interfaces with online multiplayer games using their Application Programming Interface (API), to promote in-game prosocial behavior. Compared to the systems mentioned in Section 2.5, this work puts the player's rank at the forefront of the information displayed, since it is the ultimate indicator of the player's progress. Thus, the ranking icon accompanied by the appropriate rank title (inspired by those that are present in traditional competitive ranking systems) needs to be clearly displayed at all times (i.e., on the "Overview" page, on each game page, and on the "Friends" page), so that the player is always fully aware what their rank is, how many social points they currently have, and how close they are to ranking up. The player should also be able to self-reflect on their prosocial performance effectively. To achieve this, the system provides individual pages for each game, where the "prosocial actions" are enumerated and described in detail. For "prosocial sequences", videos can accompany the description to help with understanding. The number of times each action or sequence was performed by the player is given in an intuitive way. The rank-point duo acts as the primary component of the system, intended to further encourage and stimulate prosocial behavior by the player.

Besides the ranking and self-reflection, an in-game rewards section showcases the player's progress towards game-specific cosmetic rewards based on their points in each game. This serves as a secondary component of the system, aiming to motivate prosocial behavior through content rewards. For the purposes of this thesis, these rewards are purely theoretical, since the player will not be able to make enough progress to unlock any of them during the experiment timeline, and there is no realistic way to provide actual content for published games that are not custom-made for this thesis. However, the rewards chosen for each game are real content available to be unlocked in-game and were selected so that their design fits the context (i.e., social interaction and behavior).

3.2 Definitions

To better describe the system, it is important to clearly define the various terms that are used within it, both for the purposes of the delineation of the design itself, as well as the writing of the thesis. In addition, these explicit definitions allow the player to better understand the information displayed to them by the system, without engaging with long documentation -both inside and outside of the game- that largely fails to concretely explain how the system at hand functions.

3.2.1 Prosocial rank (PSR). The titled rank of the player. The five titles are semantically ordered based on their significance and perceived value. The titles themselves follow standard naming practices found in competitive ranked modes of the majority of online multiplayer games (Figure 13), since players are likely to be already acquainted with them, and using a form of "social-related" naming practice may be perceived as silly. From lower to higher, the rankings are:

- Copper
- Bronze
- Silver
- Gold
- Platinum

It is important to note that it is not within the system's bailiwick or its general design goal (as well as the purpose of this thesis) to blueprint a high-quality and balanced ranking scheme. This would require a lot of iterative development and testing, and given the design of the user study at the core of this thesis, it would be improbable to serve any purpose. The participating players will only engage with the proposed system once, meaning that there is not enough data from a large number of gameplay sessions that would warrant a more sophisticated ranking scheme. Namely, what matters the most for this thesis are the social points and the self-reflection on the prosocial actions performed and not performed.

Figure 13: The five available ranks in the Prosocial System with their icons and titles.

3.2.2 Social points (SP). Instead of the usual "experience points" (XP) or "matchmaking rank" (MMR), the proposed system uses "social points" (SP) to measure the player's progress and assign them a rank. Players can earn SP by performing prosocial actions in any of the supported games. After earning a certain number of SPs, players rank up, unless they are already at the maximum rank of Platinum. In the event of the release of a public version that is properly integrated into games, at the start of each season, the system would reset the points earned to the minimum number required to reach the current rank, or the immediately lower rank. This aims to encourage players to continue exhibiting prosocial behavior.

3.2.3 Prosocial actions (PSA). Predefined actions in each game that are of a prosocial nature. For the purposes of this thesis and its study, these actions are defined by me, based on my personal experience with each game, as well as by reaching out to the communities for help. The communities were allowed to propose any actions they wished based on their own personal experience, attitude, and opinion. This was achieved primarily through the respective community forums, subreddits, and Discord servers. The answers were evaluated and the actions that fit the description the best were added to the system.

3.2.4 Prosocial sequences (PSS). Predefined sequences of actions in each game that are of a prosocial nature. They differ from PSAs, as the actions of these sequences when isolated do not necessarily constitute prosocial behavior. Similar to PSAs, these sequences are defined by me and the community, and were gathered in the same exact way.

A comprehensive list of all prosocial actions along with a description for each is provided in the method segment of this thesis (Section 4).

3.3 General Design

The current iteration of the system contains three types of pages. The "Overview" page, the "Friends" page, and the "Games" page. The latter page comprises the three respective pages of each available game (i.e., Helldivers II, Overwatch 2, and Rainbow Six: Siege) with the player being able to switch between them. As mentioned, the player's rank remains visible on all pages, but more so and with more detail on the "Overview" page. The player can switch between the three available page types by clicking on the appropriate tab of the page they want (top of the window). On the "Games" page, they can switch between the available games using another set of tabs. Customizable settings are not available for this version, but can be easily added to a public release version.

3.4 Overview Page

This page's focus lies on the general prosocial performance of the user (Figure 14). In its center, it clearly showcases the player's PSR in detail, including the relevant rank icon and title, and a progress bar along with their SPs indicating how close they are to the next rank. On the left side, their most recent PSA and PSS are denoted, along with the SPs earned for each game individually. On the right side, the player's previous ranks in past seasons (on a month-to-month basis) are displayed. Those ranks are the same for all players for the purposes of this thesis, despite the fact that the months and years fields are properly calculated based on the current date.

3.5 Friends Page

This page provides a list of the player's friends (e.g., through Steam or some other platform), along with their individual PSRs, SPs, recent PSAs and PSSs, and their SPs earned for each game separately (Figure 15). For the purposes of this thesis, the friend list is filled with fake names, but it is entirely possible that a publicly released version retrieves a real list, using the numerous APIs available. This is perhaps the least significant page of the system when it comes to the study, as all of its functionality -except the rank and points part that is present in all pages- is simulated and filled with information that does not pertain to the participant and their in-game performance. It still, however, plays an important role in illustrating how the system would actually function when it comes to the aspect of comparing oneself with friends.

3.6 Games Page

This page contains separate tabs for each of the three currently supported games. The player can choose for which game they want to see the relevant information by clicking on the respective tab. Each game page contains highly detailed information about the available PSAs and PSSs of each game, and the SPs earned by the player for this specific game (Figures 16, 19, and 20). The player can hover over each PSA and PSS to get a tooltip text description along with examples on how to replicate the given behavior in-game (Figure 17). The description of both the actions and sequences is often accompanied by an image or a video showcasing what it looks like in-game and how each can be performed properly (Figure 18). For sequences, descriptions illustrate how each action leads to the next to form the sequence itself. Next to each action and sequence, the player can see how many times they have performed

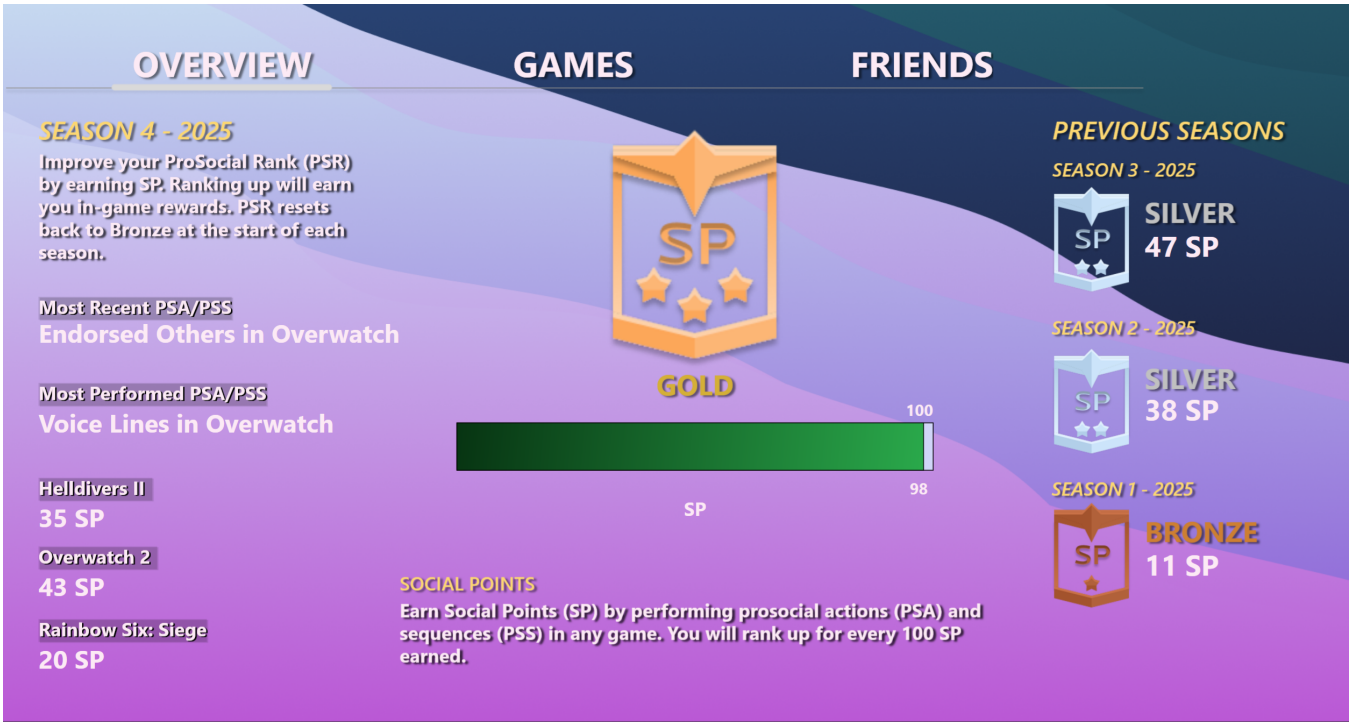

Figure 14: The Overview page in the Prosocial System. The right side contains a column of ranks and SPs for the previous three seasons. The center contains the player's current rank and their progress towards the next one. The left side, contains the most recent prosocial action/sequence performed, the prosocial action/sequence performed the most times, and the social points for each of the three supported games.

the action or sequence in-game. On the rightmost side of the page, the two in-game rewards (simulated as part of this thesis) and the SPs they require are displayed. From the perspective of the player, the actions are not differentiated from the sequences in any clear way. Doing so would likely negatively impact the system, since from the player's view, it is not relevant whether a certain in-game behavior they exhibited is classified as an action or as a sequence. Furthermore, since it would be impractical to attempt to assign different SPs to each PSA and PSS in an appropriately justifiable and sufficiently balanced manner, the simple -but satisfactory- solution to allot 1 SP per PSA and PSS performed by the player is adopted. From the player's side, it would be more confusing to have two different numerical values for each PSA or PSS (i.e., one for the SPs and one for the times it was performed), and would also render the game pages a lot larger and harder to navigate, negatively affecting self-reflection. This design has the added benefit of making comparisons between the multiple sessions for the same player, or among different players, a straightforward process.

3.7 Lack of Negative Actions

In all the Behavior Ranking Systems mentioned in Section 2.5 a strong focus on tracking negative actions is evident. As discussed in detail, the detection of negative behavior exhibited by the player results in a large variety of punishments. In contrast, the proposed system avoids tracking any negative actions or taking them into consideration when calculating the player's points and ranking. As cited in the literature review of this thesis [79], it is clear that punishment is often not as effective as reward in promoting a target behavior. In fact, the goal of the proposed system is to encourage prosocial behavior, instead of eliminating disruptive behavior by incurring a significant change in the player's behavior and attitude. As such, no negative actions are needed in its design.

3.8 Technical Details

A system of this design and objective can be implemented in a number of different ways. A website would be quite appropriate for this sort of system, would somewhat guarantee multiplatform support (e.g., Windows, Linux, macOS, OSX, etc.), and would allow relatively easy access for players. However, safely hosting a website that would in some way contain private information concerning its users (i.e., the participants of the user study of this thesis) is anything but trivial. Thus, the website route was avoided. A second option would be to create a desktop application for the task, with an executable build that is given to the user, containing only their own data. This makes multiplatform support difficult to achieve (in the context of this thesis), but this does not pose a significant hurdle, as the target group of the system is personal computer (PC) gamers, who predominantly play on Windows 10 or Windows 11 (96.10% of Steam players use a version of Windows [16]). Consequently, the system was developed in the form of a Windows Presentation

Figure 15: The Friends page of the Prosocial System. It contains a list of the player's friends (placeholders for this thesis), along with their individual PSRs, SPs, recent PSAs and PSSs, and their SPs earned for each game separately.

Platform (WPF) application. To ensure that users were not burdened with the obligation to install any piece of software on their personal computers, the Prosocial System application was built and packaged in such a way that it does not require any installation of itself or any prerequisite software. This allows the application to function out of the box, as long as the target PC has the .NET 4.8 SDK installed, which should already be present in all modern and older systems. Furthermore, special care was taken, using extensive internal testing and evaluation, to ensure that the application functioned as intended and UI/UX quality remained high on all platforms and different monitors.

The Prosocial System functions by reading all the required data from four different JavaScript Object Notation (JSON) files. There are three files for the prosocial action information (one file per supported game), and one file for the statistics concerning the most recent prosocial action and the one that was performed the most. That means that the system does not -in its present version- keep track of timestamps for the actions, which is why the aforementioned statistics have to be filled in manually (the order of performed actions is tracked during the user study; just not programmatically in the system itself). Namely, the three JSON files that concern the game information are filled in with the number of times each available action was performed by the player. The vast majority of functionalities are implemented, and only the few aspects that were mentioned in the above subsections are simulated or filled in manually.

4 Method

This section describes the design of the experiment in this thesis. The experiment is in the form of a user study, with the goal of investigating the effect that the Prosocial System described in Section 3 has on in-game prosocial behavior. The results of the study are utilized to answer RQ1 and RQ2. It is partially based on the work of Kleinman et al., since it supports the design and is of similar nature, with self-regulated learning being leveraged, as well their interview questions been borrowed [50].

4.1 Overview

To explore the effects of the novel Prosocial System, a two-phase user study is required: participants play two sessions of the same game, in between of which they interact with the Prosocial System. To maximize the comfort of the participants -as well as increase the number of those interested in participating- the sessions were conducted online through the Discord group chat platform, with the participants playing from their own personal computers and from their own space. For the first session, participants played one or two matches (depending on the selected game) which were recorded by the researcher through screen sharing on the participant's side, and notes were taken by the researcher. After the session ended, these recordings and notes were processed into results in the Prosocial System. In the second session, which always took place on a later day (i.e., at least a day apart between the two sessions), participants were given time to explore their results in detail, as well as ask any questions. Then, they played the same number

Figure 16: The Overwatch 2 Game page of the Prosocial System. It contains detailed information about the available prosocial actions/sequences in the game, and the social points earned by the player for this specific game. On the rightmost side, it indicates which two in-game rewards are available and the social points they require.

Figure 17: A descriptive text of a prosocial action, shown in the form of a tooltip in the Overwatch 2 Game page of the Prosocial System. It is displayed after the player hovers their mouse cursor over a prosocial action for a few seconds.

Figure 18: A descriptive text accompanied by the appropriate image of a prosocial action, shown in the form of a tooltip in the Helldivers II Game page of the Prosocial System. It is displayed after the player hovers their mouse cursor over a prosocial action for a few seconds.

of matches in the same game, which was again recorded in the same manner. The recordings were again processed into results to investigate how the prosocial performance changed between the two sessions. In specific parts of each session, the participants were presented with a custom questionnaire (one per session). The first focused on demographics and Tondello et al.'s gamification user type Hexad scale [74, 75]. The second questionnaire included a section on the system's effect on the participant's self-reflection surrounding prosocial behavior, a part on evaluating the various different components (i.e., rewards, self-reflection, and ranking) of the system, and a part about the general design of the system (i.e., usability).

A broad outreach on multiple social platforms, forums, and Discord servers, either directly associated with the supported games or related to gaming in general, yielded 16 participants. They were

Figure 19: The Helldivers II Game page of the Prosocial System. It contains detailed information about the available prosocial actions/sequences in the game, and the social points earned by the player for this specific game. On the rightmost side, it indicates which two in-game rewards are available and the social points they require.

allowed to choose any game they wanted from the three that were supported. The only stipulations were that they should choose a game they already own (even though Overwatch 2 is free-to-play) and that they are already familiar with. Returning players that had not played the game of choice in some time were not excluded, as long as they felt they were still up-to-date with the game's mechanics. By following these rules, it is ensured that the in-game behavior of each player is a good representation of the average player behavior in the given game (e.g., no novelty bias).

