Entertainment Computing 52 (2025) 100915

Contents lists available at ScienceDirect

Entertainment Computing

journal homepage: www.elsevier.com/locate/entcom

The enduring appeal of Mahjong: Navigating the challenges of AI integration

Ruicheng Sha [a], Weijia Shi [b,*]

[a] Beijing Normal University, No. 19 Xinjiekou Wai Street, Haidian District, Beijing 100875, China
[b] The University of Hong Kong, Pokfulam Road, Hong Kong, China

ARTICLE INFO

Keywords:
Mahjong
Game Refinement Theory
AI
Player Engagement
Game Balance

ABSTRACT

Game is an important part of entertainment. This study applies Game Refinement Theory (GRT) to the zero-sum game of Mahjong, aiming to uncover its enduring appeal in terms of player engagement and to explore whether Mahjong can continue to attract players in the AI era. By applying Game Refinement Theory (GRT), we analyze real gameplay data from the online platform Tenhou and AI self-play data by akochan, following the Riichi Competition Rules (RCR). We calculate Game Refinement (GR) values to evaluate the balance between skill and chance in Mahjong. Our findings reveal that the GR value for human players averages around 0.088, indicating high engagement, while AI players exhibit GR values within the balanced range of 0.076, all within the range of game attractiveness. Key factors such as the number of riichi, furo, and player skill levels significantly influence game balance and excitement. Importantly, even as high-level players adopt AI strategies, Mahjong remains competitive and entertaining, suggesting that AI does not diminish its inherent appeal. These insights offer practical implications for game designers and AI developers aiming to enhance player experience and maintain game balance in the AI era.

1. Introduction

Mahjong is a traditional Chinese game played with a set of 144 tiles, categorized into number and honor tiles. Each player begins with a hand of 13 tiles and, in turn, players draw and discard tiles until they complete a legal hand using a 14th tile [1]. Its origins trace back to the 19th century, evolving from the older tile game called madiao. During the mid-19th century, amidst the Taiping Rebellion and the Opium Wars, Mahjong gained popularity in southern China. Throughout its history, Mahjong has remained a resilient and popular game, reflecting broader social and cultural shifts in Chinese society [2]. In China, there is a saying: "Out of a billion people, nine hundred million play Mahjong, and the remaining hundred million are watching." The Chinese mainland introduced the "Administrative Measures for the Evaluation of Mahjong Sports Technical Skills Levels", allowing players to obtain certification. In Japan, the Nihon Mahjong League was established based on Riichi Mahjong rule sets. In Taiwan, a political party called Mahjong the Greatest Party was formed to promote the game. In suburban white communities and Jewish communities in the United States, Mahjong has also become a cultural phenomenon [3]. This enduring popularity across diverse cultural contexts highlights Mahjong's unique appeal and adaptability, making it a game that transcends regional and social boundaries. As the development of technology, online Mahjong is becoming a new choice of mahjong enthusiasts. Why does Mahjong have such cross-cultural appeal?

Meanwhile, in recent years, with the rise of AI technology, it is possible for AI to analyze incomplete information games [4]. An example of AI application in Mahjong is Suphx, developed by Microsoft, which has demonstrated performance on par with top human players in competitive settings [5]. The integration of AI into Mahjong presents both opportunities and challenges. On one hand, many people are learning to play Mahjong by playing with AI and having AI analyze game records. By observing the AI's choices during games, players can learn Mahjong strategies. On the other hand, there are concerns that widespread adoption of AI strategies might homogenize gameplay among top players. If all top players all improve their Mahjong skills through AI, the games between top players may become very similar to games between AIs. Would this style of play still be attractive or competitive?

Our study aims to address these two questions by applying Game Refinement Theory (GRT) to Mahjong. By analyzing real gameplay data from the online Mahjong platform Tenhou and comparing matches between human players and matches between AI players like akochan by

* Corresponding author.
E-mail address: weijiashi@connect.hku.hk (W. Shi).

https://doi.org/10.1016/j.entcom.2024.100915
Received 16 August 2024; Received in revised form 1 November 2024; Accepted 4 December 2024
Available online 7 December 2024

GRT, we seek to uncover the factors that contribute to Mahjong's appeal, while also using GRT to analyze Mahjong AI's game records to study whether Mahjong can still maintain its appeal in the AI era. We investigate whether players at high levels maintain their competitive drive and enjoyment of the game after adopting AI strategies. We try to provide a deeper understanding of why Mahjong remains a popular and captivating game across different cultures and generations. Ultimately, we want to identify which aspects of playing Mahjong affect the player's experience, thereby attracting more players to the activity.

The innovations of this study are as follows: The study introduces a novel method for calculating the Game Refinement (GR) value of Mahjong using real gameplay data, specifically by analyzing over 26 million human player games and 22,363 AI self-play rounds. The study also presents the first comprehensive comparison between human and AI player GR values, revealing that AI players achieve more balanced gameplay (around 0.076) compared to human players (around 0.088). Through statistical analysis, the researchers identify and quantify key factors affecting player engagement, such as Riichi declarations (21.65 % for AI vs 17.30 % for human players), defensive play strategies (particularly evident in dan grades above 16), and game progression patterns (optimal GR values achieved between 20–40 tile draws). Finally, the study demonstrates that AI integration enhances rather than diminishes game appeal, with AI players showing more balanced gameplay while maintaining competitive elements.

2. Literature Review

2.1. Game and AI

From the perspective of game theory, games can be divided into complete information games and incomplete information games, corresponding to the players' knowledge of all or partial information on the game board [6]. AI players have already defeated human players in many games [7,8], with Deep Blue and AlphaGo being the most famous examples, specializing in chess and go, respectively, both complete information games. AI has become the main training tool for professional go players, but this can lead to a reduction in creativity and personalized styles, as well as a lack of adaptability when facing non-standard or uncovered moves by the AI. Currently, human players cannot fully understand the AI's operations in go. Professional players can use AI to practice their game level [9]. Game developers can use AI to design games, and predict their difficulty and player engagement [10,11].

However, due to the randomness of the game and the vast combinations of hidden situations, the performance of AI in games with incomplete information is unlikely to be as perfect as in games with complete information [12], even the AI player's level may even be lower than that of top human players. In games with incomplete information, the influence of the AI player on human players may not be as significant as in games with complete information. However, this type of research is currently insufficient yet, lacking quantitative data from actual game records to compare player experiences between the two types of games. This paper uses Mahjong as an example of a game with incomplete information to fill this research gap.

2.2. Mahjong and game Refinement Theory (GRT)

Mahjong is a very popular game and can even be considered as a cultural phenomenon in East Asia. There has been considerable research on why Mahjong is enjoyable and appealing. Existing research primarily focuses on the benefits of Mahjong for players' mental health. For elderly individuals, Mahjong can effectively improve cognitive, psychological, and slow down functional decline [13,14].Playing Mahjong as a social activity can effectively reduce loneliness and promote mental health [15]. However, these studies do not explain the game's appeal from a gameplay perspective, limiting their usefulness for game design.

