Integrating Large Language Models in Games With A Purpose (GWAPs) for Enhanced Knowledge Elicitation

Game design paradigms for knowledge elicitation using LLMs

Maciej Lichocki[1]

Supervisor(s): Ujwal Gadiraju[1]**, Shreyan Biswas**[1]

[1]**EEMCS, Delft University of Technology, The Netherlands**

A Thesis Submitted to EEMCS Faculty Delft University of Technology,
In Partial Fulfilment of the Requirements
For the Bachelor of Computer Science and Engineering
June 23, 2024

Name of the student: Maciej Lichocki
Final project course: CSE3000 Research Project
Thesis committee: Ujwal Gadiraju, Shreyan Biswas, Ricardo Marroquim

Abstract

The swift growth of artificial intelligence has led to the development of large language models, revolutionising various scientific domains and professional fields. This research explores collaborative, cooperative, and competitive game designs, that enhance knowledge elicitation using LLMs in games with a purpose. Through a systematic literature review using the PRISMA workflow, we identify that collaborative and cooperative game designs are more effective for knowledge elicitation than competitive designs. We also address the potential and limitations of incorporating LLMs into GWAPs, such as their use as Non-Player Characters to create engaging interactions. This study provides design principles for GWAPs leveraging LLMs, offering insights for researchers and developers, and discusses ethical considerations and future research directions.

1 Introduction

The rapid advancements in artificial intelligence have led to the development of Large Language Models, significantly impacting various scientific and professional fields. A notable application of LLMs is their integration into games with a purpose, which are designed to facilitate knowledge elicitation. This category of games is particularly effective in solving large-scale computational problems requiring human involvement, without compensating the players' enjoyment [1]. This research aims to explore the game design paradigms for games with a purpose, which enhance knowledge elicitation using LLMs.

For the GWAP to successfully help solve the underlying computational problem, the game needs to be engaging and entertaining for humans, so the largest amount of data possible is gathered [1]. Therefore, we consider three of the most common game designs (collaborative, cooperative and competitive) and their applicability to GWAPs, seeking the best practices that increase the chances of the game eliciting large volumes of knowledge successfully. Next, we investigate how LLMs can be integrated into GWAPs to increase the collected data's volume further. The dynamic text generation capabilities of LLMs can be leveraged to create more engaging games, while their ability to participate in gameplay can enhance knowledge elicitation methods [2; 3].

To assist in conducting a literature survey with the research question "What collaborative, cooperative, and competitive designs can be used for knowledge elicitation using LLMs?", we have formulated the following sub-questions:

- *What are the benefits of each game design (collaborative, cooperative, competitive)?* - Understanding the benefits of different game designs helps identify which design is most effective for knowledge elicitation.
- *How can LLMs be incorporated into games?* - Exploring the integration of LLMs in games generally, provides valuable insights into implementing them in GWAPs.
- *What impacts the efficiency of knowledge elicitation?* - Investigating factors that affect the efficiency of knowledge elicitation ensures that the chosen game design is optimal for knowledge elicitation.

By addressing these sub-questions, we can help develop effective game designs for knowledge elicitation using LLMs, aligning with the primary research question. Answers to the sub-questions are provided in Findings section.

Finally, this paper summarizes all of the necessary information and provides a concise description of the design principles useful for the creation of GWAPs for knowledge elicitation using LLMs. It discusses how these models can be adapted to fit the interactive nature of games. Additionally, it addresses the ethical considerations of using human-centric games for data collection. By investigating these areas, the study aims to contribute to any domain that utilizes knowledge collection tools. It intends to assist the researchers by helping in the creation of games, that in turn help other fields solve problems that are computationally expensive or exclusively solvable by humans.

2 Methodology

To identify the relevant literature, the PRISMA workflow was applied [4]. It introduces three key stages of the process: *identification* - searching for relevant studies across various sources; *screening* - evaluating the initial list of studies to remove irrelevant records; *inclusion* - thoroughly reviewing remaining studies to confirm their quality and including them in the systematic review.