4.2 Selected Games

One of the primary goals of this thesis, was to propose a modular novel system for prosocial behavior that can be integrated into any online multiplayer game that already has a framework that supports in-game prosocial actions. To evaluate such a system, it would not be sufficient to support only a single game or even two. Thus, it became apparent that the sweet spot was three games, as this number would reinforce the modularity aspect of the system without being so high that some games end up with no players and data. The three games that were chosen are Helldivers II, Overwatch 2, and Rainbow Six: Siege. They were chosen based on various factors. Primarily, all of them have a preexisting framework with a satisfactorily high number of identifiable prosocial actions that are integrated into their gameplay loop, as opposed to games that either do not contain many (or any) such actions, or games where prosocial behavior can be seen as disruptive to the gameplay because it may impact performance or feel unnatural to the player. The second factor is their popularity, as all of the three games have high player counts, as well as active and passionate communities, both of which should increase the sampling pool of participants. At the time of writing, Helldivers II has 74, 838 players on Steam (all-time peak of 458, 709) [11], Overwatch 2 has 40, 235 (all-time peak of 75, 608) [14], and Rainbow Six: Siege has 137, 895 (all-time peak of 201, 933) [17]. These numbers are actually much higher, as they do not include all platforms (i.e., consoles) and launchers (i.e., Battle.net and Ubisoft Connect), since only Steam data is publicly available and reliable to reference. The third factor is that this triad of games are each good representations of modern online multiplayer video games, and Overwatch in particular has already been the subject of previous research [39, 76]. Lastly, while all of them fall under the multiplayer shooter category, they are very different representations of the genre, with Helldivers II being a cooperative PvE shooter, Overwatch 2 being a PvP team hero-shooter, and Rainbow Six: Siege being a classic competitive PvP shooter. In fact, there is no reason to believe the proposed system only functions for shooters, as it does not in any way rely on their characteristic mechanics.

It is also beneficial to cite a few examples of different prosocial behavior in the chosen games. All three have voice and text communication integrated, so players suggesting actions, helping and encouraging other players, and praising the actions of others, all constitute prosocial behavior. In Overwatch 2, the community was

Figure 20: The Rainbow Six: Siege Game page of the Prosocial System. It contains detailed information about the available prosocial actions/sequences in the game, and the social points earned by the player for this specific game. On the rightmost side, it indicates which two in-game rewards are available and the social points they require.

asked for examples of prosocial actions and why they interpret them as such. There were many responses, which suggested actions such as head patting people as Ramattra (character that extends their arm forwards), using "Hello"/"Thank you"/"You're welcome" voice lines, emoting with the whole team when not in a fight or during the spawn phase, using the ping system which can be used as just informational sharing or to specify danger to teammates, using the communication wheel, swapping characters to better synergize with the team composition, and endorsing your teammates. In Rainbow Six: Siege, players can coordinate their actions by communicating through the ping system and voice-chat. Because of the competitive and fast-paced nature of the game, encouraging teammates can significantly affect performance in high-stress and high-intensity conditions, allowing everyone to adapt swiftly to the rapidly changing situation. Finally, in Helldivers II, outside of common emotes, there are two-player emotes like hugging and playing rock paper scissors, as well as a quick reaction wheel for "Thank you" messages. Given the respawn-heavy nature of the game, players can use the ping system to help respawning teammates locate their dropped equipment and resources. The game also provides numerous multi-player interactions, where players have to coordinate to open a bunker door, and cooperate to complete various objectives faster and more efficiently. Equipment can also be shared, so that players can try content they have yet to unlock, as well as become better equipped to get out of a gameplay situation.

4.3 Available Prosocial Actions

To better understand what these prosocial actions are and how they are performed in the context of each game, an all-encompassing list of all the available actions follows. Some of the listed actions are likely easy to understand, even if the reader is not experienced with the respective game. However, a few may be harder to interpret by inexperienced players. This subsection aims to help alleviate this potential issue, but in case it is not enough, interested parties can easily search for these actions themselves, as videos and other representations are readily available online.

4.3.1 Helldivers II. The available prosocial actions and sequences for Helldivers II, as listed and described in the proposed Prosocial System:

- **Hug or Other Coop Emote:** perform a hug, handshake, salute, or high-five emote with a teammate
- **Rock Paper Scissors Emote:** perform a rock, paper, scissors emote with a teammate
- **Affirmative Reaction:** respond affirmatively to a teammate's ping or callout
- **Pinging Another Player's Gear:** help a respawning teammate find their dropped gear by pinging it for them
- **Opening a Buddy Bunker:** open a bunker that requires two Helldivers to open
- **Cooperating to Align a Radar:** align a Radar Objective with the help of a teammate

- **Rerouting e710 Objective:** cooperate to reroute an e710 Objective
- **Unlocking the ICBM Objective:** cooperate to complete the launch process for an ICBM Objective
- **Stimming Players that are Injured:** aid an injured Helldiver by using a stim on them
- **Using the Supply Pack to Resupply your Teammates on the Move:** resupply your fellow Helldivers with a Supply Pack
- **Team Reloading:** help fire a weapon that requires team reloading
- **Bringing High-level Equipment for Lower-level Players to Try Out:** bring higher level equipment for the newer Helldivers to try out
- **Using the Ping System:** use the ping system to indicate something to your teammates. Example: point out danger
- **Saluting Together on Extraction:** salute with the whole team on extraction
- **Sharing Equipment Requested by Other Players:** call down equipment that another Helldiver needs
- **"Thank You" Emote-Wheel Action:** use the emote-wheel to say "Thank You"
- **Praising Teammates on Voice-Chat:** praise your fellow Helldivers on voice-chat. Be encouraging
- **Praising Teammates on Text-Chat:** praise your fellow Helldivers on text-chat. Be encouraging

4.3.2 Overwatch 2. The available prosocial actions and sequences for Overwatch 2, as listed and described in the proposed Prosocial System:

- **Endorsed Teammates:** endorse your teammates after a match
- **Swapping Heroes to Synergize with the Team:** swap heroes with a teammate to maximize team performance, or change to a different one to better synergize with the rest of your teammates
- **Typing GLHF, GG, OR WP:** type a friendly GLHF, GG, or WP in the All or Team chat
- **Congratulating an Opponent that Killed You:** congratulate an opponent that bested you. Humble, even in defeat
- **Headpatting People as Ramattra:** be nice to people by headpatting them by extending your arm forwards while playing as Ramattra
- **Emoting with the Whole Team:** emote with the whole team when not in a fight, or during the spawn phase
- **Using the Ping System:** use the ping system to indicate something to your teammates. Example: point out danger
- **Cooperative Spray System:** use the spray system with a teammate. Some sprays are specifically designed to be half or part of a whole larger spray image
- **Reporting Ultimate Status:** coordinate with your team by reporting the status of your ultimate. Useful for team play and ultimate combinations, or indicating you may be able to turn a fight if you get "rezzed" (resurrected) because your ultimate is available
- **Protecting a Vulnerable Teammate:** protect a vulnerable teammate by physically body blocking or using a skill to protect them when they have been slept or crowd controlled in some way
- **In-Game Non-Verbal Communication:** use non-verbal communication to coordinate with your teammates
- **"Hello", "Thank you", "You're welcome" Voice Lines:** use the communication wheel to talk to your teammates
- **Praising Teammates on Voice-Chat:** praise your teammates in voice-chat. Be encouraging
- **Praising Teammates on Text-Chat:** praise your teammates in text-chat. Be encouraging

4.3.3 Rainbow Six: Siege. The available prosocial actions and sequences for Rainbow Six: Siege, as listed and described in the proposed Prosocial System:

- **Encouraging Last Teammate Standing:** be encouraging and helpful to the last teammate standing. Maybe they will clutch that round
- **Using the Ping System:** use the ping system to indicate something to your teammates. Example: point out danger
- **Coordinating Rotation Holes:** communicate with your teammates to suggest and coordinate rotation holes and reinforcements during the preparation phase. Remember to suggest ideas. Do not just give orders
- **Swapping Operators:** swap operators with a teammate to maximize team performance, or change to a different one to better synergize with the rest of your teammates
- **Congratulating an Opponent that Killed You:** congratulate an opponent that bested you. Humble, even in defeat
- **Placing Valkyrie Cameras on Request:** place Valkyrie cameras on a location suggested by a teammate. They may need some extra surveillance
- **Typing GLHF, GG, or WP:** type a friendly GLHF, GG, or WP in the All or Team chat
- **Non-verbal Emoting and Communication:** use your operator's movement to communicate with your teammates. Going prone and leaning is still fun
- **Praising Teammates on Voice-Chat:** praise your teammates in voice-chat. Be encouraging
- **Praising Teammates on Text-Chat:** praise your teammates in text-chat. Be encouraging

4.4 Recruitment

As mentioned before, the nature of this thesis led to the use of convenience sampling to find participants. Due to its nature as a master's thesis project study, no compensation could be offered by the researcher to those that took part, as there was no budget - personal or otherwise- available for it. The study itself was designed to ensure participant comfort and minimize any friction during the study. That is, the longest part of the study has the participant play a few matches in a game they already like and oftentimes would be playing anyway. This characteristic, coupled with the relatively short time required to participate, were used as the primary encouragement for participating, to boost recruitment efforts. While the majority of recruitment was done online through various platforms, attempts to scout out interested parties were also made in the context of in-person interactions. Namely, university students (bachelor's and master's) were made aware of the study through

a brief presentation along with a quick-response (QR) code that led them to the study's information page. Despite a few students showing interest in participating, in the end no participants were gathered from this effort.

In terms of the online recruitment, an information page for the user study was set up to provide all the necessary facts, and the relevant links to help interested parties schedule the two sessions. To keep things simple and straightforward to share, the page was put together on a Google Docs document. On the top, it included all the supported games and a reminder that participation is limited to PC players only. In all online posts, the study was promoted to the public as a study on prosocial behavior in games, usually specifically on the game related to the community it was posted to. The general structure of the study was mentioned along with the expected duration for each session, but those interested were encouraged to follow the linked information page to learn more. To avoid overloading those interested with unnecessary information, the information page was kept as short as possible, including only the information strictly essential. It comprised three short sections, filling a bit more than a single page. The first section encompassed the two ways in which someone could indicate their interest and schedule their participation: either by joining a Discord server that was specially set up for the study, or if they preferred, through personal email to the researcher. It also included a link to a read-only (to the public) Microsoft Excel spreadsheet that was kept up-to-date with all the dates and time slots available to them. Participants were asked to choose an available time slot from the sheet and then request it in one of the two aforementioned ways. The second section provided a detailed step-by-step guide on what the participants will do during the study, along with the expected duration for each. The final section contained all pertinent links for the study: a second hyperlink to the time slot spreadsheet, the consent information sheet document observing Utrecht University's informed consent guidelines for online studies (see Appendix B), and a second hyperlink to the Discord server. It is important to note that no participant (or otherwise) information of any kind was stored during their interaction with the Google Docs document and the Microsoft Excel spreadsheet, to ensure maximum privacy. The Discord server was also used only for scheduling and answering any questions, with all sessions taking place in private calls.

All three of the supported games have substantial and active communities on both Reddit and Discord. Each game was approached at a separate segment of the time allocated to the experiment. This was done to prevent the posts from being labeled as spamming behavior, given that Reddit allows only a few promotional posts in a short length of time, as well as maintain control of the process, since it is significantly harder to switch between different games in a short period of time when recording data, without the assistance of additional researchers. The recruitment commenced with Overwatch 2 through the two related subreddits. The first participants were successfully recruited from those posts, with quite a few more that followed. Subsequently, the moderators of the Discord servers of all three games were contacted, to ensure it was within the rules to post about the study. Unfortunately, after a long wait for responses, only the Overwatch Discord server gave the green light. Interest was garnered, but ultimately no participants were recruited from Discord servers. The second game community that was approached over Reddit, was Helldivers II. A few participants enlisted through these posts, but much less interest was evident. Lastly, posts on the two Rainbow Six: Siege subreddits yielded no participants. Nevertheless, two Rainbow Six: Siege players were enlisted into the study, as they were acquaintances of the researcher. Some other forums, like on Steam and on Blizzard Entertainment's website, were also reached out to, but no useful responses were received (i.e., random comments that did not have to do with the study, and people being suspicious about its intentions), quickly turning into a naive opprobrium of the researcher. Some student associations were also contacted through email, but no reply was received. The study was also promoted on a few WhatsApp groups, to no effect. Finally, a website offering "call for participants" services was utilized, but despite over 200 viewers (according to the website's metrics), nobody signed up, except one that did so accidentally and did not partake in the study.

The Overwatch community was the most interested and most willing to participate in the user study. They also expressed by far the most interest, and were excited about the outcomes of the thesis and how it could actually be used in real applications. Overall, all possible routes were explored exhaustively in the process of recruiting participants, and the author firmly believes that more time would in no way benefit this process. Despite all this and the clear difficulty in recruiting for a study that is unable to offer any compensation to its participants, the process was deemed a success in the end.

4.5 Scheduling

All participant scheduling was done online. Each was scheduled for two sessions, which took place at least a day apart. Those interested in taking part selected their own time slots from the Microsoft Excel spreadsheet that was provided to them, by contacting the researcher through personal email or by posting on the appropriate channel on the Discord server. Oftentimes, participants would opt to schedule the first session only, with the second being scheduled towards the end of the first session. Each participant was asked to choose a game they were familiar with to avoid any novelty bias. The consent information sheet was already accessible to those interested, and they were expected to have read it before proceeding with the study. Their agreement to informed consent was confirmed by them during the first session questionnaire.

4.6 First Session

During this session, participants were given a brief verbal introduction to the study and answered the first custom questionnaire at the beginning of the call. Then, they proceeded to play their game of choice, which was recorded. The purpose of this first session is to gather the data relating to prosocial behavior before any interaction with the proposed Prosocial System.

4.6.1 Setup. First, participants selected a game from the list of supported games that they reported feeling familiar and comfortable with. The researcher and the participant connected on a private Discord call so that the study could be conducted. They were asked to fill out the first questionnaire, where they confirmed their consent using a tick box, filled out the Hexad scale which maps the user's personality onto the design elements of gamified systems by

calculating their preferences [74, 75] (the instructions on the scale's proper usage were followed), and some standard demographics questions, including how experienced they were with the game of their choosing (see Appendix C). When they were finished with the questionnaire, they were asked to share their screen (application sharing only, through Discord) while they were in the chosen game. This ensured all personal information irrelevant to the study was not visible during the recording process.

4.6.2 Hexad Scale. In order to better assess the gamification aspect of the proposed Prosocial System, the Hexad Scale was utilized to assign each participant to a user type [74, 75]. This scale is based on Self-Determination Theory [37, 69] and includes six subscales, one per user type:

- **Philanthropist:** motivated by purpose, altruistic, and willing to give without expecting a reward
- **Socializers:** motivated by relatedness, want to interact with others, and create social connection
- **Free Spirits:** motivated by autonomy, freedom to express themselves, and act without external control
- **Achievers:** motivated by competence, seek to progress within a system by completing tasks, and prove themselves by tackling difficult challenges
- **Players:** motivated by extrinsic rewards, will do whatever to earn a reward within a system, independently of the type of the activity
- **Disruptors:** motivated by the triggering of change, tend to disrupt the system either directly or through others to force negative or positive changes

Each subscale has four items on a Likert scale ranging from 1 (Strongly Disagree) to 7 (Strongly Agree). Those four items are summed up together to calculate a score for each subscale. The one with the maximum score is deemed to be the assigned user type.

4.6.3 Gameplay. With the setup concluded, the researcher would start the recording, informing the participant about it as well. The participant would play one to three matches (as defined by each game), with the goal of reaching 15 to 25 minutes of recorded gameplay. More specifically, Overwatch 2 and Rainbow Six: Siege required two matches to reach the target gameplay runtime. Helldivers II was an exception, since its matches are on average 40 minutes each. Thus, only one match was played for the latter, and the resulting gameplay time was twice the average time of the other two games. The core aim was for the optimal amount of gameplay so that a sufficient representation of average player behavior in the game could be obtained.

4.6.4 Post-Gameplay. After the gameplay phase was concluded, the recording was stopped, and the participant was scheduled for the second session if they had not already chosen a time slot. Any questions about the study were answered as well.

4.7 Gameplay Data Analysis

Following the end of the first session, the relevant gameplay data concerning the participant's prosocial performance had to be extracted. The method followed was twofold. First, the gameplay that was recorded had to be carefully viewed to detect and log all prosocial actions that were performed, and to make general notes and observations that were deemed interesting (e.g., total gameplay time, patterns in behavior, and comments made by the participant during gameplay). Second, notes were taken by the researcher during gameplay, both on prosocial actions performed and any other interesting observations. This was primarily done as a precaution in case the video recording encountered an error during or after the recording itself. Shortly after retrieving all the data relative to the study, the videos were deleted to comply with ethics and privacy guidelines and rules.