The attractiveness of a game comes from its balance of competitiveness, entertainment and communication [16]. Based on this framework, the GRT serves as an effective indicator for quantifying a game's enjoyment [17]. GRT is a theory used to quantify the balance between skill and chance in games, assessing how this balance contributes to a game's excitement and engagement. Through this model, the calculated GR value measures the uncertainty and dynamic progression of a game, typically using mathematical models to evaluate the acceleration of game information. This theory has been applied to a variety of games, including board games like Chess, Go [17], doudizhu [18] and UNO [19], sports such as Soccer and Basketball [20], and traditional games like Mahjong [21], whose GR values all fall within the range of 0.07 to 0.08 [22]. By analyzing these games, GRT helps to understand what makes a game engaging and how to optimize game design for maximum player enjoyment. In previous applications of GRT to Mahjong [21], AI was used to calculate the options for each tile discard and ultimately the length of a game. Using these two data, the GR value is calculated. However, when real players are involved, it is hard to statistically account for the possible options available to players. Therefore, we will utilize Mahjong gameplay data to calculate the GR value from a different perspective, aiming to describe the balance between skill and entertainment in Mahjong.

3. Methodology

3.1. Mahjong rules

Mahjong has a variety of regional rule sets, but the basic gameplay is consistent. The standard set includes 144 tiles with Chinese characters and symbols, though some variations may exclude certain tiles. The Mahjong tile set consists of 144 tiles, including 4 each of the Bamboo (1–9 Bamboo), Character (1–9 Character), and Dot (1–9 Dot) suits, as well as 4 each of the four Wind tiles (East, South, West, North) and 3 Honor tiles (Red, Green, White). Each player starts with 13 hidden tiles and takes turns drawing and discarding tiles to complete a winning hand of 14 tiles. Players also take turns being the "oya," whose points gained or lost are typically greater. This process involves strategic tile selection and observing opponents' moves to form sets and sequences, declaring victory with a complete tile set. Japanese Riichi Mahjong, with its stable and well-established rules, provides a reliable data foundation and unique strategic depth, making it representative for Mahjong studies. Therefore, this paper uses Riichi Mahjong as an example.

In Riichi Competition Rules (RCR), when a player completes a valid tile set and wins, it is called "agari". During their turn, a player can call "chii" to form a sequence of 3 consecutive numbered tiles of the same suit by claiming a discarded tile from an opponent, specifically as part of that sequence, rather than just any 3 tiles. A player can call "pon" to form a triplet by claiming a discarded tile from an opponent, specifically to complete a set of 3 identical tiles, rather than simply having 3 identical tiles. Additionally, a player can call "kan" to form a quadruplet by having 4 identical tiles, either in their hand or when an opponent discards a tile that completes the set. Kan can be either "closed kan" (from the player's hand) or "open kan" (by claiming the opponent's discarded tile). The terms "chii," "pon," and "open kan" are known as "furo". To win, a player needs at least one "yaku". Being one tile away from winning is called "tenpai". A common yaku is "riichi," declared when a player is in tenpai with no furo. After "agari," the winner receives a score from the loser. If no one achieves agari by the end of the round, it results in a "ryuukyoku". A match consists of several rounds. Among the Mahjong tile actions, "furo", discarding, "riichi", and "closed kan" are the choices a player can make.

3.2. Data Collection for human players and AI players

Tenhou is a prominent online Mahjong platform in Japan, officially rebranded to its current name on March 1, 2007. Tenhou employs a sophisticated ranking system and offers both four-player and three-player Mahjong modes based on RCR, each with distinct rules and ranking criteria. After winning games and accumulating points, players can increase their “dan grade”. Notably, Tenhou allows users to download game records from tenhou.net/sc/raw/, providing a valuable resource for detailed analysis and study. However, for private rooms and dan rankings, only overall scores can be downloaded, and not the specific moves made by players during the matches. We can download the ranked match and phoenix table records from 2009 to 2022. Due to the extensive research already conducted on four-player Mahjong, we have chosen to focus our analysis on four-player games. We have downloaded a total of 26,406,213 rounds of Mahjong games.

Given the lack of open-source code or weights for several strong Mahjong AIs currently recognized, we selected akochan as our AI player [23]. The source code of akochan was downloaded from its official repository and compiled to guarantee we were using the most up-to-date and stable version. These self-play matches were designed to mirror the structure of real Mahjong games, with each match consisting of several rounds. The matches were conducted in a controlled environment to eliminate external variables that could influence gameplay, ensuring the data reflected pure AI decision-making processes. During these matches, every move and decision made by the AI, including tile draws, discards, and declarations of riichi and furo, was recorded. To maintain consistency with the human player data, the raw scores from these matches were normalized by dividing them by 100. We performed a preliminary statistical analysis to identify any outliers or irregularities in the AI’s performance data, ensuring the final dataset was robust and representative. Additionally, we cross-checked the results of multiple matches to verify the consistency and reliability of the AI’s performance, confirming there were no anomalies or deviations from expected behavior. We generated 194,410 self-play rounds using akochan (40,000 matches).

3.3. Game Refinement Theory (GRT) and its apply in Mahjong

From the player’s perspective, game information progress is an increasing function of the number of moves t made during the game. Notably, game information progress is not necessarily the score obtained by players during the game. The closer the game is to the end, the more complete the information obtained. In some games, such as Mahjong, the complete information is only known at the very end of the game. Therefore, a realistic model of game information progress is proposed.

$$x(t) = G\left(\frac{t}{T}\right)^n \quad (1)$$

$x(t)$ represents the game information progress, t is the number of moves made by the player during the game, T is the total moves of the game, and n is a constant determined by the perspective of the observer. Since more information is obtained as the game nears its end, assuming a relatively balanced game, the second derivative of the formula is taken, giving:

$$x''(t) = \frac{Gn(n-1)t^{n-2}}{T^n} \quad (2)$$

At $t = T$, which means the end of a match, we have:

$$x''(T) = \frac{Gn(n-1)}{T^2} \quad (3)$$

This describes the acceleration of game information progress. Therefore, we expect that the larger the value of $\frac{G}{T^2}$, the more exciting the game becomes due to the uncertainty of the game outcome. Thus, we use the square root of $\frac{G}{T^2}$ as the Game Refinement (GR) value for the game under consideration, which is:

$$R = \frac{\sqrt{G}}{T} \quad (4)$$

The GR values of several other games [24] are shown in Table 1. If the game’s GR value is higher than the range of 0.07–0.08, players may feel more entertained. A GR value within the range of 0.07–0.08 means the game has a good balance between chance and skill. A GR value lower than the range of 0.07–0.08 indicates that the game tends to be competitive [25].

Table 1
Game Refinement value for various games [24].

Game	GR value
Basketball	0.073
Soccer	0.073
Badminton	0.086
DotA ver 6.80	0.078
StarCraft II Terran	0.081

Since Mahjong involves the gain and loss of points, game information progress can be substituted with score progress. However, since Mahjong is a zero-sum game, the expected score will be close to zero, making the GR analysis using expected score meaningless. Therefore, we choose to separately compute positive scores (scores gained) and negative scores (scores lost). Due to the existence of chi, pon, and kan, the number of tiles each player draws and discards before the end of the game, and the total number of tiles discarded by all players, are not fixed. Thus, we choose to use the total number of tiles drawn by all players as the game progresses. We selected only rounds without ryuukyoku to create Fig. 1. It can be observed that as the game progresses, the score fluctuations initially decrease and then increase. The probability of a player achieving agari increases before the total number of tile draws reaches 40. After 41 tile draws, the probability of the game ending in ryuukyoku gradually increases.