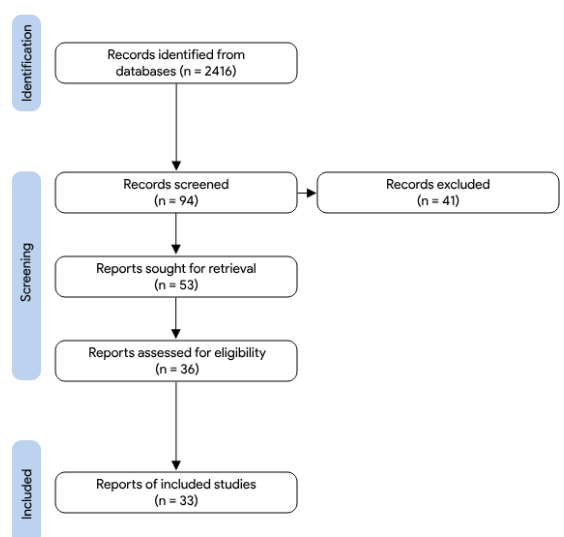

Figure 1: Applied PRISMA flow diagram

We utilized the Google Scholar and Web of Science scholarly databases [5; 6] and identified 2416 records. The search queries used were:

- `(knowledge OR information) AND (elicitation OR extraction) AND (llm* OR large language model* OR bert* OR transformer*) AND -graph*`
- `(knowledge OR information) AND (elicitation OR extraction) AND (survey OR review)`
- `(knowledge OR information) AND (elicitation OR extraction) AND efficiency`
- `(llm* OR large language model* OR bert* OR transformer*) AND (game* with a purpose OR serious game*)`
- `(collaborative OR cooperative OR competitive) AND (game* design OR game* category)`

We screened the most relevant literature identified through these search queries, assessed 36 reports as eligible, and referenced 33 of them throughout the paper. Queries more specific to the research topic, like `(knowledge OR information) AND (elicitation OR extraction) AND (llm* OR large language model* OR bert* OR transformer*) AND (collaborative OR cooperative OR competitive) AND (game* design OR game* category)` were also used, however, they yielded no results or the returned results were completely irrelevant.

The application of the PRISMA workflow, shown in Figure 1, ensured a systematic approach to identifying and selecting relevant literature. By using compound search queries, tailored to capture a wide range of studies related to knowledge elicitation, large language models, and game design we ensured the inclusivity of the selected papers.

3 Background

In this section, we provide a concise overview of Games With A Purpose, the concept of knowledge elicitation, different game designs, and the specific terminology used throughout this paper. We review related work that illustrates the practical applications and effectiveness of these approaches, highlighting the necessity for our research.

3.1 Games With A Purpose

Games with a purpose are a genre of games designed primarily to "collectively solve large-scale computational problems" as first described by Von Ahn [1]. The term arose from the observation that the massive amount of time and energy spent on playing video games by humanity, could be turned into a form of distributed human computation. GWAPs introduce a gamification element that frames tasks more as games than problems, increasing players' motivation and making the challenges presented more engaging [7]. This approach focuses on players not being aware of their contribution to the bigger problem. Examples include image labelling, web search improvement, and language translation [1].

In the context of this research, GWAPs offer a framework where players' interactions can be utilized for knowledge elicitation. By incorporating GWAP principles into the design of knowledge elicitation tasks, these activities can be transformed into enjoyable and engaging experiences that leverage the collective efforts of the participants.

3.2 Knowledge Elicitation

Knowledge elicitation is defined as the process of extracting knowledge from individuals, to make it explicit and usable in various applications, including artificial intelligence, expert systems, and decision support systems [8]. This process covers a range of techniques such as observations, interviews, and task analysis [9]. A significant challenge in knowledge elicitation is the tacit nature of expert knowledge, where experts may not be aware of the knowledge they hold or may find it difficult to articulate [10].