4.8 Second Session

In the course of this session, participants were provided with a build of the Prosocial System along with their data written in the JSON file it parses and reads. They were then walked through all the components and the pages of the system, before being asked to self-reflect on their prosocial performance during the first session. They then played the game of their choosing for the same number of matches as the initial session, which was again recorded. In the end, they completed the final questionnaire and were given time to ask the researcher any questions they had about the study.

4.8.1 Setup. This procedure did not differ much from the same one during the first session. The participants were provided with a OneDrive link to download the build of the proposed system along with their data. If any help was needed, the researcher was available to assist.

4.8.2 Prosocial System Data Exploration. After extracting the downloaded archive, participants were asked to execute the application and share its window through Discord (i.e., only the system is visible to the researcher) so that they could be guided through the system's components and pages. This walkthrough ended on the Game page of the game of their choice. The screen recording was started so that how participants explore their results in the system can be investigated, as well as keep track of their answers to the researcher's questions. They were then asked to take as long as they wished (approximately 2 minutes) to look through the list of all available prosocial actions, self-reflect on their prosocial performance by investigating where they received points and where they did not, and ask any questions that arose. After this self-reflection process, participants were asked to answer the following questions (adapted from the ones Cleary et al. used in their study of self-regulated learning in basketball players [35]), targeting the three main processes of self-reflection according to B. J. Zimmerman's Cyclical Phase Model (CPM) of self-regulated learning (SRL) [83]:

- Please evaluate your prosocial behavior during this game and elaborate on how you are reaching your conclusions (Evaluation)
- Please identify something during this game that you think you did not do particularly well, and elaborate on what you think caused your poor prosocial performance (Attribution)
- What do you think you need to do to improve your prosocial behavior in your next game (Adaptation)

This trio of questions was meant to assist participants in their self-reflection process, by stimulating them to briefly evaluate their prosocial performance, justify why and where they did not perform optimally (according to themselves), and ascertain what they

need to change to increase their prosocial performance. Before this recording was stopped, the participants were allowed to look through the rest of the system, as well as to further investigate their results if they so wished.

4.8.3 Gameplay. Following the data exploration-interview part, that recording was saved, and the gameplay recording (i.e., only the application was now shared) was started. The participants then played the same number of matches in the same game as in the first session.

4.8.4 Post-Gameplay. After the gameplay was finished and the recording was stopped, the participant filled out a questionnaire (see Appendix D) with a section about their experience with the design of the system (i.e., its usability) and a section about the effects of the system on prosocial behavior. For the former, the System Usability Scale (SUS) by J. Brooke [30] is leveraged, as it is validated, it has been used in similar applications for the evaluation of their usability, and it produces a comprehensive result. For the latter, a set of custom Likert scale questions inquire about the effect of the system on the participant's self-reflection surrounding prosocial behavior, and how much each component of the system would encourage them to be more prosocial in a game. Questions can also be asked by the participant about the experiment, as part of the exit interview.

4.9 Final Gameplay Data Analysis

The recordings were again processed into results to investigate whether the prosocial performance changed between the two sessions. These were also made available to the participants after they were extracted (does not affect the study). Besides the gameplay data, their interaction with the system was also observed and analyzed. Finally, their answers to the three SRL questions were transcribed. Similar to the first session, shortly after retrieving all the data relative to the study, the videos were deleted to comply with ethics and privacy guidelines and rules.

4.10 Prosocial System Interaction Analysis

To better assess the responses given by the participants on these questions, the SRL construct of evaluation as proposed by Cleary et al. [35] was used. In their work, they code the responses to the first question into 6 types of evaluation (e.g., "Performance of others", "Percentage of shots made", and "Use of the correct method or strategy"), on which participants relied on to assess their performance. For the second question, responses were encoded into 10 types of attribution (e.g., "Specific technique", "General technique", and "Confidence/ability"). As for the third and final question, 9 types of adaptation are used, and they are the same as the types of attribution, bar the "Confidence/ability" type. The same coding scheme was used by Kleinman et al. [50], who adapted it to video game performance (League of Legends) by slightly adjusting a few of the definitions to fit the context (i.e., from basketball to video games). This system was used qualitatively in mixed methods studies and for comparisons between different groups. For the purposes of this thesis, the same coding scheme -as it was used for performance in League of Legends- is leveraged to facilitate the qualitative analysis and discussion of the novel Prosocial System's self-reflection component (RQ1), based on the responses given by the participants. Table 1 and Table 2 contain the types and their definitions for evaluation and attribution-adaptation, respectively.

4.11 Participants

The user study was conducted with 16 participants, 2 women and 14 men, aged between 18 to 35 ($M = 25.31$, $SD = 4.26$), who played one of three available games over two sessions. Overwatch 2 had the majority of participating players with 12, while both Rainbow Six: Siege and Helldivers II had 2 each (Figure 21). In general, it became clear that members of the Overwatch community were by far the most excited and willing to take part in the study. Participants were asked whether they felt experienced in the game of their choice, on a Likert scale ranging from 1 (Strongly Disagree) to 7 (Strongly Agree). Most participants reported a high amount of experience, something that was corroborated by their subsequent gameplay (Table 3). The average was above the midpoint ($M = 5.56$) and the Std. Deviation was relatively low ($SD = 1.57$). Only 3 participants indicated an experience below the midpoint (i.e., 4), with 2 selecting an experience of 3, and 1 an experience of 2. The latter was uncertain about how to answer, as they felt they did not wish to report a high experience, despite having multiple years of experience and performing well during the gameplay session, so this particular value may not be as representative as the others.

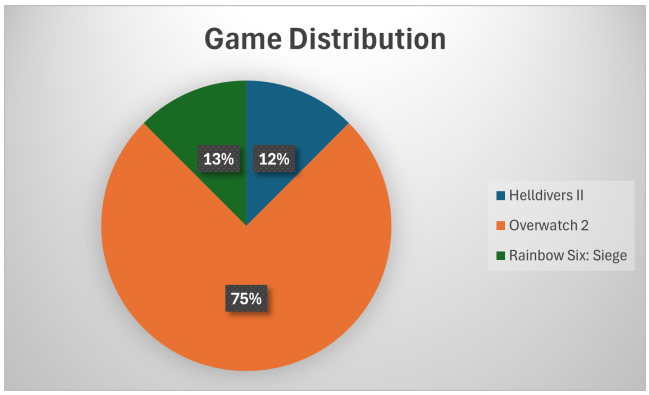

Figure 21: The distribution of participants among the games they selected. Overwatch 2 had 12 participants, Rainbow Six: Siege had 2, and Helldivers II had 2.

5 Results

This section presents the results of the study that was designed to answer the research questions posed by this thesis. It is organized in a way that follows the experiment protocol described in Section 4, starting from the questionnaire of the first session (i.e., the Hexad Scale), proceeding to the second session, and ending with the system interaction. The latter is presented last, as a lot of the relevant data stems from the final questionnaire at the end of the second session. These are followed by an evaluation of the various components of the Prosocial System, an assessment of the system's effect on prosocial behavior, and a correlation analysis between the

Table 1: SRL-based Evaluation types for the participants' responses to the assessment of their prosocial performance during the first gameplay session. These definitions were adapted from Cleary et al. [35] and Kleinman et al. [50].

Type	Definition
Performance of Others	Referred to participant responses in which they evaluated their performance by comparing it to community standards, such as how many and what types of prosocial actions are expected from you during a typical match, i.e., "I think I did well, but I could have used the ping system more for indicating danger, like others did".
Participant's Scores	Referred to participant responses in which they used their in-game scores or Prosocial System scores (e.g., how many healing points they had on Overwatch, how many social points they collected, and how high their Prosocial System rank was) as the basis of their evaluation, i.e., "I think it was great, since I got 8 social points just for stimming teammates".
Correct Method or Strategy	Referred to participant responses in which they evaluated their performance based on their execution of specific prosocial actions, especially ones that helped in-game performance as well (e.g., "Stimming Players that are Injured" and "Team Reloading" in Helldivers II), i.e., "Went well. I feel like I reported by Ultimate to coordinate with others well".
Improvement during Gameplay	Referred to participant responses in which they based their evaluation on perceptions of improvement over the course of the gameplay session (either during a match or between two matches), i.e., "I feel like I did not do well in the first match of the two because the team did not have any coordination. But I improved in the second match because of the team synergy".
Other Factors	Referred to participant responses in which they based their evaluation of their performance on anything not covered by the other codes.
They don't Know	Referred to participant responses in which they did not elaborate on how they were evaluating themselves or did not know if they were.

Table 2: SRL-based Attribution and adaptation types for the participants' responses to the justification of their shortcomings in their prosocial performance during the first gameplay session. The types for adaptation do not include the "Confidence/ability" type. These definitions were adapted from Cleary et al. [35] and Kleinman et al. [50].

Type	Definition
Specific Technique	Discussions of failure to execute or need to change/improve specific prosocial actions.
General Technique	Discussions of failure to execute or need to change/improve general in-game actions or types of prosocial actions (e.g., communicate more).
Confidence/Ability	Discussions of lack of confidence or ability to perform the task (only applied to attribution).
Focus/Concentration	Discussions of inability to or need to change/improve their ability to remain focused, concentrate on the task, or prioritize objectives.
Effort	Discussions of not trying hard enough or needing to try harder.
Practice	Discussions of lack of practice or familiarity or needing to practice more.
Rhythm	Discussions of issues with or need to change/improve patience or timing of execution.
Distractions	Discussions of external stimuli that interfered or needing to ignore these.
Other	Anything not covered by the other codes.
They don't Know	If they do not have an answer.

Hexad Scale user types and the metrics concerning the prosocial performance.

5.1 Hexad Scale

In spite of the number of participants being on the lower end of what would be ideal for a reliability analysis, one was conducted for each of the six subscales. By calculating Cronbach's α for the participants ($n = 16$), it was shown that the Philanthropist ($\alpha = 0.771$), Socializer ($\alpha = 0.746$), and Player ($\alpha = 0.708$) subscales had acceptable reliability. The remaining subscales (i.e., Free Spirit, Achiever, and Disruptor) had indicated low reliability ($\alpha < 0.4$), but this is difficult to interpret without more participants.

In terms of the distribution of the participants into user types, Philanthropists were the plurality (5 participants), followed closely

Table 3: Participant experience on selected game.

Experience Value	Frequency
Experience of 7	6
Experience of 6	4
Experience of 5	3
Experience of 3	2
Experience of 2	1

Table 4: Hexad Scale means and Std. Deviations for each subscale.

Subscale	Mean	SD
Philanthropists	23.18	3.52
Players	21.50	3.82
Socializers	22.93	3.25
Achievers	22.37	2.72
Free Spirits	21.62	2.55
Disruptors	16.68	3.45

by Players (4 participants) and Socializers (3 participants). In addition, 2 were assigned the Achiever type, while only 1 participant was considered a Free Spirit (Figure 22). Interestingly, 1 participant was found to have two equal scores that were the maximum. That designated them as a Socializer - Free Spirit user type. No Disruptors were present in our sample. This wide variety of user types is a significant advantage when it comes to evaluating the Prosocial System and its effect. Despite not having an ideally higher number of participants, the sample includes players of different types and motivations.

Notably, the subscales with the most participants assigned to them were the ones with acceptable reliability, while the ones that were assigned less frequently had low reliability scores. The mean values of all subscale scores except the Disruptor were between 21 and 23 ($M = 21.38$ for all subscales). The Disruptor subscale was the outlier, since it had a significantly lower mean ($M = 16.68$ and $SD = 3.45$), which showcases why no participant was deemed to be of the Disruptor user type. Furthermore, with a mean Std. Deviation of 3.22 (for all subscales), scores were relatively similar among participants for all subscales. Oftentimes, the assigned user type, which was chosen based on the maximum score, was only the maximum by 1 unit. On average, the difference between the maximum score and the second-highest score was 2.56, with a Std. Deviation of 1.75. This suggests that it is difficult to associate a user of a gamification system with a single user type in a concrete way, at least for our set of participants. It does, however, still provide us with valuable information on the users' motivations. Table 4 contains the means and Std. Deviations for all subscales.

5.2 First Gameplay Session

This subsection demonstrates the social points gathered by the participants in each of the three supported games during the first session. Gameplay times are also reported for each game. Table

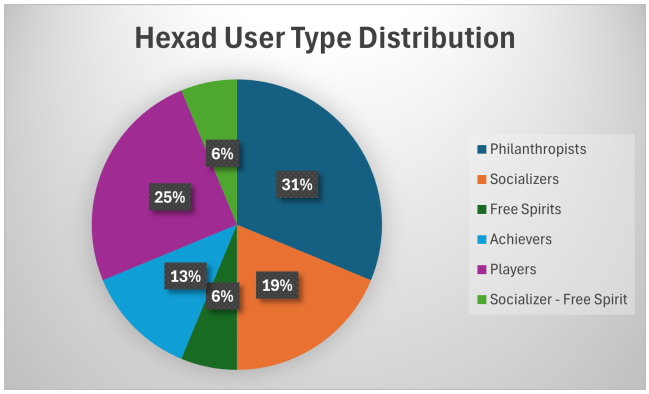

Figure 22: The distribution of participants among their user types according to the Hexad Scale. There were 5 Philanthropists, 4 Players, 3 Socializers, 2 Achievers, 1 Free Spirit, and 1 Socializer - Free Spirit.

Table 5: Means and Std. Deviations for social points and gameplay time (in minutes) for each available game during the first gameplay session.

Game	Mean *SP*	SD *SP*	Mean *Time*	SD *Time*
Helldivers II	16.00	7.07	43.00	2.82
Overwatch 2	15.83	12.17	19.41	4.10
R6: Siege	7.50	7.77	23.00	8.48

5 contains the means and Std. Deviations for social points and gameplay time for each game.

5.2.1 Helldivers II. The 2 participants that chose this game for the user study played an average of 43 minutes ($SD = 2.82$), the longest duration among all three available games. A single mission was sufficient to fit the target gameplay time, and both participants opted for a high difficulty (that they were comfortable with) during their session. Helldivers II supports 18 types of prosocial actions (see Section 4), some of which are mission and in-game randomness dependent, as the game generates missions and the included objectives -optional and not- procedurally. During the first gameplay session, 6 of the available prosocial actions were performed by at least one player (i.e., 33%). The ones not performed were: "Hug or Other Coop Emote", "Rock Paper Scissors Emote", "Affirmative Reaction", "Pinging Another Player's Gear", "Opening a Buddy Bunker" (players did not encounter one), "Cooperating to Align a Radar" (objective did not appear during gameplay), "Bringing High-level Equipment for Lower-level Players to Try Out" (players were at very high levels, since the difficulty chosen was high), "Rerouting e710 Objective" (one player encountered one, but they handled it themselves), "Using the Supply Pack to Resupply your Teammates on the Move", "Team Reloading", "Saluting Together on Extraction", and "Praising Teammates on Voice-Chat" (no one used voice-chat during gameplay). The most popular action was "Using the Ping System", with both players performing it 9 times each. The average total social points collected was 16 ($SD = 7.07$), the highest average

value over all three supported games for the first session. Table 6 includes the means and Std. Deviations for all prosocial actions performed during this first session.

5.2.2 Overwatch 2. This game had the most participants by a significant amount, with 12 opting to play it for the user study. They played an average of 19.41 minutes ($SD = 4.10$), which required 2 matches to reach, and was the shortest duration on average among the three supported games. Some players chose to play in competitive modes, but most played casual modes. Overwatch 2 supports 14 types of prosocial actions (see Section 4), with only one being hero-dependent. None are objective or game mode dependent. However, it should be noted that the "Endorsed Teammates" action was inherently limited to a maximum of 4 possible points, as 2 matches were played by each participant, each allowing up to 2 endorsements. During the first gameplay session, 11 of the available prosocial actions were performed by at least one player (i.e., 79%). The only actions that were not performed were "Congratulating an Opponent that Killed You", "Headpatting People as Ramattra" (nobody played that hero), and "Praising Teammates on Voice-Chat" (nobody used voice-chat during the first session). "Hello, Thank you, You're welcome Voice Lines" was the most highly performed action, with an average of 4.08 ($SD = 4.66$). The average total social points gathered was 15.83 ($SD = 12.17$). Table 7 contains the means and Std. Deviations for all prosocial actions that were performed during the first session.