In non-zero-sum games, the GR value in the game may be calculated using Formula 5 [26,27]:

$$R = \frac{\sqrt{winnerpoint}}{totalpoint} \quad (5)$$

Considering the characteristics of Mahjong games, we derived the formula 6:

$$R = \frac{\sqrt{\overline{P}}}{\overline{P} + \overline{N}} \quad (6)$$

where $\overline{P}$ is the average score of the winners and $\overline{N}$ is the average score loss of the losers (both positive values). The numerator reflects the average score gained by the winners, and the denominator reflects the average score difference between winners and losers. In RCR, Mahjong scores are typically multiples of 100. For analysis purposes, we normalize the game scores by dividing them by 100. The scores downloaded from the Tenhou platform are similarly adjusted, ensuring consistency in our data analysis.

4. Results

4.1. Results for human players

After obtaining the data, we conducted a statistical analysis of the matches across different years. It was observed that as time progresses, the number of rounds shows an overall increasing trend (see Fig. 2). Meanwhile, the GR value exhibits a decreasing trend but remains around 0.088. As time passed, the number of players on Tenhou increases, the matches in RCR become less entertaining and more balanced. This trend may be attributed to players gradually adopting more cautious and conservative strategies during gameplay.

We analyzed the changes in the GR value at different stages in rounds with agari and found that the trend generally shows an initial increase followed by a decrease (see Fig. 3). This phenomenon can be understood as follows: At the beginning, players with agari usually have very organized tile sets. The game progresses too quickly, and the

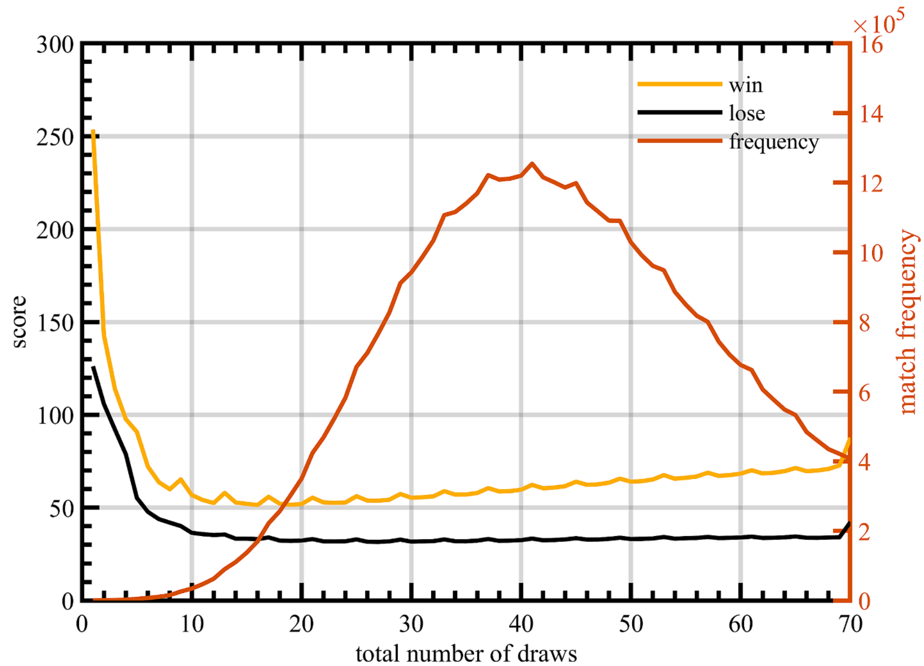

Fig. 1. Correlation between the average absolute score changes and the total number of tile draws in Mahjong games without ryuukyoku. The horizontal axis represents the total number of tiles drawn by all players before the end of the game. The left vertical axis shows the average absolute value of score changes per game, and the right vertical axis indicates the frequency of rounds. Initially, score fluctuations decrease as players build their hands, then increase as players approach winning hands. The fluctuations in the curve are influenced by the oya, who experiences higher score changes due to game rules.

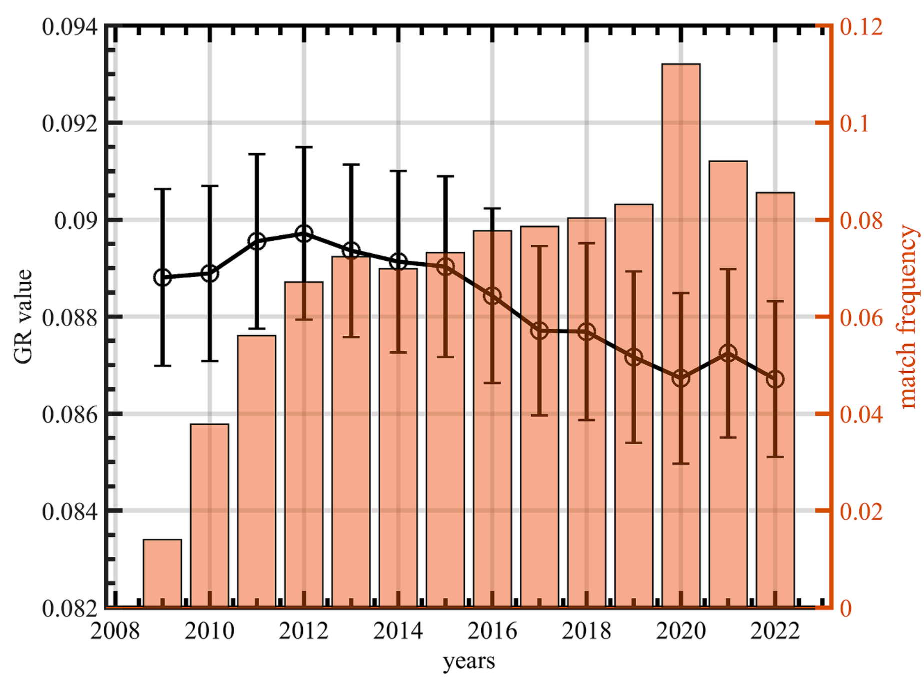

Fig. 2. Correlation between Game Refinement (GR) value and match frequency over the years from 2009 to 2022. The horizontal axis indicates the year, the left vertical axis represents the calculated GR value, and the right vertical axis indicates the frequency of rounds played each year. The GR value, based on average scores, shows a slight decreasing trend, remaining around 0.088. This suggests that as Mahjong's popularity increases, players adopt more cautious strategies, leading to more balanced but slightly less entertaining games.

information changes in the game are minimal. Therefore, the GR value is relatively low. As the game progresses, players gradually build organized tile sets, and the role of strategy becomes apparent, leading to increased uncertainty and a rising GR value. However, in the final stages of the match, as players' hands become more complete and the possibility of winning increases, uncertainty decreases again, leading to a decline in the GR value.