Another technique for knowledge elicitation involves the use of GWAPs and this research aims to address this approach by exploring how different designs can effectively capture expert knowledge.

3.3 Game Designs

Game theory categories consist of collaborative, cooperative, and competitive [11]. Each of these designs promotes distinct types of player interactions and experiences.

Collaborative

A distinctive feature of collaborative game design is its focus on team structure. In this setup, players share a unified goal, with success determined by the team's collective performance rather than individual achievements [11]. Additionally, role differentiation within the team is the primary characteristic. Players typically have varying abilities or responsibilities, ensuring that each member's unique contribution is crucial to the team's collective success [12]. The complementary nature of roles within these games encourages collaboration by requiring players to share knowledge among themselves.

Cooperative

Cooperative games are often mistaken for collaborative games, as the two terms are sometimes used interchangeably. However, the main distinction lies in the fact that "a cooperative game does not always guarantee that cooperating players will benefit equally or even benefit at all" [11]. This means, that while cooperation is required to achieve common goals, the rewards and benefits may not be distributed evenly among the players. Cooperation is encouraged by making the collective goal, which can only be achieved through teamwork, more attractive than individual objectives. The simplest form of such a game is illustrated by the Prisoner's Dilemma [13], where the shared goal is significantly more appealing than the individual ones.

Competitive

Competitive games are fundamentally different from the previous two types. They involve direct competition between players, with the primary objective being to outperform or outlast opponents, and feature Player vs Player dynamics, as well as ranking systems to track individual performance [14]. Unlike collaborative and cooperative games, competitive games often create a zero-sum environment where one player's or team's gain equates to another player's or team's loss [15]. The appeal of competitive games keeps players engaged as they strive to improve and ascend the leaderboards [16].

3.4 Related Work

Our work contributes to the growing field of human computation and knowledge elicitation through the usage of LLMs and gamification. No literature surveys have been identified that align with these goals. However, several papers do exist that detail the implementation of GWAPs in different domains.

One example is a game called "FindItOut" [17], which elicits diverse types of knowledge from players through a "Guess Who?" inspired game. It effectively generates valuable knowledge for AI tasks and captures tacit knowledge, enhancing AI systems with the information extracted.

A more direct approach was used at the University of Oulu in Finland in an experiment to elicit requirements for software development [18]. This experiment involved using games to gather requirements from stakeholders. Participants, including students from 14 nationalities, were divided into teams and engaged in game-based elicitation sessions, which proved effective in improving the quantity and quality of requirements. The use of serious games facilitated better communication and collaboration, particularly among less-experienced participants, highlighting the potential of this method in future software projects.

The absence of literature surveys on enhancing knowledge elicitation from GWAPs using LLMs motivates the need for our research. While the aforementioned implementations make effective use of GWAPs, they lack a clear rationale for the chosen game design concerning the type of knowledge being elicited. Additionally, they do not incorporate Large Language Models, which could significantly improve the quality and quantity of the extracted information.

4 Findings

In this section, we present the main findings that address the research questions outlined in the Introduction. We examine the benefits of the three main game designs, explore ways of incorporating Large Language Models into GWAPs, and analyze the factors that impact the efficacy of knowledge elicitation.

4.1 Benefits of Different Game Design Approaches

The design choice is crucial for game development, therefore in this section, we explore the benefits of various game design strategies. Each approach - collaborative, cooperative, or competitive, offers unique advantages that enhance player experience and development.

Collaborative

As stated in the Game Designs section, collaborative game design emphasizes teamwork. According to Zagal et al. [11], this design encourages players to improve their communication skills and collective problem-solving abilities. They note that players' focus shifts towards group success rather than individual achievements, which tends to be more satisfying and enjoyable. Additionally, collaborative games can foster a sense of responsibility and selflessness among players. However, the team structure, with its emphasis on group responsibility for success, can sometimes lead to issues where one player dominates decision-making. Hamalainen et al. [12] highlight the importance of clear role division in collaborative games, noting that a lack of defined roles can result in one player assuming the responsibilities of all team members.