5.2.3 Rainbow Six: Siege. Like Helldivers II, this game had 2 participants that chose to play it for the user study. The gameplay session lasted an average of 23 minutes ($SD = 8.48$), which took both players 2 matches to reach. One participant preferred to play in ranked mode, while the other chose casual mode. This fact is quite significant, since players are more likely to engage with each other through voice-chat on ranked mode, which is what happened in this case as well. In this first session, 5 of the available 10 prosocial actions were performed by at least one player (i.e., 50%). The actions that were not performed were: "Coordinating Rotation Holes", "Congratulating an Opponent that Killed You", "Swapping Operators", "Placing Valkyrie Cameras on Request" (no participant played the Valkyrie operator in this session), and "Non-verbal Emoting and Communication". The most popular prosocial action was "Praising Teammates on Voice-Chat", which was performed by just one participant 6 times. Among both players, "Typing GLHF, GG, or WP" was the most performed, with an average of 1.5 ($SD = 0.70$). The average total social points collected was 7.5 ($SD = 7.77$), the lowest among the three supported games for this first session. Table 8 contains the means and Std. Deviations for all prosocial actions that were performed during the first session.

5.3 Second Gameplay Session

This subsection demonstrates the social points gathered by the participants in each of the three supported games in the course of the second session. Gameplay times are also reported for each game. Table 9 contains the means and Std. Deviations for social points and gameplay time for each game. The same structure as the relevant subsection for the first session is followed.

5.3.1 Helldivers II. The 2 participants that chose this game for the user study played an average of 44 minutes ($SD = 5.65$). Similar to the first session, a single mission was sufficient to reach the target gameplay time, and both participants opted for a high game difficulty. During the second gameplay session, 11 of the available prosocial actions (total of 18) were performed by at least one player (i.e., 61%). The ones not performed were: "Pinging Another Player's Gear", "Cooperating to Align a Radar" (objective did not appear during gameplay), "Bringing High-level Equipment for Lower-level Players to Try Out" (players were at very high levels, since the difficulty chosen was high), "Stimming Players that are Injured", "Using the Supply Pack to Resupply your Teammates on the Move", "Team Reloading", and "Praising Teammates on Voice-Chat" (no one used voice-chat during gameplay). The most popular action was "Using the Ping System", with players performing it an average of 16 times ($SD = 1.41$). The average total social points collected was 25.50 ($SD = 3.53$). Table 6 includes the means and Std. Deviations for all prosocial actions performed during this second session.

5.3.2 Overwatch 2. In the second gameplay session, the 12 participants opting for this game played an average of 20.83 minutes ($SD = 2.91$), which again required 2 matches to reach. Those who chose to play in competitive modes in the first session did the same for the second. The ones that preferred casual modes chose similarly. During the second gameplay session, 13 of the available prosocial actions (total of 14) were performed by at least one player (i.e., 93%). The only action that was not performed was "Headpatting People as Ramattra", despite one participant commenting that they would like to try it, but ultimately did not pick the Ramattra hero. "Hello, Thank you, You're welcome Voice Lines" was again the most highly performed action, with an average of 10.75 ($SD = 6.82$). The average total social points gathered was 33.75 ($SD = 14.32$), the highest average value over all three supported games in the second session. Table 7 contains the means and Std. Deviations for all prosocial actions that were performed during the second session.

5.3.3 Rainbow Six: Siege. For the second gameplay session, the 2 participants that selected Rainbow Six: Siege for the user study played an average of 22.50 minutes ($SD = 3.53$), which again took both players 2 matches to reach. The participant that preferred to play in ranked mode during the first session made the same choice in the second, while the other chose casual mode again. During the second session, 8 of the available prosocial actions (total of 10) were performed by at least one player (i.e., 80%). The two actions that were not performed were: "Coordinating Rotation Holes" and "Non-verbal Emoting and Communication". The most popular prosocial action was "Praising Teammates on Voice-Chat", which was performed by just one participant 18 times. Among both players, "Using the Ping System" was the most performed, with an average of 4.5 ($SD = 0.70$). The average total social points collected was 19.50 ($SD = 17.67$), remaining the lowest among the three supported games for the second session as well. Table 8 contains the means and Std. Deviations for all prosocial actions that were performed during the second session.

Table 6: Helldivers II means and Std. Deviations for each prosocial action that was performed at least once either during the first or second gameplay session. The ones that were not performed are not included to save space. Mean values of 0.50 usually indicate that only one of the players performed the actions a single time.

Action	*First Session* Mean	*First Session* SD	*Second Session* Mean	*Second Session* SD
Hug or Other Coop Emote	0.00	0.00	0.50	0.70
Using the Ping System	9.00	0.00	16.00	1.41
Rock Paper Scissors Emote	0.00	0.00	0.50	0.70
Affirmative Reaction	0.00	0.00	0.50	0.70
Opening a Buddy Bunker	0.00	0.00	0.50	0.70
Rerouting e710 Objective	0.00	0.00	0.50	0.70
Unlocking the ICBM Objective	0.50	0.70	0.50	0.70
Stimming Players that are Injured	4.00	5.65	0.00	0.00
Saluting Together on Extraction	0.00	0.00	1.50	0.70
Sharing Equipment Requested by Other Players	1.50	0.70	1.50	2.12
"Thank You" Emote-Wheel Action	0.50	0.70	1.50	2.12
Praising Teammates on Text-Chat	0.50	0.70	2.00	1.41

Table 7: Overwatch 2 means and Std. Deviations for each prosocial action that was performed at least once either during the first or second gameplay session. The ones that were not performed are not included to save space. The low mean value for "Voice-Chat" is due to only a single participant performing it.

Action	*First Session* Mean	*First Session* SD	*Second Session* Mean	*Second Session* SD
Endorsed Teammates	2.41	1.56	2.75	1.76
Swapping Heroes to Synergize with the Team	0.91	0.90	1.25	1.13
Typing GLHF, GG, OR WP	0.50	0.90	1.41	0.90
Congratulating an Opponent that Killed You	0.00	0.00	0.33	0.77
"Hello", "Thank you", "You're welcome" Voice Lines	4.08	4.66	10.75	6.82
In-Game Non-Verbal Communication	1.33	0.88	1.91	0.79
Emoting with the Whole Team	0.50	1.00	1.16	1.46
Using the Ping System	2.58	3.77	6.75	6.78
Cooperative Spray System	0.16	0.38	0.25	0.45
Reporting Ultimate Status	1.33	2.26	3.08	4.29
Protecting a Vulnerable Teammate	1.66	2.46	2.25	3.72
Praising Teammates on Voice-Chat	0.00	0.00	0.08	0.28
Praising Teammates on Text-Chat	0.33	1.15	1.75	2.83

Table 8: Rainbow Six: Siege means and Std. Deviations for each prosocial action that was performed at least once either during the first or second gameplay session. The ones that were not performed are not included to save space.

Action	*First Session* Mean	*First Session* SD	*Second Session* Mean	*Second Session* SD
Encouraging Last Teammate Standing	1.00	1.41	0.50	0.70
Typing GLHF, GG, or WP	1.50	0.70	1.50	0.70
Congratulate Enemy	0.00	0.00	0.50	0.70
Swapping Operators	0.00	0.00	1.00	1.41
Placing Valkyrie Cameras on Request	0.00	0.00	0.50	0.70
Using the Ping System	1.00	1.41	4.50	0.70
Praising Teammates on Voice-Chat	3.00	4.24	9.00	12.72
Praising Teammates on Text-Chat	1.00	1.41	2.00	2.82

5.4 Interaction with the Prosocial System

The evaluation of the interaction with the proposed system was twofold: the system's usability was assessed using the validated SUS scale, and the SRL coded responses (see Section 4) were used to gain insight into how participants self-reflected on their prosocial performance from the first session of gameplay.

Table 9: Means and Std. Deviations for social points and gameplay time (in minutes) for each available game during the second gameplay session.

Game	Mean *SP*	SD *SP*	Mean *Time*	SD *Time*
Helldivers II	25.50	3.53	44.00	5.65
Overwatch 2	33.75	14.32	20.83	2.91
R6: Siege	19.50	17.67	22.50	3.53

5.4.1 System Usability. The SUS items about the novel Prosocial System were the first part of the second -and last- questionnaire, which was given to participants at the end of the second session. A reliability analysis was performed, by calculating Cronbach's α for the participants ($n = 16$). The SUS items had acceptable reliability ($\alpha = 0.828$). The SUS scores were calculated -according to the relevant guidelines- from the answers of each participant. The overall scores were high on the SUS scale, with an average of 82.96 ($SD = 11.03$). According to the scale, this gives the proposed system a usability grade of A [29], the highest grade (given to scores ≥ 80.30). It is worth noting that one of the participants likely answered incorrectly to some of the negatively loaded questions opposite to their answers on the positive loaded ones, with a final outlier score of 55, but this was not an issue. Overall, both the SUS score and general comments made by the participants indicated that the system's design and development were successful in their goal of producing a user-friendly system.

5.4.2 Self-Regulated Learning Types. To answer RQ1, a qualitative analysis is used to assess how participants self-reflected on their prosocial performance, the responses to the three SRL questions mentioned in Section 4 for evaluation, attribution, and adaptation, were recorded. Using the coding scheme used by several previous works (see Table 1 and Table 2), the responses were categorized to provide a more holistic perspective of their distribution. Tables 10, 11, and 12 present the frequency for each type of response. Many participants cited the performance generally expected of them as a basis of evaluation (i.e., "I could use the ping system more like my teammates did."), or specified a gameplay function that they felt they performed (or did not perform) adequately (i.e., "It went well, especially when it comes to emoting, spraying, and saying hello to all teammates."), while some considered some other factor. The latter was the evaluation type preferred by those that indicated they were satisfied by their prosocial performance. In terms of the attribution phase, 4 answered that they could not pinpoint anything they did not do sufficiently (i.e., "They don't Know" types), while the majority determined that some specific (e.g., pinging for danger) or general technique (e.g., overall communication) was not performed enough by them during the first session. Lastly, for the adaptation phase, the majority of participants (i.e., 11 out of 16) reported specific prosocial actions that they often learned from the Prosocial System as their goal for the second gameplay session, while a few focused on general aspects, like communication and coordination. It is evident that the manner in which the novel Prosocial System visualizes and describes prosocial actions helped participants expand the range of prosocial actions in their toolbox.

Table 10: Frequency of each type of evaluation response given during the interaction with the Prosocial System.

Type of Evaluation	Frequency
Performance of Others	5
Participant's Scores	1
Correct Method or Strategy	4
Improvement during Gameplay	1
Other Factors	5

Table 11: Frequency of each type of attribution response given during the interaction with the Prosocial System.

Type of Attribution	Frequency
Specific Technique	4
General Technique	4
Confidence/Ability	1
Focus/Concentration	1
Effort	1
Other	1
They don't Know	4

Table 12: Frequency of each type of adaptation response given during the interaction with the Prosocial System.

Type of Adaptation	Frequency
Specific Technique	11
General Technique	4
Focus/Concentration	1

5.5 Prosocial System Components

The last source of results comes from the set of 7 questions about the Prosocial System's functionality. These custom questions were presented to participants after the SUS items as part of the second session questionnaire, in the form of a Likert scale ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). They aim to evaluate the components of the system, as well as the general attitude participants had towards the system and its potential integration into their favorite video games. Table 13 presents the items along with the responses given by the participants. The responses to all items had high scores (i.e., between 4.18 and 4.50, and a maximum $SD = 0.91$). Item 1 shows that the novel Prosocial System helped participants understand prosocial actions better and discover ones they did not know about. Responses to items 2, 3, 5, and 6 indicate that all components of the Prosocial System encouraged participants to be more prosocial during their second gameplay session, suggesting that the design and development of the system were successful in their goal to encourage players to be more prosocial after their interaction with it. According to item 4, the Prosocial System accomplished its goal to help players self-reflect on their prosocial performance. Lastly, item 7 shows that a complete integration of the system into the participants' favorite games would further encourage them to be more prosocial in those games.

Table 13: The custom questions about the Prosocial System, as well as the means and Std. Deviations of the participants' responses (Likert scale, ranging from 1 to 5).

Item	Mean	SD
1. Using the Prosocial System as a whole, allowed me to better understand what constitutes a prosocial action that is possible in the game of my choice.	4.18	0.75
2. Using the Prosocial System as a whole, encouraged me to be more prosocial and perform more prosocial actions in my second gameplay session.	4.18	0.54
3. The score and ranking functionality (SP and rank) of the Prosocial System encouraged me to be more prosocial and perform more prosocial actions in my second gameplay session.	4.25	0.68
4. The Prosocial System helped me self reflect on my prosocial performance of my first gameplay session.	4.18	0.91
5. Self-reflection helped encourage me to be more prosocial and perform more prosocial actions in my second gameplay session.	4.31	0.47
6. A real reward functionality in the Prosocial System would encourage me to be more prosocial and perform more prosocial actions.	4.50	0.81
7. If the Prosocial System was fully integrated into my favorite games, I would be encouraged to be more prosocial and perform more prosocial actions.	4.50	0.51

Table 14: Mean and Std. Deviation of the social points collected by participants in the second session as a percentage of the first session (i.e., 100% are the points collected in the first session, and 200% in the second session would indicate a doubling of the points), for each of the three available games.

Game	SP Percentage Mean	SP Percentage SD
Helldivers II	171.21%	53.56%
Overwatch 2	282.74%	134.92%
Rainbow Six: Siege	298.08%	73.43%

5.6 Effect on Prosocial Behavior

The first factor in answering RQ2, is the change in the social points for each game between the two sessions. Because the points collected -in either session- in each game differ (i.e., Overwatch 2 players generally had more social points than players of the other two games), the best way to illustrate the change between sessions overall, is to report the second session's points in terms of percentages of the first sessions (e.g., doubling the points would mean a percentage of 200% in the second session). Table 14 presents the average percentages for each of the three games, where the significant universal increase in points is evident.

While this information clearly indicates a sheer increase in points, a statistical analysis should also take place to confirm the findings. For this purpose, a paired t-test was conducted to compare the social points collected by the participants in the two sessions. The test showed a statistically significant difference ($t(15) = -7.3672$, $p = 2.34e^{-06}$) between the two sessions (mean difference of 16.12 points), which confirms that the second session of gameplay led to more points compared to the first (Figure 23). In fact, every participant gathered more points in the second session for every available game.

Some prosocial actions varied significantly in how many times they were performed between the two sessions. For Helldivers II, "Stimming Players that are Injured" was not performed at all during the second session, while it was performed 8 times by a single participant in the first. The reason is that the same participant switched to an entirely different play-style for the second session, which did not allow them to perform this particular action. On the other hand, "Using the Ping System" was performed 7 times more on average ($SD = 1.41$) in the second session. The rest of the actions only differed negligibly, if at all. For Overwatch 2, "Using the Ping System" increased by 6.67 on average ($SD = 5.88$) and "Hello, Thank you, You're welcome Voice Lines" increased by 4.17 on average ($SD = 3.97$), with the rest of the actions increasing slightly between the sessions. "Praising Teammates on Voice-Chat" was also performed for the first time in the second session. In fact, one participant used voice-chat to praise a teammate and shield them from the verbal abuse of another teammate who was being toxic. The participant did this despite not using voice-chat in the first session and commenting that they usually avoid using voice-chat at all. For Rainbow Six: Siege, the "Using the Ping System" action was performed 3.50 times more on average ($SD = 0.70$) during the second session, while the "Praising Teammates on Voice-Chat" was performed 6 times more on average ($SD = 8.48$). The rest of the actions increased by a slight margin, except "Encouraging Last Teammate Standing" which was performed 1 time in the second session instead of 2 times in the first.

As mentioned before, the percentages of available prosocial actions that were performed at least once in each session were kept track of and analyzed. On average, among the three games, the percentage of available prosocial actions performed increased by 24% in the second session (Table 15). To better validate the aforementioned result, a paired t-test was conducted to compare the percentage of available prosocial actions in each game that were performed at least once in each of the two sessions. This test showed a statistically significant ($t(2) = -4.8925$, $p = 0.0393$) difference between the means of the two sessions (Figure 24). Moreover, in the two sessions combined, 80% of prosocial actions were performed for Rainbow Six: Siege, 93% for Overwatch 2, and 67% for Helldivers II.