Riichi is a distinctive feature of Mahjong under the RCR rules, which generally encourage riichi. In the game, the decision to riichi contrasts with the decision to furo. The analysis of the data (see Fig. 4) reveals the dynamics of riichi declarations within Mahjong games. Lower numbers of riichi declarations are more frequent. The line graph depicting the GR

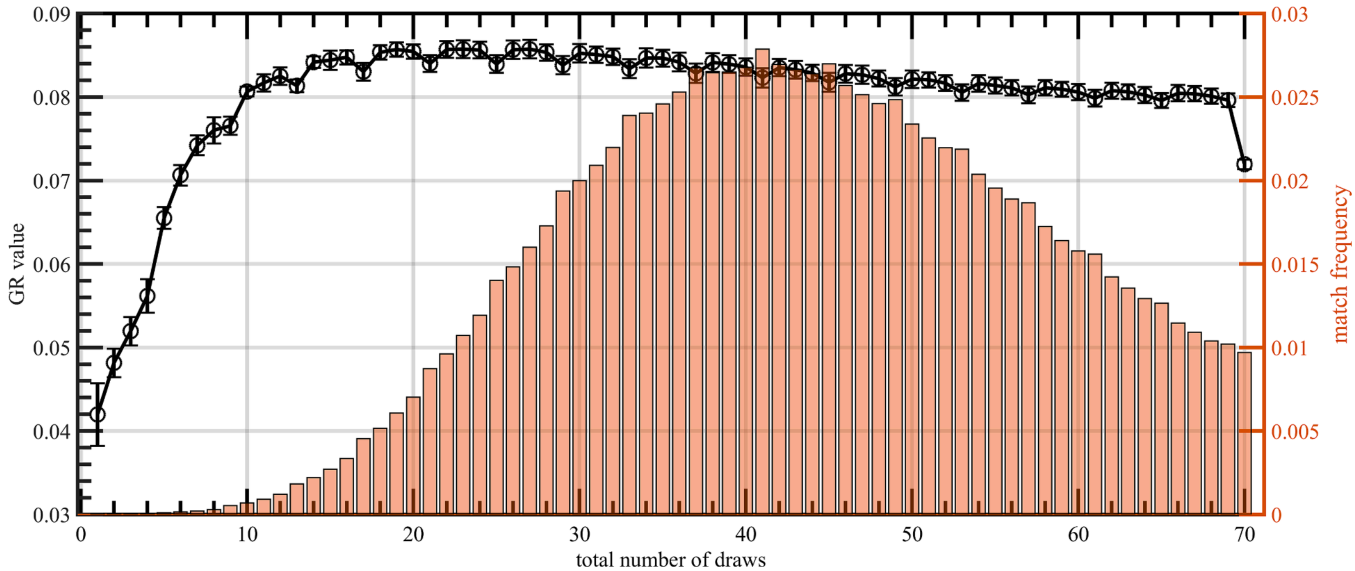

Fig. 3. Correlation between the GR value, match frequency, and the number of tiles draws in Mahjong games ending with an agari. The horizontal axis shows the total tile draws, the left vertical axis displays the GR value, and the right vertical axis indicates the frequency of rounds. The GR value increases initially, reflecting growing uncertainty and strategic depth, then decreases as players near the endgame and uncertainty diminishes. Fluctuations are partly due to the oya's higher score potential. This figure illustrates the impact of game length on excitement and balance in Mahjong.

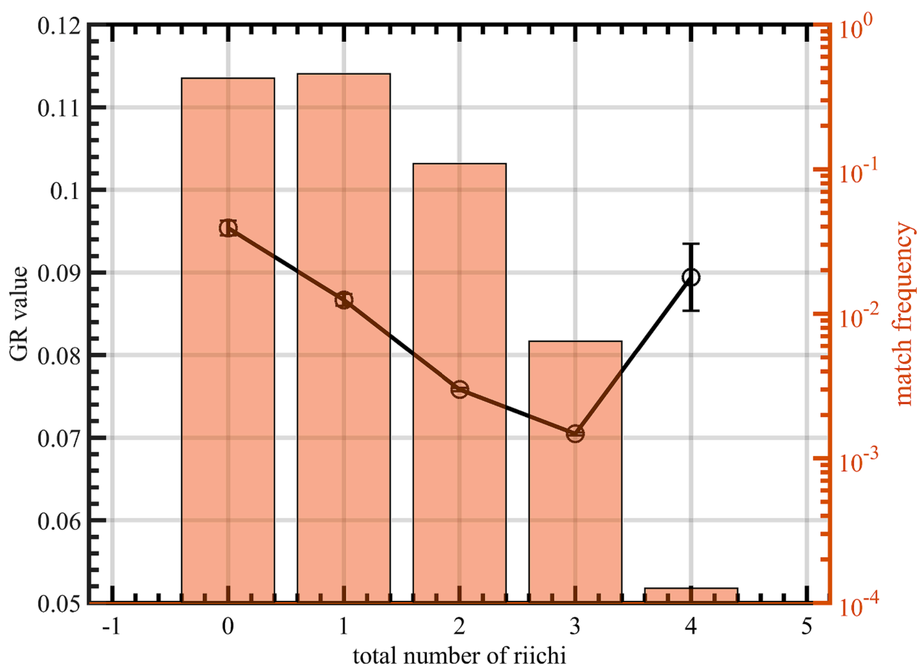

Fig. 4. Correlation between the GR value, match frequency, and the number of riichi declarations in Mahjong games. The horizontal axis represents the count of riichi declarations per game (0 to 4), the left vertical axis shows the GR value, and the right vertical axis (logarithmic scale) indicates the frequency of rounds. As the number of riichi declarations increases, the GR value decreases from about 0.09 to 0.07, indicating that games with more riichi are more balanced but potentially less thrilling. The rarity of games where all four players declare riichi is also highlighted. This figure emphasizes how riichi declarations influence game dynamics and player engagement.

value indicates a gradual decrease from about 0.09 to about 0.07 as the number of riichi declarations increases from zero to three. Specially, in cases where all four players declare riichi, the special rule "Sucha Riichi" comes into play, leading to ryuukyoku if the last player successfully declares riichi. Therefore, games with more riichi tend to be more balanced, with GR value between 0.07 and 0.08.

Furo represents another strategic choice for players. It is observed that the GR value increases when the number of furo in a game ranges from 0 to 5 but then decreases when the number of furo exceeds 6 (see Fig. 5). The GR value consistently remains above 0.08, indicating uncertainty in the games. Therefore, Mahjong does discourage an excessive number of furo. Considering that games with many furo are actually quite rare, games with fewer furo are more balanced.

Dan grade is an important factor influencing the GR value. A sudden change is observed at dan grade 15 to 16 (see Fig. 6). When players' dan grades are above 16, they can enter the "Senior table" in Tenhou. This phenomenon can be understood as follows: At lower dan grades, players are less inclined to play defensively, making it relatively easy to