Cooperative

Cooperative and collaborative designs share similar benefits when chosen for game development. Cooperative games promote enjoyment by encouraging players to assist one another in achieving goals that are difficult to solve individually [19]. This cooperative element has been shown to enhance players' sense of accomplishment, engagement, and overall enjoyment [20].

Hinitz et al. [21] conducted a study suggesting that cooperative games effectively enhance cooperative behaviours, including sharing, assisting, and working together towards a common goal. The study also found that such games led to an increase in positive social behaviours and a decrease in aggression among young children. These findings underscore the potential of cooperative games to make use of a more supportive and less competitive environment, which can be particularly beneficial in GWAPs aimed at more sensitive players.

Another example of the positive impact of cooperative games is demonstrated by Morschheuser et al. [22]. Through features such as communication tools, shared objectives, and complementary abilities, players are encouraged to discuss strategies, help each other, and build relationships. These interactions can positively impact players' social lives outside of the game. The potential for transferring in-game cooperative skills to real-life suggests further applications of GWAPs in experimental studies involving players.

Competitive
Competitive games offer numerous benefits that differ significantly from those provided by other game designs. The competitive aspect greatly increases the motivation of participants, which in turn positively impacts players' performance and dedication to the game [23]. The motivation in competitive games is closely tied to point-based systems and leaderboards, which rely on comparisons to other players [16]. Unfortunately, comparisons can have both positive and negative impacts. Zuo [24] demonstrated that individuals who compare themselves to others, particularly those they perceive as better off, may experience lower self-esteem, mental health issues, and dissatisfaction with their attributes.

In addition to boosting motivation, competitive design significantly enhances player engagement [23]. The thrill of competition and the desire to outperform others pushes players to invest more time and effort into the game. This increased engagement not only improves players' skills and performance but also integrates them into the game's community, creating a sense of belonging [25]. This increased level of involvement can lead to a more enjoyable gaming experience.

Apart from game categories, other factors influence player experience, a crucial aspect of GWAPs. Grudpan et al. [19] introduce the term game premise "which is the story behind the game is one of the dramatic elements and impacts the engagement of players". The authors reported on a study examining the influence of game premise on player experience in cooperative games. They concluded that choosing a negative premise increases the sense of achievement and excitement. Additionally, the negative premise led to a higher number of cooperative actions, such as knowledge sharing, indicating more active engagement in cooperative tasks.

There are also general pitfalls that can affect all game designs. If the game's outcome is not appealing and satisfactory, players may become disengaged and lose interest quickly [11]. Similarly, if the challenges presented throughout the gameplay are repetitive, predictable, or too easy, players may become bored and uninterested in continuing the playthrough.

4.2 Incorporating LLMs in Games

Large Language Models have numerous applications in various fields, in this section however we will focus solely on ways of incorporating them in games. Based on these examples, their application could be translated to GWAPs as well.

Rational players
Fan et al. [26] conducted an experiment analyzing the rationality of three OpenAI LLMs across three different games [27]. The LLMs participated in the Dictator Game to test the ability to build desires, Rock-Paper-Scissors to assess the ability to refine beliefs, and the Ring-Network Game to evaluate the ability to take optimal actions.

From the first game, the authors concluded that LLMs are capable of forming desires based on common preferences (such as equality) but struggle with uncommon preferences (like altruism). This suggests that LLMs have difficulty applying exceptional behaviour, potentially due to a lack of training data. In Rock-Paper-Scissors, LLMs performed well with familiar patterns but failed to adapt to new ones, likely also due to limitations in the training data. In the Ring-Network Game, the authors found that LLMs struggled to replicate the decision-making processes and actions typical of humans in game scenarios. As a result, they emphasized the necessity "to decouple human behaviour from LLMs in game theory" [26]. The authors also identified a research gap that needs to be addressed before proper incorporation of LLMs as rational players.