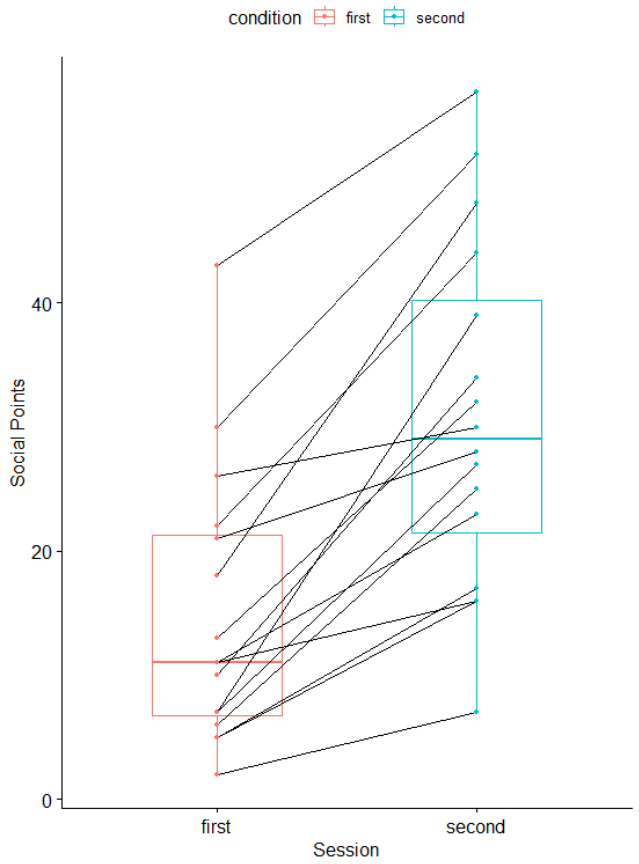

Figure 23: The social points that were collected between each session, plotted in pairs for each participant. The lines clearly indicate the increase in social points gathered in the second session for all participants.

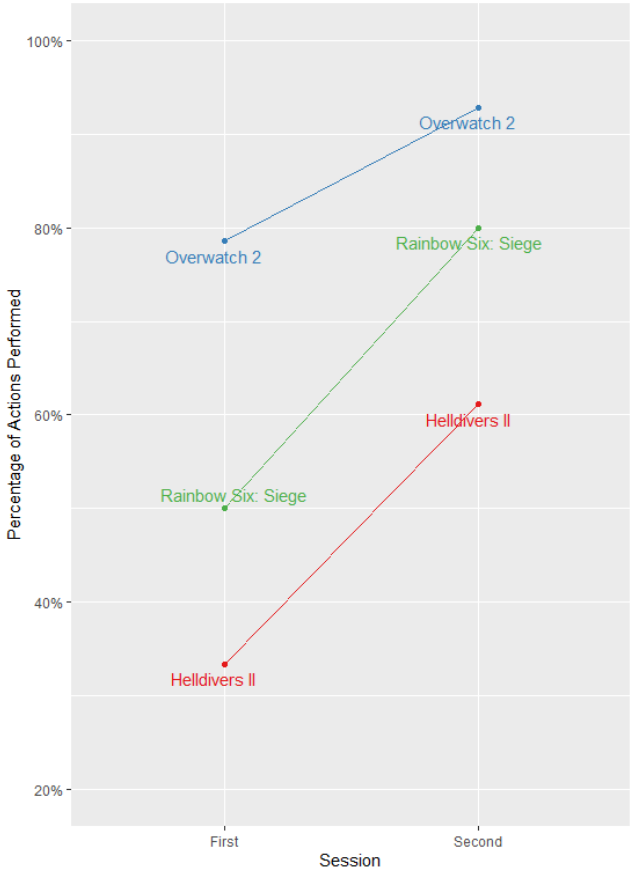

Figure 24: The percentage of available prosocial actions that were performed between each session, plotted in pairs for the three available games.

Table 15: Percentages of available prosocial actions performed at least once for each available game during the first and second gameplay sessions.

Game	% in First Session	% in Second Session
Helldivers II	33%	61%
Overwatch 2	79%	93%
R6: Siege	50%	80%

5.7 User Types & Prosocial Performance

In an attempt to associate the participants' user type (according to the Hexad scale) with the points they collected, the scores on each of the 6 user types of each participant were used (i.e., in a continuous format), instead of just the type with the maximum score for each participant. Correlation tests (i.e., with a correlation matrix) were conducted between the 6 user type scores, the points gathered in the first session, the points gathered in the second session, the difference in the points between the sessions, and the same difference as a percentage. Figure 25 showcases the correlations, where the ones with $p < 0.05$ are marked. The most interesting relationships are between Philanthropist score and Socializer score, between Philanthropist score and point difference, and between second session points and Player score. The first relationship shows that participants with a high Philanthropist score generally had a high Socializer score ($r = 0.51$, $p = 0.04$), which is validated by the fact that for most Philanthropists (i.e., maximum score was the Philanthropist score) their second-highest score was in the Socializer type. In the Philanthropist score and point difference correlation, it is shown that participants with a high Philanthropist score generally had a high point difference ($r = 0.46$), but it is not statistically significant ($p = 0.07$ when $p \leq 0.05$ is preferred). Finally, the relationship between the second session points and the Player score, indicates that participants with a high Player score generally collected fewer points in the second gameplay session ($r = -0.53$, $p = 0.03$). This observation can be attributed to the fact that Players are primarily encouraged by rewards, and while the proposed system had some mock-up rewards present, it was the least realized component of the system, and thus less likely to encourage prosocial behavior in the context of this user study. Figure 26, Figure 27, and Figure 28 demonstrate the Philanthropist score -

Socializer score, the Philanthropist score - point difference, and the second session points - Player score relationships, respectively.

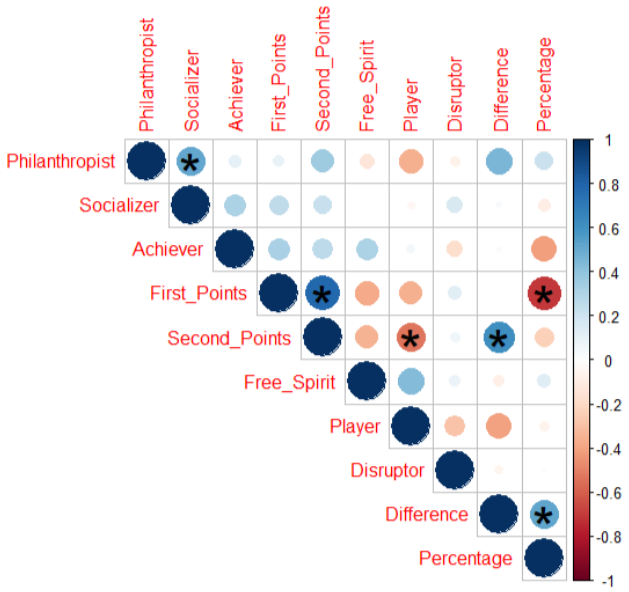

Figure 25: The correlation plot between the 6 user type scores, the points gathered in the first session, the points gathered in the second session, the difference in the points between the sessions, and the same difference as a percentage. The marked correlations indicate a $p < 0.05$.

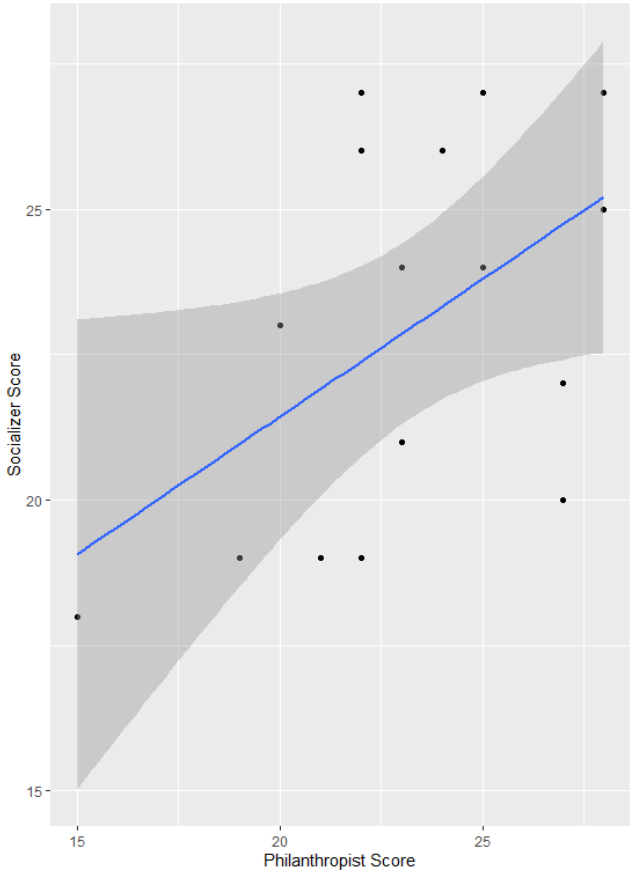

Figure 26: The relationship between the Philanthropist user type score and the Socializer user type score.

6 Discussion

The purpose of this section is to draw conclusions based on the results presented in Section 5. As the aforementioned section clearly demonstrates, most results are of a quantitative nature, together with some equally interesting qualitative ones. All the inferences that follow stem from statistical tests performed on the data, as well as general observations wherever statistical tests were not applicable. To ensure comprehensibility, the structure of this section observes the order of the two research questions, introduced in Section 1. This is followed by a return to the research gap, a write-up on real-world implications, and a deliberation on possible encouragements for participants to increase their points.

6.1 Self-Reflection

RQ1 concerns the effect of the Prosocial System on the player's self-reflection when it comes to their in-game prosocial behavior. The high SUS score of the proposed system (i.e., grade of *A*) is a vital factor that significantly supports the user-friendliness of the system, which by extent suggests that lack of usability is not an element to consider when we evaluate the system's effects. Assessing the exact magnitude of the effect on self-reflection is difficult without a specific baseline of comparison. The only system that provides a small fragment of analogous functionality is the Reputation System found in Rainbow Six: Siege, by tracking positive text-chat messages sent by the player (see Section 2.5). However, since text-chat tracking is but a small fraction of the proposed system's capabilities, and no similar system in terms of purpose exists in commercial (or otherwise) video games, the baseline here is the absence of such functionality. The primary source of quantitative data on self-reflection are the responses to the final questionnaire (Table 13), in particular items 1, 4, and 5. The latter ($M = 4.31$ and $SD = 0.47$) clearly indicates that self-reflection itself was a major source of encouragement for participants to improve their prosocial behavior. Furthermore, the high score on item 4 ($M = 4.18$ and $SD = 0.91$) shows that the proposed system helped enhance the self-reflection process of the participants between the two sessions. Item 1 can also be considered to be part of self-reflection, since the relevant responses ($M = 4.18$ and $SD = 0.75$) suggest that participants better understood what constitutes a prosocial action in the three available games and how those can be performed, because of the novel visualization design and functionality of the proposed system.

In the qualitative analysis, the responses to the SRL questions also indicate the system's impact on self-reflection. Many participants were able to comprehensively explain what they did not do well in terms of prosocial performance, which actions they should

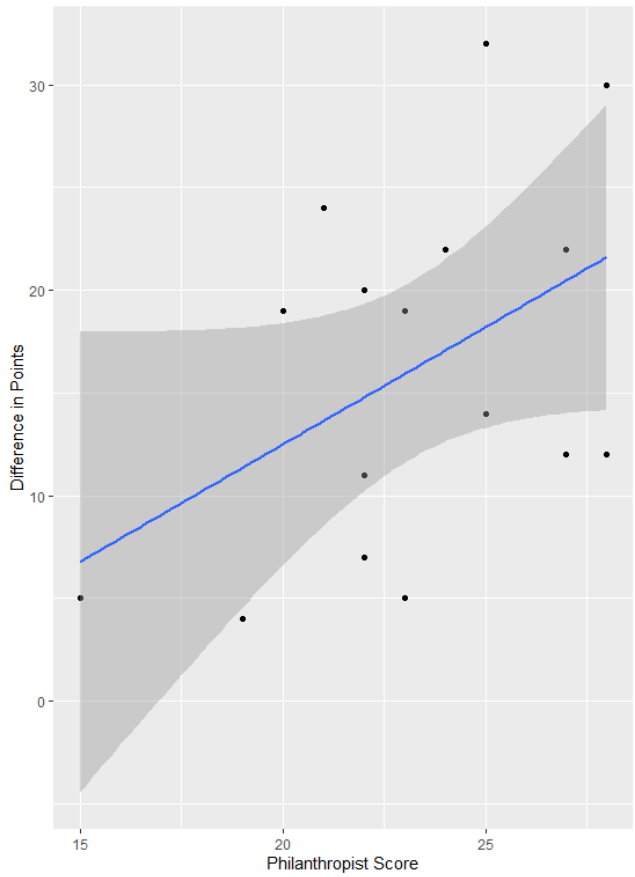

Figure 27: The relationship between the Philanthropist user type score and the difference in social points between the two gameplay sessions.

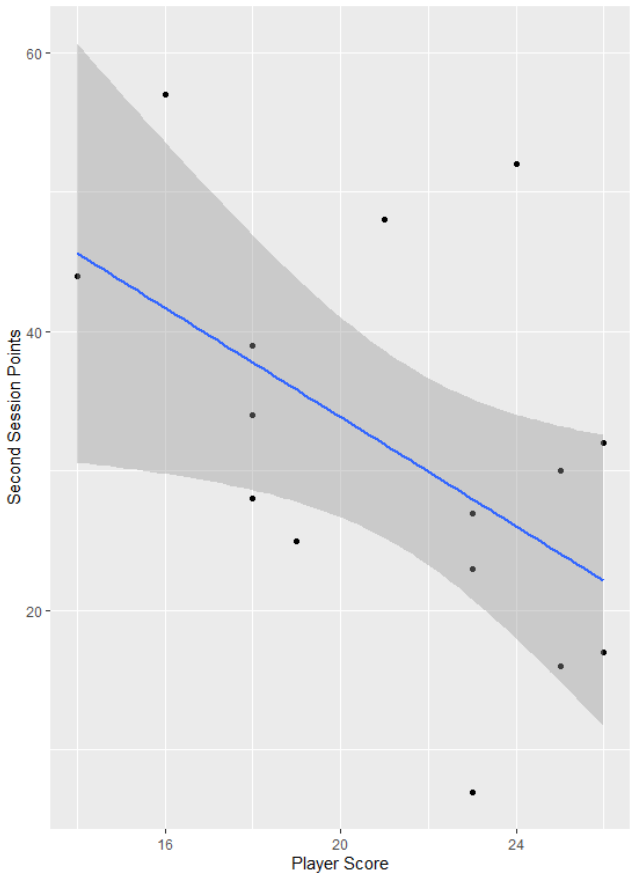

Figure 28: The relationship between the Player user type score and the social points gathered during the second gameplay session.

improve in, and which new actions they wanted to attempt. Without the system tracking those actions and visualizing them, this introspection task would likely be challenging.

Overall, the answer to RQ1 is that the novel Prosocial System of this thesis improved self-reflection on the participants' prosocial behavior during the first session, and supported them in discovering different prosocial actions and how to perform them during the second session. The SRL questions during the course of their interaction with the system were also easy to answer because of the visualization provided by the system.

6.2 Prosocial Behavior

Perhaps the most significant aspect of this thesis concerns RQ2, which pertains to how the interaction with the proposed system affects the user's in-game prosocial behavior. While inferring about the long-term effects of the system is not an explicit part of this master's thesis, the user study that was conducted provides substantial results from which we can draw some interesting conclusions. The goal here was to investigate the change in social points between the two sessions. The study was designed so that the only variable between the two sessions was the interaction with the Prosocial System. Thus, it is a safe assumption to make that no other external factors had a significant effect on the participants' behavior in the second session (e.g., gameplay times did not vary significantly, participants played the same game for both sessions, and the researcher did not explicitly encourage the participants to change their behavior in any way in the second session).

In Helldivers II, the average total SP almost doubled. In Overwatch 2 and Rainbow Six: Siege, the average total SP nearly tripled. It is worth noting, that on Overwatch 2, on which most of the data was gathered, there were participants who quadrupled or even quintupled their points in the second session. Furthermore, even participants with a considerable collection of points in the first session, substantially increased their points in the second, with one participant going from 22 SP to 44 SP, and another from 30 SP to 52 SP. Moreover, as displayed by the paired box plot in Figure 23, the points were increased in the second session for every single participant and every single game.