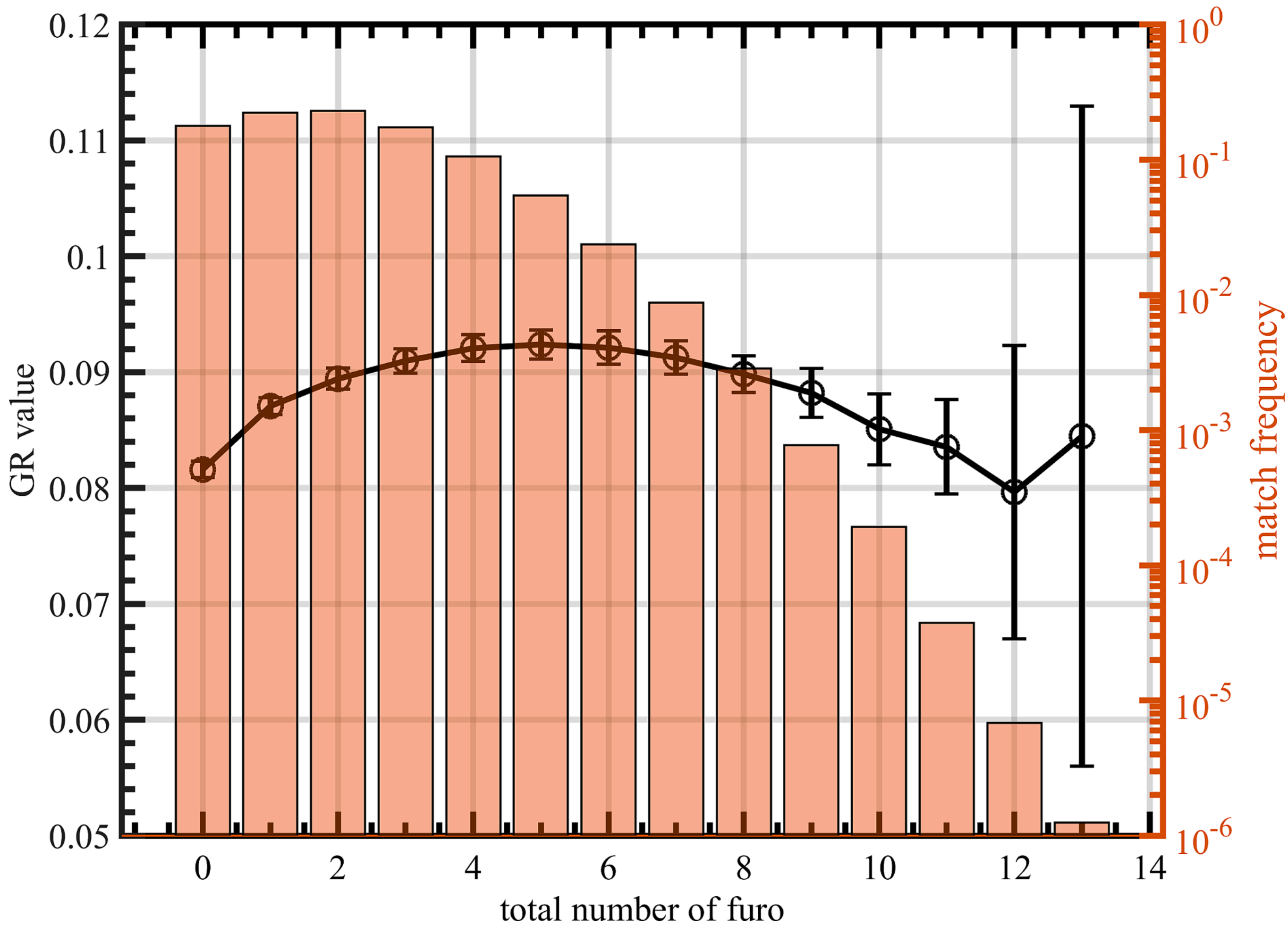

Fig. 5. Correlation between the GR value, match frequency, and the number of furo (calls like chi, pon, or open kan) in Mahjong games. The horizontal axis indicates the total furo count per game, the left vertical axis shows the GR value, and the right vertical axis (logarithmic scale) represents match frequency. The GR value increases with furo counts from 0 to 5, suggesting enhanced uncertainty and excitement, but decreases when furo counts exceed 6, indicating diminished balance. This demonstrates that while furo can add complexity, excessive furo may detract from game appeal. The low frequency of high furo counts also points to player tendencies towards balanced play.

complete tile set after learning the Mahjong rules. At higher dan grades, the proportion of defensive play increases, making it more difficult for players to complete tile sets. The GR values, as well as the average scores and losses for players with low and high dan grades, are shown in Table 2. The GR values for games in lower dan grades are closer to the range of 0.07–0.08. Therefore, for beginners, Mahjong is more balanced. This also explains why Mahjong has strong appeal to beginners.

Additionally, the oya is a unique feature of Mahjong games. We separately calculated the average positive score and negative score for players when they are and are not the oya, and computed the corresponding GR values. As shown in Table 3, the GR value for the oya is lower, closer to the range of 0.07 to 0.08, making games more balanced when players are the oya.

4.2. Results for AI players

To address the questions posed at the beginning of the paper, we configured akochan and calculated its GR value during self-play. We processed the data in a manner similar to the above analysis. The skill level of akochan is approximately between dan grade 15–16 on Tenhou. We found that the overall average GR value of the AI player in Mahjong is 0.076, which is lower than the 0.088 average of human players.

We observed that as the Mahjong game progresses, the GR value initially increases and then decreases (see Fig. 7). Notably, around the 20th tile draw, the GR value for AI players aligns with that of human players. However, the GR value for AI players soon declines sharply, falling below that of human players. This trend suggests that after a certain stage, the scores of the tiles completed by AI players increase rapidly and surpass those of human players. Consequently, the GR value for AI players in the game approaches the range of 0.07–0.08, contributing to a more balanced Mahjong game. Games with a lower number of tile draws are less frequent, which lead to increased statistical variance in the results.

In AI self-play Mahjong games, we observed that the proportion of riichi declarations is higher, increasing from 17.30 % for human players to 21.87 % for AI players. This results in AI players winning with more score, likely due to their superior game strategies compared to human players (see Fig. 8). The average GR value of all rounds is 0.076. Consequently, the GR value for AI players tends to fall within the 0.07–0.08 range, making the Mahjong game more balanced.

The furo ratio for human and AI players is 35.52 % and 35.96 %, which are quite close. Additionally, the trend in GR value changes is similar (see Fig. 9). Therefore, we can consider that the furo ratio and decisions made by human players are relatively reasonable. Games with a higher number of furo are less frequent, which lead to increased statistical variance in the results.

As shown in Table 4, regardless of whether the player is the oya, the absolute values of scores gained and lost by AI players in Mahjong games