LLM-powered environments
Adekanye et al. [28] explore an innovative application of Large Language Models in generating synthetic environments for self-driving car scenarios. Their approach involves using LLMs to dynamically create realistic driving scenes within game engines like Unity and Unreal [29; 30]. The primary objective is to enhance the testing and development of autonomous vehicles by providing them with diverse and complex environments that reflect real-world driving conditions.

The methodology includes: generating driving scenarios based on user input or automatically, extracting configuration parameters for game engine setup, dynamically configuring game engines to replicate the scenarios, involving users in the simulation for interactive feedback, and integrating these synthetic environments into autonomous systems for training [28]. This research underscores the potential of Large Language Models in creating testing grounds for self-driving technology. This application implies LLMs' potential in generating realistic and challenging game environments, representing a significant advancement in the field. Such environments could be adapted for various gaming applications, providing a rich and dynamic user experience. This research is still ongoing and requires more time and investigation to become fully validated and practically applicable.

Non-player character
Cox et al. [2] conducted a comprehensive analysis of player feedback on the use of Large Language Models to generate dialogue for Non-Player Characters in the video game "Vaudeville". The incorporation of LLMs allows NPCs to engage in open-ended, dynamic conversations with players, enhancing immersion and realism. Players reported enjoying the ability to converse freely with NPCs, making the characters feel more lifelike and improving the game's replayability, as each interaction was unique and each playthrough felt original.

However, several challenges were identified. One major issue is the occurrence of LLM hallucinations, where NPCs provide inconsistent or incorrect information. This has led to player confusion, especially when the hallucinations involved critical game details, breaking immersion by introducing wrong or inappropriate elements. Another issue is the lack of memory between interactions, which requires players to repeat information, diminishing the immersive experience.

Cox et al. [2] provided several guidelines for integrating LLMs into games as NPCs. Developers must ensure that crucial details remain consistent throughout interactions to avoid player confusion. NPCs should have an incorporated memory to remember previous interactions. The evasiveness of NPC's answers must be well-balanced. To improve accessibility, both text and voice inputs should be supported. Additionally, NPCs should have conversational freedom to engage in a wide range of topics, not just game-related ones. By following these insights, developers can maximize player experience and immersion, creating more engaging NPC interactions that leverage the full potential of LLM technology in games.

4.3 Knowledge Elicitation Efficiency Factors

In 1990, Burton et al. [31] published an article evaluating the efficacy of various knowledge elicitation techniques through a series of experiments. The main results indicated that the choice of knowledge elicitation technique significantly influences its efficiency. The authors found that protocol analysis was the least efficient among the techniques tested, due to its time complexity and lower information yield. Laddering and card sorting, labelled as contrived techniques, exhibited better performance by taking less time and providing more comprehensive data. Additionally, the study revealed that different techniques might elicit different types of knowledge, suggesting that a combination of them could be the most effective approach for knowledge elicitation.

The time required to conduct the knowledge elicitation session and the subsequent time needed to transcribe and analyze the data played a significant role in the quality of the data elicited. Furthermore, the quality of transcription and analysis of the raw data from the sessions was identified as crucial for the latter stages of the process. An intriguing factor highlighted by the authors was the context in which the knowledge elicitation process occurs. They emphasized that the efficiency of knowledge elicitation varies, depending on whether the process takes place in real-world settings or controlled laboratory environments. They also identified expert-specific factors such as familiarity with the elicitation technique, articulation and interaction skills, and level of expertise. However, these factors were not the focus of our research. Understanding and optimizing these methodological factors can lead to more effective and efficient knowledge elicitation processes.

5 Discussion

In this section, we discuss the findings from the reviewed literature, place them in a broader context, and reflect on the conclusions.

5.1 Implications

In the Findings section, we presented material on the main topic of this research. Here, we reflect on these findings and discuss the existing implications of the work.