The second type of data worth discussing is which prosocial actions varied significantly in how many times they were performed between the two sessions. Most actions somewhat increased in frequency, while only a few were performed fewer times or not

at all in the second session. In order to draw concrete conclusions about player preferences when it comes to prosocial actions, more gameplay sessions would need to be conducted. Some actions were, in fact, only possible if a specific play-style was chosen (e.g., equipment selected in Helldivers II, or operators/heroes picked in Rainbow Six:Siege/Overwatch 2). However, the large number of participants in Overwatch 2 provides enough information to make some interesting observations. More specifically, the definitively high scores -across both sessions- in the "Hello, Thank you, "You're welcome Voice Lines" and "Using the Ping System" actions indicate that participants strongly preferred prosocial actions that do not disrupt the flow of gameplay (i.e., get in the way of performance) and that are easier to perform in the gameplay loop. Pinging is an intuitive and useful action that most players are used to. The same is true for the voice lines, since they are facilitated by the game's communication wheel interface, and they do not interrupt what the player is currently doing.

A significant component in the method of this thesis is learning new prosocial actions and being encouraged to perform them, increasing both prosocial behavior in-general and its variety. As mentioned in Section 5, by the end of both gameplay sessions, 80% of available prosocial actions had been performed at least once for Rainbow Six: Siege, 93% for Overwatch 2, and 67% for Helldivers II. For the latter, the result is likely lower because it included the highest number of situational prosocial actions. These observations demonstrate that the Prosocial System was successful in its goal to increase the variety of prosocial behavior through its design. Finally, the fact that such a high percentage of the available prosocial actions were performed in all three games, corroborates the choices made when selecting the prosocial actions themselves.

The significant increase in social points gathered, along with the substantial rise in the percentage of available prosocial actions that were performed during the second gameplay session, establish that the proposed Prosocial System was successful in its ultimate goal to encourage players to be more prosocial in online multiplayer games. Therefore, the answer to RQ2 is that the novel Prosocial System had a positive effect on prosocial behavior exhibited during gameplay.

6.3 Filling the Research Gap

This work succeeds in its goal to assess how a system that tracks the player's in-game prosocial actions and visualizes using a points-ranking scheme impacts the player's self-reflection on their prosocial behavior, and how it affects their subsequent prosocial behavior. By building upon designs found in commercial games [3, 4, 7], the system places the positive actions at the forefront and allows the negative to be handled by the existing systems. This focus enhances the players' ability to self-reflect on their prosocial performance. Furthermore, the results suggest that, like in an academic setting [56], self-reflection in the context of video games can have a significant effect on player behavior. While Kleinman et al. focused on the effect on performance improvement of esports players [50], this work provides similar insights on prosocial performance, successfully addressing the identified research gap. It also establishes the fact that limiting analogous research to a single use case (i.e., a single video game) is generally unnecessary. Given the substantial number of video games of interest, considering modularity and extendability in the design of the research (e.g., the system that is being evaluated) is a major factor in how the results can find real-world use. It is exactly because of this characteristic of modularity that the novel system proposed in this thesis can theoretically be implemented for actual applications outside of research.

6.4 Real-World Implications

Considering the prominence of disruptive behavior in online multiplayer video games (see Section 2.1), the idea of games actively encouraging prosocial behavior in their gameplay loop is one worth considering. Those games provide a framework (or environment) that connects people who are often complete strangers, and allows them to interact with each other in a variety of ways during gameplay. A system similar to the one proposed in this thesis can play the role of promoting positive interactions among those players, while the existing systems that concentrate on negative behavior can focus on detecting and punishing disruptive conduct. This symbiosis is likely to foster positive and social environments in games, and create strongly connected communities. Moreover, because exhibiting prosocial behavior is by itself emotionally rewarding [19, 21, 38], the general happiness of players in real-life can also benefit. Whether the Prosocial System is integrated into the game, or is provided as a separate platform interfacing with multiple games (i.e., like the one proposed here), it is bound to have a positive effect on player experience and motivation to play the game, both of which are severely impacted by disruptive behavior. As to whether promoting prosocial behavior in such a way can combat toxicity directly (e.g., reduce it), it is difficult to know without researching the system's long-term impact and its effects on disruptive behavior. This kind of research is greatly encouraged by the author.

6.5 What Motivates Participants to Be More Prosocial?

All the components of the system were designed specifically to succeed in being encouraging, as they are based on proven designs like ranking systems and self-reflection techniques. It is thus, to some extent, expected that users of the Prosocial System will be motivated to increase their prosocial behavior and their social points. This hypothesis was corroborated by the user study that was conducted. However, what actually motivated the participants of this study to increase their points in the second session is difficult to concretely answer. In a real-world scenario, players have access to their data in the system after every gameplay session. They, therefore, can clearly reflect on whether they collected more points in the last few sessions compared to some previous ones, they can observe their rank increasing, and they can get real in-game rewards as they continue using the system. In contrast, those that participated in the aforementioned user study interacted with the system only once. Consequently, they were able to reflect on their prosocial performance of the first session, and learn how many points they scored and what their rank was, before following through with the second session of gameplay. All participants knew that they would not be shown their data from that second session in the same manner, unless they specifically asked the researcher about it after the conclusion of the experiment, which nobody did. It is

also noteworthy, that the metrics used by the proposed system are not particularly familiar to players. Damage-per-second, healing-per-second, and assists are only a few of the metrics players are utterly acquainted with. On the other hand, the novel system of this thesis introduces prosocial actions as metrics, which players have to get accustomed to. Even so, participants of the user study were able to learn those prosocial actions in a very short period of time. So, why actively try to perform better (i.e., collect more points), if you will never know if you actually succeeded? One possible answer to this question is that players believe that they can somewhat infer whether they performed better or not, which is enough of a motivator for most. People wish to attempt to succeed, even if they cannot be certain about the result. Gamer culture can be another factor supporting this. It could be, then, that just the fact that gamers are aware they are being evaluated on some type of performance in the context of their gameplay, is enough of a stimulus to push them to improve that performance, even if the end-result does not become available to them.

7 Limitations & Future Work

Despite the beneficial results of this work, there are certain limitations to be considered. First, the user study that was conducted and that provided most of the data to draw conclusions from had 16 participants. While that number is not prohibitively low, more participants would enable researchers to collect more data on the novel system's effect on in-game prosocial behavior, as well as to perform more statistical tests and with higher certainty in their results. Second, while plenty of data was gathered on Overwatch 2 from 12 participants, Helldivers II and Rainbow Six: Siege were only played by 2 participants each. Thus, there was a notable lack of information concerning prosocial behavior in those games, making it difficult to consider the relevant results as concretely representative. The lack of further participation in those two games was a consequence of the convenience sampling method and likely the lack of a reward in exchange for participation. Gathering more participants to fill this gap using a more targeted sampling strategy would allow for more statistical tests to be performed, especially when it comes to analysis based on the game chosen by the participant, and would provide a better understanding on the system's effect on these last two games. Lastly, as is the nature of a master's thesis project, this research's user study had a relatively short duration. This is not ideal, as investigating the long-term effects of the use of the proposed Prosocial System would be quite beneficial. A long-term study would have participants play several sessions (over several days) and their data recorded, before using the system. Then, the effects of the system would be investigated by providing the participants with their data on the system for an equal number of sessions.

In the future, to corroborate these findings, and further delve into the impact of such a system on prosocial behavior in online multiplayer video games, more research is needed. It goes without saying, that more participants would significantly help reinforce the findings of this work. Specifically, targeting a more uniform distribution of the participants among the available games would be of great importance. Another interesting aspect to explore is the evolution of the system's functionality by automating most of the prosocial action tracking. This is not seen as a limitation of the present work, as it does not impact the study in any measurable way (e.g., tracking the actions manually is trivial for someone who is familiar with the prosocial actions available in the game being played). However, extending this study to a significant number of more participants would greatly benefit from a percentage of automation to ease the workload of the researchers. It should be noted, though, that automating most of the tracking is no trivial task without some sort of Software Development Kit or Application Programming Interface provided by the developers of the video game of interest. It would also be beneficial to extend the current system so that it provides a unique page for prosocial performance in the last gameplay session, along with the "overall performance of the season" page. This was not needed for the present study, as participants only had one session for which they needed to view their data. Furthermore, while this study proved the modular aspect of its novel Prosocial System, adding more supported games may deliver some additional information. Some popular games like NetEase Games' Marvel Rivals, which is similar in concept and design to Overwatch, or Coffee Stain Studios' cooperative game Satisfactory, should be easy to integrate into the Prosocial System without modifying the existing source code too heavily.

8 Conclusion

This master's thesis presents a novel system that visualizes the player's in-game prosocial behavior in the form of a points-ranking user interface, that works with online multiplayer video games. This Prosocial System's design allows its user to explore the in-game prosocial actions they performed during a gameplay session of a specific game, discover which prosocial actions are available in that game and how they are accomplished, and stimulates and helps them self-reflect on their general prosocial behavior. By leveraging traditional ranking systems that are commonplace in most online multiplayer games in its design, the proposed system aims to encourage its users to be more prosocial in their gameplay sessions by handing out points for each prosocial action performed, assigning ranks to users based on the number of points collected, allowing users to compare their statistics with friends (mock-up for the purposes of this study), and earn unique in-game rewards based on their points on each game (mock-up for the purposes of this study). The system followed a modular design, allowing it to be used with data from virtually any online multiplayer game that supports prosocial actions. The primary goal of this thesis was to investigate the effect of the system on users' self-reflection on their in-game prosocial behavior (RQ1), and its impact on users' encouragement to be more prosocial in their gameplay sessions (RQ2).

A user study with 16 participants was conducted to evaluate the Prosocial System and answer the two posed research questions. The participants played two gameplay sessions of the game of their choosing out of three choices (i.e., Helldivers II, Overwatch 2, and Rainbow Six: Siege). Their prosocial behavior during the first session was used as input for the system, which was then explored by participants, and in the second session the system's effect on their behavior was analyzed. The Hexad scale was used to identify user

types among participants, with the majority identified as Philanthropists, closely followed by Players and Socializers. Correlation tests suggested that the Philanthropist and Socializer scores were closely related, and that Players scored fewer points in the second session compared to their first. The Prosocial System itself was overall rated very highly in terms of its usability (SUS grade of *A*). The percentage of available prosocial actions that were performed increased significantly in the second session for all three games (shown by a paired t-test), showing that during the self-reflection phase the system's visualization and description of such actions helped and stimulated participants to explore different aspects of prosocial behavior in their gameplay. During that self-reflection phase, the participants were asked the three self-regulated learning questions (i.e., evaluation, attribution, and adaptation) to help them self-reflect. Their responses indicated that most participants cited the prosocial performance that is expected of them by others when evaluating their performance, while specific and general prosocial actions were mentioned when talking about attribution and adaptation for the second gameplay session. Moreover, the social points (i.e., how many prosocial actions were performed in total) increased significantly in the second session (shown by a paired t-test), with all participants collecting more points (often more than triple) during their second gameplay, in all three games. This strongly suggests that the novel Prosocial System's components successfully encouraged all participants to be more prosocial in their gameplay, despite the mock-up nature of a few of its components (e.g., the rewards functionality). Finally, the participants rated all components of the system equally in terms of how much they encouraged them to be more prosocial during their gameplay, and reported that if such a system was actually integrated in their favorite video games, they would feel more compelled to be more prosocial in those games.

Overall, the results of this master's thesis are promising and could pave the path for more research on encouraging prosocial behavior in online multiplayer video games. Beyond research, the adaptation of such a system in commercial video games is likely to increase the players' satisfaction and general product success.

Acknowledgments

A heartfelt thank you to the people who supported me in general during this master's thesis. Many thanks to all the participants who made this study possible, taking part out of sheer interest and not because of any reward. Special thanks to Julian Frommel for the supervision and assistance throughout this thesis.

References

[1] 2018. *Endorsements | overwatch-archive.fandom.com.* Retrieved April 9, 2025 from https://overwatch-archive.fandom.com/wiki/Endorsements

[2] 2020. *Commend | Dota 2 Wiki.* Retrieved April 14, 2025 from https://dota2.fandom.com/wiki/Commend

[3] 2020. *DevBlog: Reputation System.* Retrieved April 9, 2025 from https://www.ubisoft.com/en-us/game/rainbow-six/siege/news-updates/1tjFZoNSJycILiLUjMyw7X/devblog-reputation-system

[4] 2022. *Reputation System Beta: Reputation Standing & Fair Play Program Beta.* Retrieved April 9, 2025 from https://www.ubisoft.com/en-us/game/rainbow-six/siege/news-updates/3pf2Hhz7nKS9CiRj5XdHRv/reputation-system-beta-reputation-standing-fair-play-program-beta

[5] 2023. *Conduct Summary | Dota 2 Wiki.* Retrieved April 14, 2025 from https://dota2.fandom.com/wiki/Conduct_Summary

[6] 2023. *Matchmaking | Dota 2 Wiki.* Retrieved April 14, 2025 from https://dota2.fandom.com/wiki/Matchmaking#Behavior_Score_and_Account_Flags

[7] 2023. *Reputation System | Rainbow Six Wiki.* Retrieved April 9, 2025 from https://rainbowsix.fandom.com/wiki/Reputation_System

[8] 2024. *Endorsements | overwatch.fandom.com.* Retrieved April 9, 2025 from https://overwatch.fandom.com/wiki/Endorsements

[9] 2024. *"This is a necessary feature" - Valorant community asks to add an honor system to the game.* Retrieved April 14, 2025 from https://bo3.gg/valorant/news/this-is-a-necessary-feature-valorant-community-asks-to-add-an-honor-system-to-the-game

[10] 2025. *Commend | liquipedia - Dota 2.* Retrieved April 14, 2025 from https://liquipedia.net/dota2/Commend

[11] 2025. *HELLDIVERS 2 | Steam charts.* Retrieved April 18, 2025 from https://steamdb.info/app/553850/charts/

[12] 2025. *Honor | League of Legends Wiki — wiki.leagueoflegends.com.* Retrieved April 8, 2025 from https://wiki.leagueoflegends.com/en-us/Honor

[13] 2025. *HonorHistorical | League of Legends Wiki — wiki.leagueoflegends.com.* Retrieved April 8, 2025 from https://wiki.leagueoflegends.com/en-us/Honor/Historical

[14] 2025. *Overwatch 2 | Steam charts.* Retrieved April 18, 2025 from https://steamdb.info/app/2357570/charts/

[15] 2025. *Player Behavior Summary | liquipedia - Dota 2.* Retrieved April 14, 2025 from https://liquipedia.net/dota2/Player_Behavior_Summary

[16] 2025. *Steam Hardware & Software Survey: March 2025.* Retrieved April 16, 2025 from https://store.steampowered.com/hwsurvey/Steam-Hardware-Software-Survey-Welcome-to-Steam

[17] 2025. *Tom Clancy's Rainbow Six Siege | Steam charts.* Retrieved April 18, 2025 from https://steamdb.info/app/359550/charts/

[18] Sonam Adinolf and Selen Turkay. 2018. Toxic Behaviors in Esports Games: Player Perceptions and Coping Strategies. In *Proceedings of the 2018 Annual Symposium on Computer-Human Interaction in Play Companion Extended Abstracts* (Melbourne, VIC, Australia) *(CHI PLAY '18 Extended Abstracts)*. Association for Computing Machinery, New York, NY, USA, 365–372. https://doi.org/10.1145/3270316.3271545

[19] Lara B Aknin, Tanya Broesch, J Kiley Hamlin, and Julia W Van de Vondervoort. 2015. Prosocial behavior leads to happiness in a small-scale rural society. *Journal of experimental psychology: General* 144, 4 (2015), 788.

[20] Lara B Aknin, Julia W Van de Vondervoort, and J Kiley Hamlin. 2018. Positive feelings reward and promote prosocial behavior. *Current Opinion in Psychology* 20 (2018), 55–59. https://doi.org/10.1016/j.copsyc.2017.08.017 Early Development of prosocial behavior.

[21] Lara B Aknin, Ashley V Whillans, Michael I Norton, and Elizabeth W Dunn. 2019. Happiness and prosocial behavior: An evaluation of the evidence. *World happiness report* 2019 (2019), 67–86.

[22] Donna Armellino, Manish Trivedi, Isabel Law, Narendra Singh, Mary Ellen Schilling, Erfan Hussain, and Bruce Farber. 2013. Replicating changes in hand hygiene in a surgical intensive care unit with remote video auditing and feedback. *American journal of infection control* 41, 10 (2013), 925–927.

[23] Delia Baldassarri and Maria Abascal. 2020. Diversity and prosocial behavior. *Science* 369, 6508 (2020), 1183–1187. https://doi.org/10.1126/science.abb2432 arXiv:https://www.science.org/doi/pdf/10.1126/science.abb2432

[24] Jay Belsky. 2008. Rewards are better than punishment: Here's why. *Psychology Today* (2008).