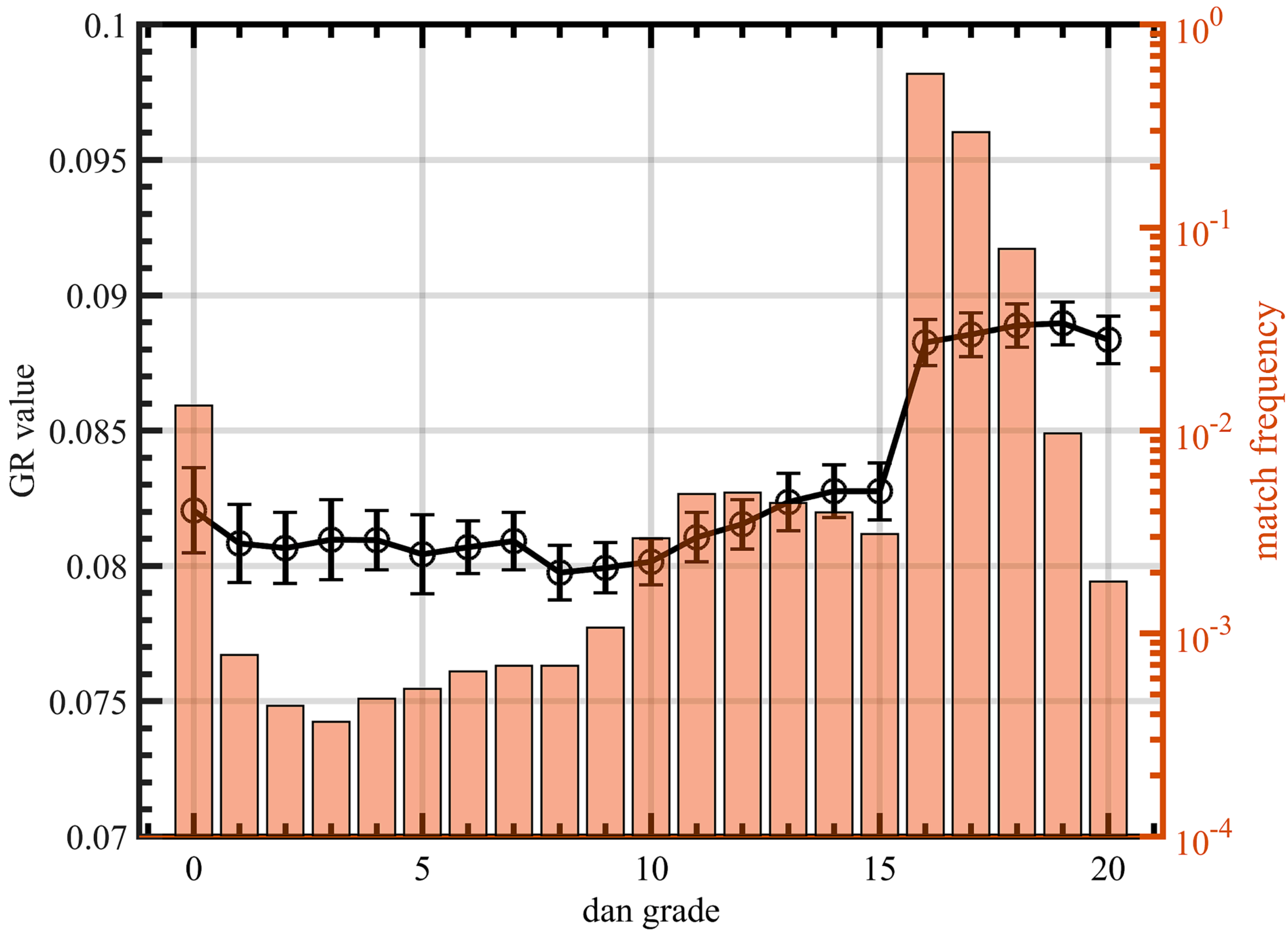

Fig. 6. Correlation between the GR value, match frequency, and skill level (dan grade) in Mahjong games. The horizontal axis represents the dan grade, ranging from beginner to advanced levels, the left vertical axis shows the GR value, and the right vertical axis (logarithmic scale) indicates the frequency of rounds. A notable increase in GR value occurs around dan grades 15 to 16, where players can enter the "Senior table" on Tenhou. Higher dan grades correspond to higher GR values. Due to only ranked match and phoenix table records are open for download, matches for lower dan grades are relatively less. However, even at the lowest dan grade of 3, there are still 38,425 recorded matches.

Table 2
Comparison of GR Values, Average Positive Scores, and Negative Scores Between High-Rank (≥16) and Low-Rank (<16) Dan Grade Players in Tenhou Mahjong.

Dan grade	Positive Score	Negative Score	GR value
>=16	53.2190	29.9861	0.0877
<16	57.8070	35.1824	0.0818

Table 3
Comparison of GR Values, Average Positive Scores, and Negative Scores for Oya or not Oya in Mahjong Games.

oya	Positive Score	Negative Score	GR value
no	46.4113	28.9507	0.0904
yes	65.5024	32.9420	0.0822

are higher than those of human players. This further contributes to lower GR value, enhancing the game's balance.

5. Discussion and Conclusion

5.1. Factors contributing to the balance of Mahjong and game design

Through data analysis, we have identified several key trends and conclusions regarding Mahjong games under the RCR rules. By examining the influence of various selected factors on the GR value, we have uncovered insights: towards the end of the game, playing as the oya, declaring riichi, and having fewer furo in the game make the game more balanced. For beginners with lower dan grades, the GR value is in a more balanced range, making Mahjong more attractive to new players. Moreover, AI players have a higher rate of riichi declarations and greater score fluctuations per game compared to human players, ultimately leading to a lower GR value for AI players. This trend differs from the variations observed between low and high dan grade players, which suggests that high dan grade players who learn Mahjong strategies from AI might make the game more balanced. In the AI era, players learning Mahjong techniques from artificial intelligence will not diminish the game's inherent appeal.

We found that AI players tend to use riichi strategies more frequently, resulting in greater score fluctuations and a more balanced

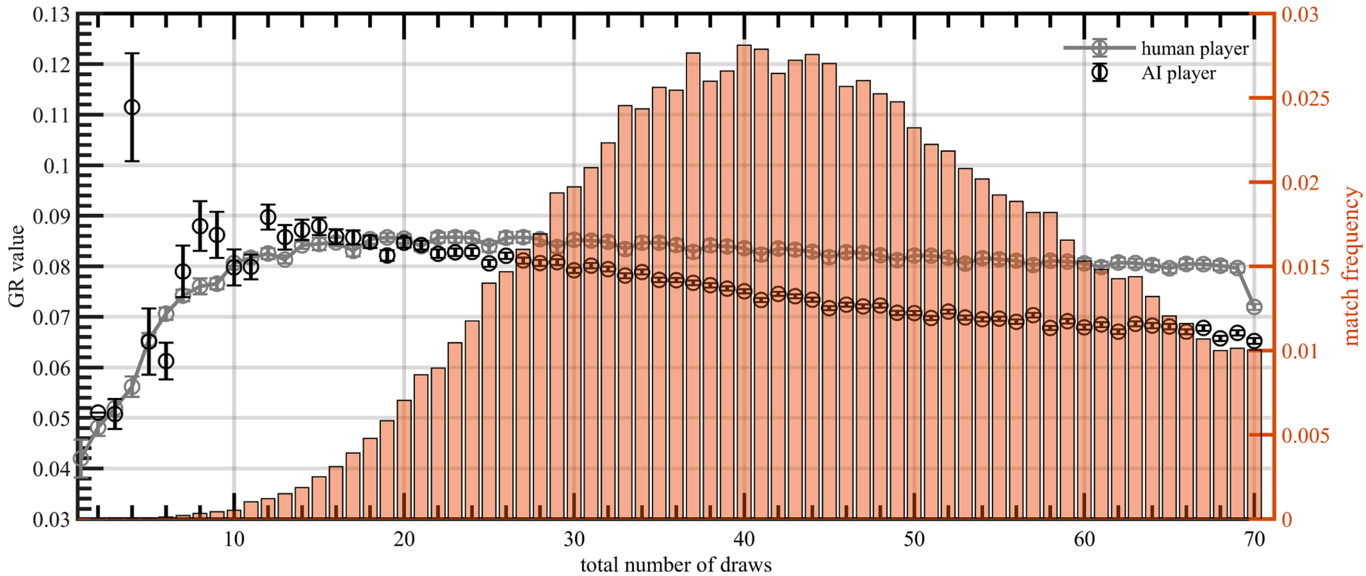

Fig. 7. Correlation between the GR value, match frequency, and total tile draws in AI self-play Mahjong games using akochan. The horizontal axis represents total tile draws, the left vertical axis shows the GR value, and the right vertical axis indicates match frequency. Initially, the GR value for AI players aligns with that of human players but declines after approximately 20 tile draws, falling below human GR values. This suggests that AI players form winning hands more efficiently, leading to larger score gains, with a GR value stabilizing in the 0.07–0.08 range, indicating a more balanced game.