In subsection 4.1, we explored the benefits of collaborative, cooperative, and competitive game designs. Competitive games may be particularly useful when the knowledge to be elicited is derived from the game's premise, such as individual behaviour during a pandemic or decision-making in a hostage situation. Collaborative and cooperative game designs incorporate elements of teamwork, making them suitable for eliciting knowledge from both the game's premise and the interactions and group behaviour of players.
Additionally, regardless of whether a competitive or one of the casual designs is chosen, player motivation does not change significantly. Therefore, the choice of design may not heavily impact player engagement and playtime if the goal is to maximize these factors [32].

In subsection 4.2, we identified the LLM's current limitations in acting as rational players. We also highlighted the existing research gap in LLM-powered environments. Due to these limitations and the unsuitability of NPCs for competitive games, we believe that LLMs currently lack practical applications in this category. Unfortunately, these limitations also affect the application of LLMs in collaborative and cooperative games. However, the use of NPCs remains promising in these games due to their adventurous nature and less competitive environment.

For the development of GWAPs, these findings provide valuable guidelines. They suggest that developers can focus on selecting game designs based on the type of knowledge they wish to elicit rather than on player motivation. This approach can simplify the design process and lead to more effective and targeted GWAPs.

5.2 Research Process

Reflecting on the literature review process, several aspects were crucial to the success of this research. Using the PRISMA workflow established a systematic and thorough approach to finding relevant literature, ensuring that the study was built on a solid foundation of current knowledge. The search queries were carefully formulated to cover a broad spectrum of topics related to game design, integration of LLMs, and knowledge elicitation. These queries helped in capturing a wide range of relevant studies, making sure that the review was extensive.
The methodologies and approaches used in this research were effective in several ways. Listing the benefits of collaborative, cooperative, and competitive game designs provided a

solid foundation for understanding how different game designs can impact player engagement and the efficiency of knowledge elicitation. Focusing on LLMs offered valuable insights into the current capabilities and limitations of these models in game environments. This focus highlighted the potential of LLMs in enhancing GWAPs and identified critical gaps in current research, guiding future studies. Listing the factors affecting knowledge elicitation efficiency underscored the importance of selecting appropriate methods for efficient data collection. Understanding how different techniques vary in terms of time required, data quality, and context emphasized the necessity of choosing and potentially combining methods based on the specific goals and settings of the research.

6 Responsible Research

In this section, we consider the ethical aspects of the research and the reproducibility of our methods.

6.1 Bias and Representation

To mitigate the bias, we considered work from sources across various fields, including psychology, computer science, sociology and gaming. These sources come from diverse regions and institutions, ensuring a wide range of geographical and institutional perspectives. Additionally, we included research from both established and emerging scholars to represent a variety of viewpoints.

By diversifying our sources in these ways, we aimed to reduce potential bias in our literature selection and analysis. Unfortunately, we were unable to address the bias introduced by selecting only English-language literature, due to the unreliability of translation methods.

6.2 Integrity and Plagiarism

During the research, we ensured that academic integrity remained a top priority, consistently guiding our methodology and analysis. We critically evaluated each source for credibility, and relevance and transparently documented our methods. In the process of screening, we utilized OpenAI's ChatGPT[1] to efficiently assess reports for eligibility. Additionally, we used it to rephrase and help formulate certain sentences during the writing process. We avoided direct pasting and carefully checked the credibility of the information provided. We acknowledged all contributions with proper citations, adhering to the IJCAI style [33]. These practices contributed to high standards of academic integrity, making our work more credible and reliable.

7 Conclusions and Future Work

In this review, we delved into the topic of incorporating Large Language Models into Games With a Purpose for enhanced knowledge elicitation. We introduced the research question, which focused on identifying collaborative, cooperative, and competitive designs suitable for knowledge elicitation using LLMs, and provided all necessary background information. The methodology of the research was clearly explained, detailing the frameworks and queries used.