[25] Nicole A Beres, Julian Frommel, Elizabeth Reid, Regan L Mandryk, and Madison Klarkowski. 2021. Don't You Know That You're Toxic: Normalization of Toxicity in Online Gaming. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 438, 15 pages. https://doi.org/10.1145/3411764.3445157

[26] Jonathan Z. Berman, Emma E. Levine, Alixandra Barasch, and Deborah A. Small. 2015. The Braggart's Dilemma: On the Social Rewards and Penalties of Advertising Prosocial Behavior. *Journal of Marketing Research* 52, 1 (2015), 90–104. https://doi.org/10.1509/jmr.14.0002 arXiv:https://doi.org/10.1509/jmr.14.0002

[27] David Boud, Rosemary Keogh, and David Walker. 2013. *Reflection: Turning experience into learning.* Routledge.

[28] Samuel Bowles and Herbert Gintis. 2011. *A Cooperative Species: Human Reciprocity and Its Evolution.* Princeton University Press, 280. https://doi.org/10.1515/9781400838837

[29] John Brooke. 1995. SUS: A quick and dirty usability scale. *Usability Eval. Ind.* 189 (11 1995).

[30] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[31] Alessandro Canossa, Dmitry Salimov, Ahmad Azadvar, Casper Harteveld, and Georgios Yannakakis. 2021. For Honor, for Toxicity: Detecting Toxic Behavior through Gameplay. *Proc. ACM Hum.-Comput. Interact.* 5, CHI PLAY, Article 253 (Oct. 2021), 29 pages. https://doi.org/10.1145/3474680

[32] Thomas Chesney, Iain Coyne, Brian Logan, and Neil Madden. 2009. Griefing in virtual worlds: causes, casualties and coping strategies. *Information Systems*

Journal 19, 6 (2009), 525–548. https://doi.org/10.1111/j.1365-2575.2009.00330.x arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1365-2575.2009.00330.x

[33] Chipteck. 2024. *Honor FAQ.* Retrieved November 30, 2024 from https://support-leagueoflegends.riotgames.com/hc/en-us/articles/115008474148-Honor-FAQ

[34] Frederick Choi, Charlotte Lambert, Vinay Koshy, Sowmya Pratipati, Tue Do, and Eshwar Chandrasekharan. 2025. Creator Hearts: Investigating the Impact Positive Signals from YouTube Creators in Shaping Comment Section Behavior. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 1212, 18 pages. https://doi.org/10.1145/3706598.3713521

[35] Timothy J Cleary, Barry J Zimmerman, and Tedd Keating. 2006. Training physical education students to self-regulate during basketball free throw practice. *Research Quarterly for Exercise and sport* 77, 2 (2006), 251–262.

[36] Edward L Deci and Richard M Ryan. 2012. Self-determination theory. *Handbook of theories of social psychology* 1, 20 (2012), 416–436.

[37] Edward L Deci and Richard M Ryan. 2013. *Intrinsic motivation and self-determination in human behavior.* Springer Science & Business Media.

[38] Sara Erreygers, Heidi Vandebosch, Ivana Vranjes, Elfi Baillien, and Hans De Witte. 2019. Feel good, do good online? Spillover and crossover effects of happiness on adolescents' online prosocial behavior. *Journal of Happiness Studies* 20, 4 (2019), 1241–1258.

[39] Joey R Fanfarelli. 2018. Expertise in professional overwatch play. *International Journal of Gaming and Computer-Mediated Simulations (IJGCMS)* 10, 1 (2018), 1–22.

[40] Marc Guitart-Masip, Emrah Duzel, Ray Dolan, and Peter Dayan. 2014. Action versus valence in decision making. *Trends in cognitive sciences* 18, 4 (2014), 194–202.

[41] Weszt Hart. 2024. *25 Examples of Prosocial Behavior in Games.* Retrieved November 30, 2024 from https://digitalthrivingplaybook.org/example/25-examples-of-prosocial-behavior-in-games/

[42] Weszt Hart. 2024. *Chapters Method for Prosocial Design in Session-Based Multiplayer Games.* Retrieved November 30, 2024 from https://digitalthrivingplaybook.org/method/chapters-method-for-prosocial-design-in-session-based-multiplayer-games/

[43] Weszt Hart. 2024. *Introduction to Disruptive Behaviors.* Retrieved November 30, 2024 from https://digitalthrivingplaybook.org/big-idea/introduction-to-disruptive-behaviors/

[44] Weszt Hart. 2024. *Prosocial Behaviors | Checklist.* Retrieved November 30, 2024 from https://digitalthrivingplaybook.org/method/prosocial-behaviors-checklist/

[45] Weszt Hart. 2024. *Social Rewards and Other Big Ideas Behind the RPM Method.* Retrieved November 30, 2024 from https://digitalthrivingplaybook.org/big-idea/social-rewards-and-other-big-ideas-behind-the-rpm-method/

[46] Robert A Hinde and Jo Groebel. 1991. *Cooperation and prosocial behaviour.* Cambridge University Press.

[47] Bailey R. House, Joan B. Silk, Joseph Henrich, H. Clark Barrett, Brooke A. Scelza, Adam H. Boyette, Barry S. Hewlett, Richard McElreath, and Stephen Laurence. 2013. Ontogeny of prosocial behavior across diverse societies. *Proceedings of the National Academy of Sciences* 110, 36 (2013), 14586–14591. https://doi.org/10.1073/pnas.1221217110 arXiv:https://www.pnas.org/doi/pdf/10.1073/pnas.1221217110

[48] Blake Kammermann, Selen Türkay, Daniel Johnson, and Stephanie J Tobin. 2024. Virtual Rewards, Real World Impact? Investigating the Prosocial Effects of Videogame Reward Salience. *International Journal of Human–Computer Interaction* (2024), 1–22.

[49] Yoon Jeon Kim and Valerie J. Shute. 2015. The interplay of game elements with psychometric qualities, learning, and enjoyment in game-based assessment. *Computers & Education* 87 (2015), 340–356. https://doi.org/10.1016/j.compedu.2015.07.009

[50] Erica Kleinman, Jason Xu, Johannes Pfau, and Magy Seif El-Nasr. 2024. "Trust the Process": An Exploratory Study of Process Visualizations for Self-Reflection in League of Legends. *Proc. ACM Hum.-Comput. Interact.* 8, CHI PLAY, Article 346 (Oct. 2024), 28 pages. https://doi.org/10.1145/3677111

[51] Bastian Kordyaka, Katharina Jahn, and Bjoern Niehaves. 2020. Towards a unified theory of toxic behavior in video games. *Internet Research* 30, 4 (2020), 1081–1102.

[52] Yubo Kou. 2020. Toxic Behaviors in Team-Based Competitive Gaming: The Case of League of Legends. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play* (Virtual Event, Canada) *(CHI PLAY '20)*. Association for Computing Machinery, New York, NY, USA, 81–92. https://doi.org/10.1145/3410404.3414243

[53] Yubo Kou and Xinning Gui. 2014. Playing with strangers: understanding temporary teams in league of legends. In *Proceedings of the First ACM SIGCHI Annual Symposium on Computer-Human Interaction in Play* (Toronto, Ontario, Canada) *(CHI PLAY '14)*. Association for Computing Machinery, New York, NY, USA, 161–169. https://doi.org/10.1145/2658537.2658538

[54] Haewoon Kwak, Jeremy Blackburn, and Seungyeop Han. 2015. Exploring Cyberbullying and Other Toxic Behavior in Team Competition Online Games. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. Association for Computing Machinery, New York, NY, USA, 3739–3748. https://doi.org/10.1145/2702123.2702529

[55] Charlotte Lambert, Koustuv Saha, and Eshwar Chandrasekharan. 2025. Does Positive Reinforcement Work?: A Quasi-Experimental Study of the Effects of Positive Feedback on Reddit. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 1213, 16 pages. https://doi.org/10.1145/3706598.3713830

[56] Magdeleine DN Lew and Henk G Schmidt. 2011. Self-reflection and academic performance: is there a relationship? *Advances in Health Sciences Education* 16 (2011), 529–545.

[57] Holin Lin and Chuen-Tsai Sun. 2005. The 'white-eyed'player culture: Grief play and construction of deviance in MMORPGs. In *Proceedings of DiGRA 2005 Conference: Changing Views: Worlds in Play.*

[58] De Liu, Xun Li, and Radhika Santhanam. 2013. Digital Games and Beyond: What Happens When Players Compete? *MIS Quarterly* 37, 1 (2013), 111–124. http://www.jstor.org/stable/43825939

[59] Stefania Maggi, Lori J. Irwin, Arjumand Siddiqi, and Clyde Hertzman. 2010. The social determinants of early child development: An overview. *Journal of Paediatrics and Child Health* 46, 11 (2010), 627–635. https://doi.org/10.1111/j.1440-1754.2010.01817.x arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1111/j.1440-1754.2010.01817.x

[60] Charles K Monge and TC O'Brien. 2022. Effects of individual toxic behavior on team performance in League of Legends. *Media Psychology* 25, 1 (2022), 82–105.

[61] Jennifer A Moon. 2013. *A handbook of reflective and experiential learning: Theory and practice.* Routledge.

[62] Joaquim A. M. Neto, Kazuki M. Yokoyama, and Karin Becker. 2017. Studying toxic behavior influence and player chat in an online video game. In *Proceedings of the International Conference on Web Intelligence* (Leipzig, Germany) *(WI '17)*. Association for Computing Machinery, New York, NY, USA, 26–33. https://doi.org/10.1145/3106426.3106452

[63] TEAL Center Fact Sheet No. 2012. Self-regulated learning. *Just Write! Guide* (2012), 29.

[64] Mike C Parent, Teresa D Gobble, and Aaron Rochlen. 2019. Social media behavior, toxic masculinity, and depression. *Psychology of Men & Masculinities* 20, 3 (2019), 277.

[65] Kathie L. Pelletier. 2010. Leader toxicity: An empirical investigation of toxic behavior and rhetoric. *Leadership* 6, 4 (2010), 373–389. https://doi.org/10.1177/1742715010379308 arXiv:https://doi.org/10.1177/1742715010379308

[66] Susanne Poeller, Martin Johannes Dechant, Madison Klarkowski, and Regan L. Mandryk. 2023. Suspecting Sarcasm: How League of Legends Players Dismiss Positive Communication in Toxic Environments. *Proc. ACM Hum.-Comput. Interact.* 7, CHI PLAY, Article 374 (Oct. 2023), 26 pages. https://doi.org/10.1145/3611020

[67] Michael I Posner and Mary K Rothbart. 2000. Developing mechanisms of self-regulation. *Development and psychopathology* 12, 3 (2000), 427–441.

[68] Matt Purslow. 2020. *Rainbow Six Siege's New Two-Year Roadmap Promises Toxicity Tools, House Rework, and Next-Gen.* Retrieved April 9, 2025 from https://www.ign.com/articles/rainbow-six-sieges-new-two-year-roadmap-promises-toxicity-tools-house-rework-and-crazy-events

[69] Richard M Ryan and Edward L Deci. 2000. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American psychologist* 55, 1 (2000), 68.

[70] Amit Sheth, Valerie L. Shalin, and Ugur Kursuncu. 2022. Defining and detecting toxicity on social media: context and knowledge are key. *Neurocomputing* 490 (2022), 312–318. https://doi.org/10.1016/j.neucom.2021.11.095

[71] John Suler. 2004. The online disinhibition effect. *Cyberpsychology & behavior* 7, 3 (2004), 321–326.

[72] Zazie Todd. 2022. *Purr: The Science of Making Your Cat Happy.* Greystone Books Ltd.

[73] Sian Tomkinson and Benn Van Den Ende. 2022. 'thank you for your compliance': Overwatch as a disciplinary system. *Games and culture* 17, 2 (2022), 198–218.

[74] Gustavo F. Tondello, Alberto Mora, Andrzej Marczewski, and Lennart E. Nacke. 2019. Empirical validation of the Gamification User Types Hexad scale in English and Spanish. *International Journal of Human-Computer Studies* 127 (2019), 95–111. https://doi.org/10.1016/j.ijhcs.2018.10.002 Strengthening gamification studies: critical challenges and new opportunities.

[75] Gustavo F. Tondello, Rina R. Wehbe, Lisa Diamond, Marc Busch, Andrzej Marczewski, and Lennart E. Nacke. 2016. The Gamification User Types Hexad Scale. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play* (Austin, Texas, USA) *(CHI PLAY '16)*. Association for Computing Machinery, New York, NY, USA, 229–243. https://doi.org/10.1145/2967934.2968082

[76] Tanja Välisalo and Maria Ruotsalainen. 2019. " I never gave up" engagement with playable characters and esports players of overwatch. In *Proceedings of the 14th International Conference on the Foundations of Digital Games.* 1–6.

[77] Dorothy E Warner and Mike Raiter. 2005. Social Context in Massively-Multiplayer Online Games (MMOGs): : Ethical Questions in Shared Space. *The International Review of Information Ethics* 4 (Dec. 2005), 46–52. https://doi.org/10.29173/irie172

[78] Rafael Wittek and René Bekkers. 2015. Altruism and prosocial behavior, sociology of. *International encyclopedia of the social & behavioral sciences* 1 (2015), 579–583.

[79] Junhui Wu, Shenghua Luan, and Nichola Raihani. 2022. Reward, punishment, and prosocial behavior: Recent developments and implications. *Current Opinion in Psychology* 44 (2022), 117–123. https://doi.org/10.1016/j.copsyc.2021.09.003
[80] Nick Yee. 2016. The Gamer Motivation Profile: What We Learned From 250,000 Gamers. In *Proceedings of the 2016 Annual Symposium on Computer-Human Interaction in Play* (Austin, Texas, USA) *(CHI PLAY '16)*. Association for Computing Machinery, New York, NY, USA, 2. https://doi.org/10.1145/2967934.2967937
[81] Nick Yee, Nicolas Ducheneaut, and Les Nelson. 2012. Online gaming motivations scale: development and validation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Austin, Texas, USA) *(CHI '12)*. Association for Computing Machinery, New York, NY, USA, 2803–2806. https://doi.org/10.1145/2207676.2208681
[82] Greyson K Young. 2022. How much is too much: the difficulties of social media content moderation. *Information & Communications Technology Law* 31, 1 (2022), 1–16.
[83] Barry J Zimmerman. 1998. Developing self-fulfilling cycles of academic regulation: an analysis of exemplary instructional models. (1998).

A Ethics and Privacy Quick Scan

A copy of the filled out Ethics and Privacy Quick Scan follows on the next page.

Response Summary:

Section 1. Research projects involving human participants

P1. Does your project involve human participants? This includes for example use of observation, (online) surveys, interviews, tests, focus groups, and workshops where human participants provide information or data to inform the research. If you are only using existing data sets or publicly available data (e.g. from X, Reddit) without directly recruiting participants, please answer no.

- Yes

Recruitment

P2. Does your project involve participants younger than 16 years of age?

- No

P3. Does your project involve participants with learning or communication difficulties of a severity that may impact their ability to provide informed consent?

- No

P4. Is your project likely to involve participants engaging in illegal activities?

- No

P5. Does your project involve patients?

- No

P6. Does your project involve participants belonging to a vulnerable group, other than those listed above?

- No

P8. Does your project involve participants with whom you have, or are likely to have, a working or professional relationship: for instance, staff or students of the university, professional colleagues, or clients?

- Yes

P9. Is it made clear to potential participants that not participating will in no way impact them (e.g. it will not directly impact their grade in a class)?

- Yes

Informed consent

PC1. Do you have set procedures that you will use for obtaining informed consent prior to collecting data from all participants, including (where appropriate) parental consent for children or consent from legally authorized representatives? (See suggestions for information sheets and consent forms on the website.)

- Yes

PC2. Will you tell participants that their participation is voluntary?

- Yes

PC3. Will you obtain explicit consent for participation?

- Yes

PC4. Will you obtain explicit consent for any sensor readings, eye tracking, photos, audio, and/or video recordings?

- Yes

PC5. Will you tell participants that they may withdraw from the research at any time and for any reason?

- Yes

PC6. Will you give potential participants time to consider participation?

- Yes

PC7. Will you provide participants with an opportunity to ask questions about the research before consenting to take part (e.g. by providing your contact details)?

- Yes

PC8. Does your project involve concealment or deliberate misleading of participants?

- No

Section 2. Data protection, handling, and storage

The General Data Protection Regulation imposes several obligations for the use of **personal data** (defined as any information relating to an identified or identifiable living person) or including the use of personal data in research.