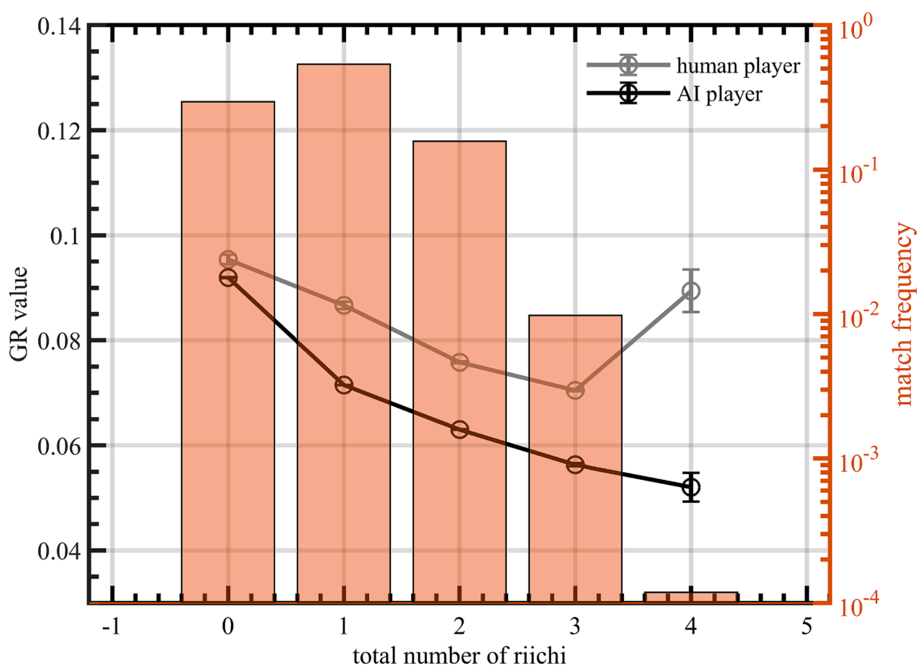

Fig. 8. Correlation between the GR value, match frequency, and the riichi declarations in AI self-play Mahjong games. The horizontal axis shows the number of riichi declarations per game, the left vertical axis displays the GR value, and the right vertical axis (logarithmic scale) indicates match frequency. AI players have a higher riichi declaration rate (21.87%) compared to human players (17.30%), leading to games with higher-scoring hands and GR values within the balanced range of 0.07–0.08. This figure demonstrates how AI strategies favoring riichi influence game balance and excitement.

game experience. This finding highlights AI's potential to alter game strategies and balance. Game designers could use AI analysis to fine-tune game rules, ensuring the game remains balanced and engaging even at higher levels of play. We observed that dan grades impact the GR value, emphasizing the importance of learning curve design for long-term player engagement. Game designers can leverage this finding to create more detailed progression systems, ensuring players maintain interest as their skills improve. For example, designers could implement staged goals and rewards to encourage players to continuously learn and refine their strategies while maintaining game balance at each stage. These insights can be extended to other types of games, such as strategy games, card games, and even some eSports. By carefully considering these factors, game designers can create more engaging, balanced, and long-lasting gaming experiences.

5.2. Limitations and Future research Directions

Although this study focuses Mahjong based on RCR, different regions and cultures have various Mahjong rule sets. Future research could compare GR values across different rule sets to analyze their impact on

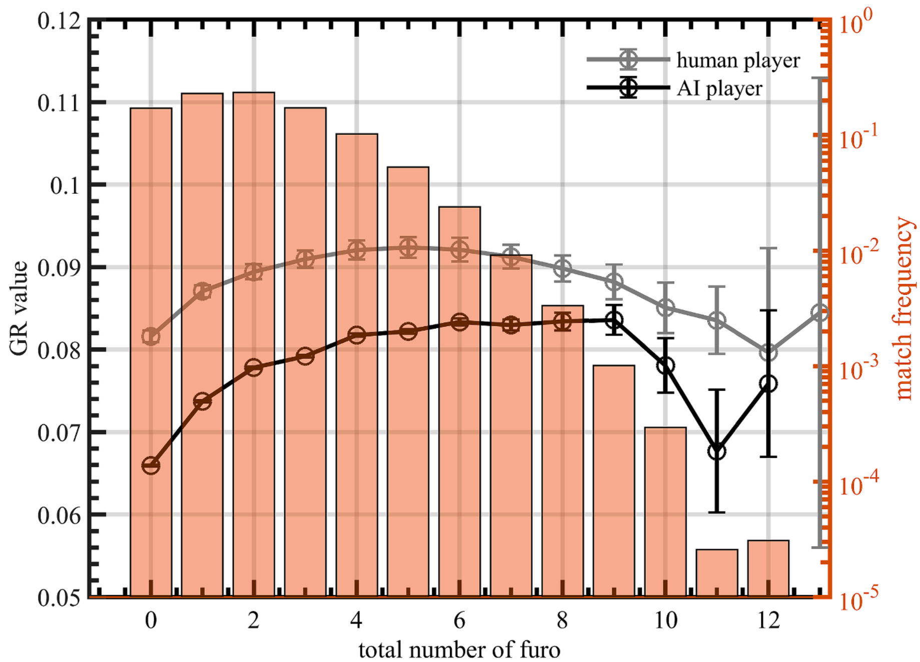

Fig. 9. Correlation between the GR value, match frequency, and the furo counts in AI self-play Mahjong games. The horizontal axis represents the number of furo per game, the left vertical axis shows the GR value, and the right vertical axis (logarithmic scale) indicates match frequency. The furo ratio for AI players (35.96%) is close to that of human players (35.52%), and the GR value trends are similar. This suggests that AI decisions regarding furo are aligned with human tendencies.

Table 4
Comparison of GR Values, Average Positive Scores, and Negative Scores for Oya or not Oya in AI Mahjong Games.

oya	Average Positive Score	Average Negative Score	Average GR value
no	49.2860	30.7225	0.0876
yes	70.9609	34.8965	0.0798

game balance and appeal, offering further insights into optimal game design strategies. For example, Sichuan Mahjong is known for its entertainment [28]. Chinese Standard Mahjong is different from RCR, it encourages the players to adopt aggressive strategies, which makes the competition watchable and attractive [29]. These may result in higher GR values in such rule sets.

Aside from Suphx, there are other AIs, such as Mortal [30] and NAGA [31]. Mortal uses the core logic of akochan discussed in this article and has made improvements, thus performing better than akochan. NAGA's dan grade is around 18, which is superior to the AI utilized in this paper. However, the weights of Mortal are not open source, both code and weights of NAGA and Suphx are not open source, so this study could not utilize and analyze data from these AIs. Additionally, there are criticisms of akochan's overly aggressive riichi strategies. As AI technology advances, more sophisticated Mahjong AIs with advanced strategies and techniques may emerge. Future research can leverage these stronger AIs to recalculate and analyze GR values for a better understanding of game balance. Similarly, this research approach could be applied to other games, such as Go and MOBA games, to provide valuable insights into their respective game dynamics and balance. Additionally, the number of AI self-play games in our study was relatively limited, which may have contributed to a higher variance in the calculated GR values. In future research, increasing the sample size of AI self-play matches would be beneficial to reduce this variance and provide more robust results. This expansion of the dataset would allow for a more precise and reliable analysis of AI gameplay patterns and their impact on game balance.