To aid understanding of the findings, we explained the crucial concepts behind them. We discovered that collaborative and cooperative game designs share similar benefits for knowledge elicitation, while competitive game design is less effective due to its lack of an adventurous nature and more vying environment. We identified the current limitations of LLMs in acting as rational players or powering game environments while underlying their promising application as Non-Player Characters. Main efficiency factors for knowledge elicitation were listed, including the choice of technique, time required for the elicitation session, and the context of the elicitation process.

We concluded that the choice of game design for a Game With A Purpose is less important, as long as the primary focus is on player engagement rather than the type of knowledge being elicited. Furthermore, we believe that Large Language Models can currently be incorporated only into collaborative and cooperative Games With a Purpose as Non-Player Characters, due to the aforementioned limitations.

We reflected on the research process, explained the existing biases in the research and justified the chosen mitigation strategies. Additionally, we emphasized the importance of academic integrity.

Through the research process, we have identified potential directions for future work. The main issues that remain unresolved relate to the field of Incorporating LLMs in Games. As explained, LLMs are not yet reliable enough to serve as rational players and further development is needed. Currently, they struggle with forming desires based on uncommon preferences, adapting to new patterns, and replicating the decision-making processes and actions typical of humans. Furthermore, the research concerning LLM-powered environments by Adekanye et al. [28] is still ongoing, with no results available yet. This indicates a substantial gap in the field, presenting numerous promising opportunities for further exploration.

Acronyms

GWAP Game With A Purpose. 1, 2, 3, 4, 5, 6

LLM Large Language Model. 1, 3, 4, 5, 6

NPC Non-Player Character. 4, 5, 6

PvP Player vs Player. 3

[1] https://chatgpt.com/

References

[1] Luis Von Ahn. Games with a purpose. *Computer*, 39:92–94, 2006.

[2] Samuel Rhys Cox and Wei Tsang Ooi. Conversational interactions with npcs in llm-driven gaming: Guidelines from a content analysis of player feedback. In *International Workshop on Chatbot Research and Design*, pages 167–184. Springer, 2023.

[3] Jen tse Huang, Eric John Li, Man Ho Lam, Tian Liang, Wenxuan Wang, Youliang Yuan, Wenxiang Jiao, Xing Wang, Zhaopeng Tu, and Michael R Lyu. How far are we on the decision-making of llms? evaluating llms' gaming ability in multi-agent environments. *arXiv preprint arXiv:2403.11807*, 2024.

[4] Elizabeth Stovold, Deirdre Beecher, Ruth Foxlee, and Anna Noel-Storr. Study flow diagrams in cochrane systematic review updates: an adapted prisma flow diagram. *Systematic reviews*, 3:1–5, 2014.

[5] Google Scholar — scholar.google.com. https://scholar.google.com.

[6] Web of Science — webofscience.com. https://www.webofscience.com/wos.

[7] Juho Hamari, Jonna Koivisto, and Harri Sarsa. Does gamification work?–a literature review of empirical studies on gamification. In *2014 47th Hawaii international conference on system sciences*, pages 3025–3034. Ieee, 2014.

[8] Nigel R Shadbolt, Paul R Smart, J Wilson, and S Sharples. Knowledge elicitation. *Evaluation of human work*, pages 163–200, 2015.

[9] Nancy J Cooke. Varieties of knowledge elicitation techniques. *International journal of human-computer studies*, 41(6):801–849, 1994.

[10] Jeremy Howells. Tacit knowledge. *Technology analysis & strategic management*, 8(2):91–106, 1996.

[11] J P Zagal, J Rick, and I Hsi. Collaborative games: lessons learned from board games. *Simulation Amp; Gaming*, 37:24–40, 2006.

[12] Raija H Hämäläinen, Mikko Niilo-Rämä, Timo Lainema, and Kimmo Oksanen. How to raise different game collaboration activities: The association between game mechanics, players' roles and collaboration processes. *Simulation & Gaming*, 49(1):50–71, 2018.

[13] W. Poundstone. *Prisoner's Dilemma: John von Neumann, Game Theory, and the Puzzle of the Bomb*. Knopf Doubleday Publishing Group, 1992.