D1. Are you gathering or using personal data (defined as any information relating to an identified or identifiable living person)?

- No

Section 3. Research that may cause harm

Research may cause harm to participants, researchers, the university, or society. This includes when technology has dual-use, and you investigate an innocent use, but your results could be used by others in a harmful way. If you are unsure regarding possible harm to the university or society, please discuss your concerns with the Research Support Office.

H1. Does your project give rise to a realistic risk to the national security of any country?

- No

H2. Does your project give rise to a realistic risk of aiding human rights abuses in any country?

- No

H3. Does your project (and its data) give rise to a realistic risk of damaging the University's reputation? (E.g., bad press coverage, public protest.)

- No

H4. Does your project (and in particular its data) give rise to an increased risk of attack (cyber- or otherwise) against the University? (E.g., from pressure groups.)

- No

H5. Is the data likely to contain material that is indecent, offensive, defamatory, threatening, discriminatory, or extremist?

- No

H6. Does your project give rise to a realistic risk of harm to the researchers?

- No

H7. Is there a realistic risk of any participant experiencing physical or psychological harm or discomfort?

- No

H8. Is there a realistic risk of any participant experiencing a detriment to their interests as a result of participation?

- No

H9. Is there a realistic risk of other types of negative externalities?

- No

Section 4. Conflicts of interest

C1. Is there any potential conflict of interest (e.g. between research funder and researchers or participants and researchers) that may potentially affect the research outcome or the dissemination of research findings?

- No

C2. Is there a direct hierarchical relationship between researchers and participants?

- No

Section 5. Your information.

This last section collects data about you and your project so that we can register that you completed the Ethics and Privacy Quick Scan, sent you (and your supervisor/course coordinator) a summary of what you filled out, and follow up where a fuller ethics review and/or privacy assessment is needed. For details of our legal basis for using personal data and the rights you have over your data please see the University's privacy information. Please see the guidance on the ICS Ethics and Privacy website on what happens on submission.

Z0. Which is your main department?

- Information and Computing Science

Z1. Your full name:

Panagiotis Vrettis

Z2. Your email address:

p.vrettis@students.uu.nl

Z3. In what context will you conduct this research?

- As a student for my master thesis, supervised by::
  Julian Frommel

Z5. Master programme for which you are doing the thesis

- Game and Media Technology

Z6. Email of the course coordinator or supervisor (so that we can inform them that you filled this out and provide them with a summary):

j.frommel@uu.nl

Z7. Email of the moderator (as provided by the coordinator of your thesis project):

j.frommel@uu.nl

Z8. Title of the research project/study for which you filled out this Quick Scan:

Investigating the impact of a ranking & reward system in promoting prosocial behavior in online multiplayer games

Z9. Summary of what you intend to investigate and how you will investigate this (200 words max):

The video game industry has consistently grown to a giant entertainment medium, rivaling the movie, literature, and music industries. In online multiplayer video games, disruptive behavior like toxicity and other harmful conduct have plagued player experience from the introduction of such games. Unfortunately, the few proposed solutions have failed to mitigate, let alone eliminate, the problem. In this master's thesis, I propose a novel approach to alleviate the issue. I will design and develop a custom modular tool that tracks a player's in-game prosocial behavior and ranks them based on points they gain through their prosocial actions. I will evaluate the effect of this tool by conducting a user study, where participants play a game of their choice (Helldivers II, Overwatch 2, or Rainbow Six: Siege), and I record and analyze their prosocial actions. The tool will then be given to the participants for them to self-reflect on their prosocial actions, as well as explore what such actions are available in the chosen game. A second gameplay session will be then conducted and recorded (along with a custom questionnaire that the participant will answer), to investigate whether an observable increase in prosocial behavior occurred.

Z10. In case you encountered warnings in the survey, does supervisor already have ethical approval for a research line that fully covers your project?

- Not applicable

Scoring

- Privacy: 0
- Ethics: 0

B Consent Information Sheet

A copy of the consent information sheet follows on the next page. It observes the informed consent guidelines for online studies of Utrecht University.

Universiteit Utrecht

Welcome to the online questionnaire for the study "Investigating the Impact of a Ranking & Reward System in Promoting Prosocial Behavior in Online Multiplayer Games"

1. Introduction

You are being asked to participate in the scientific research for the master's thesis "**Investigating the Impact of a Ranking & Reward System in Promoting Prosocial Behavior in Online Multiplayer Games**" as part of the Game & Media Technology master's programme of Utrecht University.

2. What is the background and purpose of this study?

The video game industry has consistently grown to a giant entertainment medium, rivaling the movie, literature, and music industries. In online multiplayer video games, disruptive behavior like toxicity and other harmful conduct have plagued player experience from the introduction of such games. Unfortunately, the few proposed solutions have failed to mitigate, let alone eliminate, the problem. In this master's thesis, I propose a novel approach to alleviate the issue. I will design and develop a custom modular tool that tracks a player's in-game prosocial behavior and ranks them based on points they gain through their prosocial actions. I will evaluate the effect of this tool by conducting a user study, where participants play a game of their choice (Helldivers II, Overwatch 2, or Rainbow Six: Siege), and I record and analyze their prosocial actions manually (a complete version would automate this task in real-time).

3. Who will carry out the study?

This study is conducted by Panagiotis Vrettis as part of his master's thesis (p.vrettis@students.uu.nl).

It is supervised by Dr. Julian Frommel (j.frommel@uu.nl).

4. How will the study be carried out?

This study will be conducted in two sessions per participant.

In the first session, the participant will enter a Discord call with the researcher on the decided-upon time. The participant will first sign the informed consent form and answer a few demographic questions online (approx. 5 minutes to complete). Then, the participant will start the game of their choosing (from the three available) share their screen to the researcher and complete a game match (approx. 20 minutes to complete). The researcher will be recording the participant's screen (including audio) during the gameplay session. After gameplay, the researcher will stop the recording, and the participant will stop sharing their screen. Any questions are answered, and a new date will be decided upon for the second session (unless such a date has already been set).

In the second session, the participant will enter a Discord call with the researcher on the decided-upon time. The participant is provided with the Prosocial System loaded with their data and they are requested to share their screen. The researcher will then record the interaction

with the system for analysis. During this time, the participant is requested to explore their data in the system and answer a few questions about their interpretation of the given data (approx. 7 minutes to complete). The participant will then switch to the game of their choice and play another match, while the researcher is recording (approx. 20 minutes to complete). Post gameplay, the recording is stopped, and the participant is asked to complete a few questions about the system's usability and effectiveness in an online questionnaire (approx. 7 minutes to complete) and they are given the chance to ask any questions about the research. This concludes the study.

5. What will we do with your data?

The demographic data gathered consists of the age, gender, and experience in the chosen game. This data is completely anonymous and will be stored securely until the end of the study as is.

The audiovisual data gathered from the screen recordings will be analysed by the researcher to identify all prosocial actions performed during gameplay. Those numbers will be retained until the end of the study. The videos will be deleted post-analysis, but will be stored securely before deletion. No identifiable information found in the videos will be seen outside of the analysis stage and only by the researchers.

All questionnaire answers will be stored until the end of the study.

You also consent to allow the de-identified data to be used in future publications and other scholarly means of disseminating the findings from this research project.

6. What are the possible risks and benefits

This study does not carry any foreseeable risks. Potential benefits include learning new prosocial actions in multiple games, as well as improve in-game prosocial behavior.

7. What are your rights?

Participation is voluntary. We are only allowed to collect your data for our study if you consent to this. If you decide not to participate, you do not have to take any further action. You do not need to sign anything. Nor are you required to explain why you do not want to participate. If you decide to participate, you can always change your mind and stop participating at any time, including during the study. You will even be able to withdraw your consent after you have participated by sending an email to Panagiotis Vrettis. However, if you choose to do so, we will not be required to undo the processing of your data that has taken place up until that time. The personal data we have obtained from you up until the time when you withdraw your consent will be erased (where personal data is any data that can be linked to you, so this excludes any already anonymized data).

8. Approval of this study

This study has been allowed to proceed by the Research Institute of Information and Computing Sciences on the basis of an Ethics and Privacy Quick Scan. If you have a complaint about the way this study is carried out, please send an email to: ics-ethics@uu.nl. If you have any complaints or questions about the processing of personal data, please send an email to the Faculty of Sciences Privacy Officer: privacy-beta@uu.nl. The Privacy Officer will also be able to assist you in exercising

the rights you have under the GDPR. For details of our legal basis for using personal data and the rights you have over your data please see the University's privacy information at www.uu.nl/en/organisation/privacy.

9. More information about this study?

If you have any questions or concerns about this research, please contact Panagiotis Vrettis at p.vrettis@students.uu.nl or my supervisor Dr. Julian Frommel at j.frommel@uu.nl.

10. Do you consent to participate in the study?

You will be asked to confirm your consent when you fill out the first phase questionnaire before any data collection begins.

C First Session Questionnaire

A copy of the first session questionnaire, including the Hexad scale items and the demographics questions.

Prosocial System - Thesis | Phase 1

Start of Block: Demographics

C1 I have read the consent form given to me by the researcher, and I consent to continue with my participation in this research.

- ○ Yes (1)

Page Break

DTitle Consent and Demographics

*

D1 Enter your unique ID number given to you by the researcher

_____

*

Password Enter the password given to you by the researcher

_____

*

D2 How old are you (in years)?

_____

D3 What option best describes your gender?

- Man (1)
- Woman (2)
- Non-binary (3)
- Prefer not to say (4)
- Prefer to self describe (5)

_____

End of Block: Demographics

Start of Block: Game Related

GTitle Game Related

G1 Which game did you choose?

- ○ Helldivers II (1)
- ○ Overwatch 2 (2)
- ○ Rainbow Six: Siege (3)

G2 You are experienced with this game

- ○ Strongly disagree (1)
- ○ Disagree (2)
- ○ Somewhat disagree (3)
- ○ Neither agree nor disagree (4)
- ○ Somewhat agree (5)
- ○ Agree (6)
- ○ Strongly agree (7)

End of Block: Game Related

Start of Block: Hexad Scale

HTitle How well does each of the following sentences describes you?

P1 It makes me happy if I am able to help others.

- ○ Strongly disagree (1)
- ○ Disagree (2)
- ○ Somewhat disagree (3)
- ○ Neither agree nor disagree (4)
- ○ Somewhat agree (5)
- ○ Agree (6)
- ○ Strongly agree (7)

P2 I like helping others to orient themselves in new situations

- ○ Strongly disagree (1)
- ○ Disagree (2)
- ○ Somewhat disagree (3)
- ○ Neither agree nor disagree (4)
- ○ Somewhat agree (5)
- ○ Agree (6)
- ○ Strongly agree (7)

P3 I like sharing my knowledge.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

P4 The wellbeing of others is important to me.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

S1 Interacting with others is important to me.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

S2 I like being part of a team.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

S3 It is important to me to feel like I am part of a community.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

S4 I enjoy group activities.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

F1 It is important to me to follow my own path.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

F2 I often let my curiosity guide me.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

F3 I like to try new things.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

F4 Being independent is important to me.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

A1 I like defeating obstacles.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

A2 It is important to me to always carry out my tasks completely.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

A3 It is difficult for me to let go of a problem before I have found a solution.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

A4 I like mastering difficult tasks.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

D1 I like to provoke.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

D2 I like to question the status quo.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

D3 I see myself as a rebel.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

D4 I dislike following rules.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

R1 I like competitions where a prize can be won.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

R2 Rewards are a great way to motivate me.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

R3 Return of investment is important to me.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

R4 If the reward is sufficient I will put in the effort.

- Strongly disagree (1)
- Disagree (2)
- Somewhat disagree (3)
- Neither agree nor disagree (4)
- Somewhat agree (5)
- Agree (6)
- Strongly agree (7)

End of Block: Hexad Scale

D Second Session Questionnaire

A copy of the second session questionnaire, including the SUS items and the custom questions about the components of the Prosocial System and the encouragement they provided.

Prosocial System - Thesis | Phase 2

Start of Block: Demographics

*

D1 Enter your unique ID number given to you by the researcher

_____

*

Password Enter the password given to you by the researcher

_____

End of Block: Demographics

Start of Block: SUS

STitle System Usability Related

S1 I think that I would like to use this system frequently.

- Strongly disagree (1)
- Somewhat disagree (2)
- Neither agree nor disagree (3)
- Somewhat agree (4)
- Strongly agree (5)

S2 I found the system unnecessarily complex.

- ○ Strongly disagree (1)
- ○ Somewhat disagree (2)
- ○ Neither agree nor disagree (3)
- ○ Somewhat agree (4)
- ○ Strongly agree (5)

S3 I thought the system was easy to use.

- ○ Strongly disagree (1)
- ○ Somewhat disagree (2)
- ○ Neither agree nor disagree (3)
- ○ Somewhat agree (4)
- ○ Strongly agree (5)

S4 I think that I would need the support of a technical person to be able to use this system.

- ○ Strongly disagree (1)
- ○ Somewhat disagree (2)
- ○ Neither agree nor disagree (3)
- ○ Somewhat agree (4)
- ○ Strongly agree (5)

S5 I found the various functions in this system were well integrated.

- ○ Strongly disagree (1)
- ○ Somewhat disagree (2)
- ○ Neither agree nor disagree (3)
- ○ Somewhat agree (4)
- ○ Strongly agree (5)

S6 I thought there was too much inconsistency in this system.

- ○ Strongly disagree (1)
- ○ Somewhat disagree (2)
- ○ Neither agree nor disagree (3)
- ○ Somewhat agree (4)
- ○ Strongly agree (5)

S7 I would imagine that most people would learn to use this system very quickly.

- ○ Strongly disagree (1)
- ○ Somewhat disagree (2)
- ○ Neither agree nor disagree (3)
- ○ Somewhat agree (4)
- ○ Strongly agree (5)

S8 I found the system very cumbersome to use.

- Strongly disagree (1)
- Somewhat disagree (2)
- Neither agree nor disagree (3)
- Somewhat agree (4)
- Strongly agree (5)

S9 I felt very confident using the system.

- Strongly disagree (1)
- Somewhat disagree (2)
- Neither agree nor disagree (3)
- Somewhat agree (4)
- Strongly agree (5)

S10 I needed to learn a lot of things before I could get going with this system.

- Strongly disagree (1)
- Somewhat disagree (2)
- Neither agree nor disagree (3)
- Somewhat agree (4)
- Strongly agree (5)

End of Block: SUS

Start of Block: Future

FTitle System Effect Related

F1 Using the Prosocial System as a whole, allowed me to better understand what constitutes a prosocial action that is possible in the game of my choice.

- Strongly disagree (1)
- Somewhat disagree (2)
- Neither agree nor disagree (3)
- Somewhat agree (4)
- Strongly agree (5)

F2 Using the Prosocial System as a whole, encouraged me to be more prosocial and perform more prosocial actions in my second gameplay session.

- Strongly disagree (1)
- Somewhat disagree (2)
- Neither agree nor disagree (3)
- Somewhat agree (4)
- Strongly agree (5)

F3 The score and ranking functionality (SP and rank) of the Prosocial System encouraged me to be more prosocial and perform more prosocial actions in my second gameplay session.

- ○ Strongly disagree (1)
- ○ Somewhat disagree (2)
- ○ Neither agree nor disagree (3)
- ○ Somewhat agree (4)
- ○ Strongly agree (5)

F4 The Prosocial System helped me self reflect on my prosocial performance of my first gameplay session.

- ○ Strongly disagree (1)
- ○ Somewhat disagree (2)
- ○ Neither agree nor disagree (3)
- ○ Somewhat agree (4)
- ○ Strongly agree (5)

F5 Self-reflection helped encourage me to be more prosocial and perform more prosocial actions in my second gameplay session.

- ○ Strongly disagree (1)
- ○ Somewhat disagree (2)
- ○ Neither agree nor disagree (3)
- ○ Somewhat agree (4)
- ○ Strongly agree (5)

F6 A real reward functionality in the Prosocial System would encourage me to be more prosocial and perform more prosocial actions.

- Strongly disagree (1)
- Somewhat disagree (2)
- Neither agree nor disagree (3)
- Somewhat agree (4)
- Strongly agree (5)

F7 If the Prosocial System was fully integrated into my favorite games, I would be encouraged to be more prosocial and perform more prosocial actions.

- Strongly disagree (1)
- Somewhat disagree (2)
- Neither agree nor disagree (3)
- Somewhat agree (4)
- Strongly agree (5)

End of Block: Future

Received 26 May 2025