CRediT authorship contribution statement

Ruicheng Sha: Writing – review & editing, Methodology, Data curation. **Weijia Shi:** Writing – original draft, Software, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

References

[1] S. Li, X. Yan, Let's Play Mahjong!, (2019). http://arxiv.org/abs/1903.03294 (accessed October 30, 2024).
[2] M. Greene, The Game People Played: Mahjong in Modern Chinese Society and Culture, *Cross-Currents* 5 (2016) 1–27, https://doi.org/10.1353/ach.2016.0001.
[3] A. Heinz, *Mahjong: a Chinese game and the making of modern American culture*, Oxford University Press, New York, 2021.
[4] Y. Zheng, S. Li, A Review of Mahjong AI Research, in: Proceedings of the 2020 2nd International Conference on Robotics, Intelligent Control and Artificial Intelligence, ACM, Shanghai China, 2020: pp. 345–349. DOI: 10.1145/3438872.3439104.
[5] J. Li, S. Koyamada, Q. Ye, G. Liu, C. Wang, R. Yang, L. Zhao, T. Qin, T.-Y. Liu, H.-W. Hon, Suphx: Mastering Mahjong with Deep Reinforcement Learning, (2020). http://arxiv.org/abs/2003.13590 (accessed October 30, 2024).
[6] R. Gibbons, An Introduction to Applicable Game Theory, (n.d.).
[7] G. Robertson, I. Watson, A Review of Real-Time Strategy Game AI, *AI Mag.* 35 (2014) 75–104, https://doi.org/10.1609/aimag.v35i4.2478.
[8] P. Bory, Deep new: The shifting narratives of artificial intelligence from Deep Blue to AlphaGo, *Convergence: the International Journal of Research into New Media Technologies* 25 (2019) 627–642, https://doi.org/10.1177/1354856519829679.
[9] J. Kang, J.S. Yoon, B. Lee, How AI-Based Training Affected the Performance of Professional Go Players, in: CHI Conference on Human Factors in Computing Systems, ACM, New Orleans LA USA, 2022: pp. 1–12. DOI: 10.1145/3491102.3517540.

[10] S. Roohi, C. Guckelsberger, A. Relas, H. Heiskanen, J. Takatalo, P. Hämäläinen, Predicting Game Difficulty and Engagement Using AI Players, Proceedings *ACM Hum.-Comput. Interact.* 5 (2021) 1–17, https://doi.org/10.1145/3474658.
[11] C. Guerrero-Romero, S.M. Lucas, D. Perez-Liebana, in: Using a Team of General AI Algorithms to Assist Game Design and Testing, IEEE, Maastricht, 2018, pp. 1–8, 10.1109/CIG.2018.8490417.
[12] Q. Yin, J. Yang, K. Huang, M. Zhao, W. Ni, B. Liang, Y. Huang, S. Wu, L. Wang, AI in Human-computer Gaming: Techniques, Challenges and Opportunities, Machine Intelligence Research 20 (2023) 299–317, https://doi.org/10.1007/s11633-022-1384-6.
[13] K. Machishima, M. Sakamoto, C. Oshima, K. Kitajima, T. Koba, R. Gohara, A. Sato, J. Higashiuchi, E. Horikawa, * The Effect of Mahjong on Healthy Elderly Persons' Cognitive Function, *Arch. Clin. Neuropsychol.* 29 (2014) 505, https://doi.org/10.1093/arclin/acu038.5.
[14] Z.C.K. Tse, Y. Cao, B.K.H. Chau, M.K. Yeung, C. Leung, D.H.K. Shum, Does Playing Mahjong Benefit Older Individuals? A Scoping Review, Journal of Prevention of Alzheimer's Disease (2024), https://doi.org/10.14283/jpad.2024.102.
[15] R. Wang, Z. Chen, Y. Zhou, L. Shen, Z. Zhang, X. Wu, Melancholy or mahjong? Diversity, Frequency, Type, and Rural-Urban Divide of Social Participation and Depression in Middle- and Old-Aged Chinese: A Fixed-Effects Analysis, *Social Science & Medicine* 238 (2019) 112518, https://doi.org/10.1016/j.socscimed.2019.112518.
[16] H. Iida, *Fairness, Judges and Thrill in Games*, (n.d.).
[17] H. Iida, N. Takeshita, J. Yoshimura, A Metric for Entertainment of Boardgames: Its Implication for Evolution of Chess Variants, in: R. Nakatsu, J. Hoshino (Eds.), *Entertainment Computing*, Springer US, Boston, MA, 2003: pp. 65–72. DOI: 10.1007/978-0-387-35660-0_8.
[18] Y. Gao, W. Li, Y. Xiao, M.N.A. Khalid, H. Iida, Nature of Attractive Multiplayer Games: Case Study on China's Most Popular Card Game—DouDiZhu, *Information* 11 (2020) 141, https://doi.org/10.3390/info11030141.
[19] A. Ramadhan, N.U. Maulidevi, H. Iida, Game Refinement Theory and Multiplayer Games: Case Study Using UNO, (2015).
[20] A. Pratama Sutiono, R. Ramadan, P. Jarukasetporn, J. Takeuchi, A. Purwarianti, H. Iida, A Mathematical Model of Game Refinement and Its Applications to Sports Games, *EAI Endorsed Transactions on Creative Technologies 2* (2015) 150095, https://doi.org/10.4108/eai.20-10-2015.150095.
[21] H. Iida, K. Takahara, J. Nagashima, Y. Kajihara, T. Hashimoto, An Application of Game-Refinement Theory to M.h. Jong, in: M. Rauterberg (Ed.), *Entertainment Computing – ICEC*, Springer, Berlin Heidelberg, Berlin, Heidelberg, 2004, pp. 333–338, 10.1007/978-3-540-28643-1_41.
[22] G. Naying, G. Yuexian, M.N.A. Khalid, H. Iida, A computational game experience analysis via game refinement theory, *Telematics and Informatics Reports* 9 (2023) 100039, https://doi.org/10.1016/j.teler.2022.100039.
[23] critter-mj, *akochan*, (n.d.). https://github.com/critter-mj/akochan (accessed October 30, 2024).
[24] Z. Zeliang, *Game Refinement Theory and its Application to Fighting Game and Action Game*, (n.d.).
[25] S. Xiong, L. Zuo, H. Iida, Possible Interpretations for Game Refinement Measure, in: N. Munekata, I. Kunita, J. Hoshino (Eds.), *Entertainment Computing – ICEC 2017*, Springer International Publishing, Cham, 2017: pp. 322–334. DOI: 10.1007/978-3-319-66715-7_35.
[26] H. Iida, M.N.A. Khalid, *A Paradigm Shift from Optimal Play to Mental Comfort: A Perspective from the Game Refinement Theory*, (n.d.).
[27] L. Jiangzhou, A. Primanita, M.N.A. Khalid, H. Iida, Analyzing the Improvement Process of Table Tennis Using the Game Refinement Theory, in: *Proceedings of the Sriwijaya International Conference on Information Technology and Its Applications* (SICONIAN 2019), Atlantis Press, Palembang, Indonesia, 2020. DOI: 10.2991/aisr.k.200424.067.
[28] Y. Lu, W. Li, W. Li, Official International Mahjong: A New Playground for AI Research, *Algorithms* 16 (2023) 235, https://doi.org/10.3390/a16050235.
[29] ailab-pku, *Chinese-Standard-Mahjong* (n.d.). https://github.com/ailab-pku/Chinese-Standard-Mahjong (accessed October 30, 2024).
[30] Mortal, (n.d.). https://github.com/Equim-chan/Mortal (accessed October 30, 2024).
[31] NAGA, (n.d.). https://dmv.nico/ja/articles/mahjong_ai_naga/ (accessed October 30, 2024).