[14] Erik Spek, Stefan Göbel, Ellen Yi-Luen Do, Esteban Clua, and Jannicke Baalsrud Hauge. Entertainment computing and serious games, 2019.

[15] Ilan Adler, Constantinos Daskalakis, and Christos H Papadimitriou. A note on strictly competitive games. In *Internet and Network Economics: 5th International Workshop, WINE 2009, Rome, Italy, December 14-18, 2009. Proceedings 5*, pages 471–474. Springer, 2009.

[16] De Liu, Xun Li, and Radhika Santhanam. Digital games and beyond: What happens when players compete? *Mis Quarterly*, pages 111–124, 2013.

[17] Agathe Balayn, Gaole He, Andrea Hu, Jie Yang, and Ujwal Gadiraju. Finditout: A multiplayer gwap for collecting plural knowledge. pages 14–18, 2021.

[18] Hadi Ghanbari, Jouni Similä, and Jouni Markkula. Utilizing online serious games to facilitate distributed requirements elicitation. *Journal of Systems and Software*, 109:32–49, 2015.

[19] S Grudpan, D Alexandrovky, J B Hauge, and R Malaka. Exploring the effect of game premise in cooperative digital board games. *Entertainment Computing and Serious Games*, pages 214–227, 2019.

[20] Anastasiia Beznosyk, Peter Quax, Karin Coninx, and Wim Lamotte. The influence of cooperative game design patterns for remote play on player experience. In *Proceedings of the 10th asia pacific conference on Computer human interaction*, pages 11–20, 2012.

[21] A K Bay-Hinitz, R B Peterson, and H R Quilitch. Cooperative games: a way to modify aggressive and cooperative behaviors in young children. *Journal of Applied Behavior Analysis*, 27:435–446, 1994.

[22] B Morschheuser, M Riar, J Hamari, and A Maedche. How games induce cooperation? a study on the relationship between game features and we-intentions in an augmented reality game. *Computers in Human Behavior*, 77:169–183, 2017.

[23] N Çağıltay, E Ozcelik, and N Ozcelik. The effect of competition on learning in games. *Computers Education*, 87:35–41, 2015.

[24] Angie Zuo. *Measuring up: Social comparisons on Facebook and contributions to self-esteem and mental health*. PhD thesis, 2014.

[25] Jessica Chesler, Suzanne McLaren, Britt Klein, and Shaun Watson. The effects of playing nintendo wii on depression, sense of belonging and social support in australian aged care residents: a protocol study of a mixed methods intervention trial. *BMC geriatrics*, 15:1–8, 2015.

[26] Caoyun Fan. Can large language models serve as rational players in game theory? a systematic analysis. *Proceedings of the Aaai Conference on Artificial Intelligence*, 2024.

[27] Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. Language models are unsupervised multitask learners. *OpenAI blog*, 1:9, 2019.

[28] Oluwanifemi Adebayo Moses Adekanye. Llm-powered synthetic environments for self-driving scenarios. *Proceedings of the Aaai Conference on Artificial Intelligence*, 2024.

[29] Afzal Hussain, Haad Shakeel, Faizan Hussain, Nasir Uddin, and Turab Latif Ghouri. Unity game develop-

ment engine: A technical survey. *Univ. Sindh J. Inf. Commun. Technol*, 4(2):73–81, 2020.

[30] Andrew Sanders. *An introduction to Unreal engine 4*. AK Peters/CRC Press, 2016.

[31] AM Burton, NR Shadbolt, G Rugg, and AP Hedgecock. The efficacy of knowledge elicitation techniques: a comparison across domains and levels of expertise. *Knowledge acquisition*, 2(2):167–178, 1990.

[32] Amy Walker. Gaming addiction, motivation, and identity: investigating the effects of competitive play in a multiplayer online battle arena game. 2017.

[33] Proceedings of the international joint conference on artificial intelligence (ijcai), 1989.