Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Artificial Intelligence and Analytics for Evaluating Trick-Taking Games: A Narrative Review with a Case Study on an Estimation Variant

Amr Hussein Reda Aly Lotfy
amarhussein262@gmail.com

ARTICLE INFO

Received: 02 Sept 2025

Revised: 25 Oct 2025

Accepted: 03 Nov 2025

ABSTRACT

This review examined how artificial intelligence (AI) and analytics had been used to support the design and evaluation of trick-taking card games, using the author's patented Estimation variant as a case lens. A narrative approach had been followed to bring together three key themes: AI methods for decision-making under hidden information, AI-driven balancing and automated playtesting, and the use of telemetry to analyze gameplay data with limited human trials. The literature had shown that search-based agents, evolutionary algorithms, and behavioural clustering were effective in simulating play, testing fairness, and evaluating balance. These methods had then been synthesized into a practical playbook for the patented Estimation variant, outlining how simulations could filter weak rule sets and guide targeted human playtests. The review concluded that this staged, data-driven process provided a transparent and reproducible pathway for evaluating novel rule changes, ensuring fairness and skill expression while protecting intellectual property.

Keywords*:* trick-taking games; Estimation; imperfect information; game balancing; automated playtesting; narrative review; patent

Introduction

Designing and improving card games where players cannot see all the information, like trick-taking games, is difficult. In these games, players make decisions without knowing everything their opponents hold, and they must guess, plan, and adjust as the game goes on. Traditionally, new versions of such games have been tested only with small groups of people. While this helps, it is not always enough to find balance problems that appear with different numbers of players, trump rules, or changes in scoring. Recent progress in artificial intelligence (AI) shows that computers can now simulate these situations at a much larger scale, giving game designers new tools to study fairness, difficulty, and how much a game rewards skill. Work on methods like Perfect Information Monte Carlo (PIMC) and Information-Set Monte Carlo Tree Search (ISMCTS) shows that AI can still make good decisions even when information is hidden (Cowling et al., 2012; Long et al., 2010). These methods have already been tested in complex games like Bridge (Ginsberg, 2001).

One family of card games that is especially interesting is Estimation-style games, which are similar to Oh Hell. In these games, players must predict exactly how many tricks they will win, and they score points for being right. It is a system that makes the game exciting and at the same time, highly sensitive to any slight modifications in the rules. To illustrate, an increase in the number of players, altering rules on the functionality of trump suits or the scoring rules can make the game easily balanced. It is due to this that Estimation can be used to study the use of AI to assist in designing and testing new rule sets (Pagat, 2025).

Scientists have already begun studying similar games. Others involve the use of AI search together with machine learning to make decisions that are better suited in trick-taking games (Solinas et al., 2019;

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Solinas et al., 2023). New versions of Monte Carlo Tree Search have been experimented by others, which are more successful in the management of hidden information (Rebstock et al., 2024). These works demonstrate that AI can not only perform such games quite well but also can be employed as the means to assess whether new rules present fair or unfair conditions. As an example, AI can tell whether some of bidding strategies prevail in the game or whether some versions of rules are trivial or too difficult.

The other direction of research is on game balancing and automated playtesting. In this case, AIs are employed to play a large number of games and then quantify the sense of fairness and interest in the game. As an illustration, research regarding the playing card game Top Trumps has demonstrated that the use of computer simulation can be employed to balance out decks to make them fair (Volz et al., 2016). Likewise, AI agents can be used to playtest different games and simulate various player behavioral patterns and determine whether a game is fun and balanced or not, before human players even get a chance to play (Holmgård et al., 2018). These methods suggest a clear workflow: run large computer simulations, collect data such as win rates or how often bids succeed, and then use these results to adjust the rules.

Even though there is already useful work in this area, the research is scattered across different games and methods. Some studies focus on Bridge, others on Oh Hell, and others on general balancing techniques. What is missing is a review that brings these ideas together and shows how they can be applied to Estimation-style games, especially new versions such as those in the patented design. This review will aim to fill that gap. It will (i) explain how AI has been used in games with hidden information, (ii) describe methods for balancing and testing new rules, and (iii) show how these methods can guide the design of Estimation variants. Following best practice for narrative reviews (Baethge et al., 2019), this paper will focus on explaining ideas in clear language, connecting them to the Estimation case, and creating a practical playbook for designers who want to test and balance new trick-taking games.

Background & Case Lens: Estimation

Trick-taking card games are a type of game where each player plays one card per round, and the highest card (often depending on a trump suit) wins that round. Because players cannot see each other's hands, they make choices with hidden information, which makes these games complex and sometimes unpredictable. Small changes in rules, such as the number of players, the order of trump suits, or scoring rules, can strongly affect how fair and enjoyable the game feels. A well-known game in this family is *Oh Hell* (often called Estimation), where players must guess exactly how many rounds or "tricks" they will win, and they score only if their prediction is correct (Pagat, n.d.). This exact-bidding rule makes the game exciting, but it also makes balance difficult, since success depends not only on skill but also on luck and how rules are set.

The author's patent introduces a new version of Estimation that extends the classic rules. In this version, the game can be played with a larger number of players (five to seven instead of the usual four) and may include extra suits or new ways of defining trump. It also keeps the "not-equal" bidding rule, which forces the last player to choose a different prediction from the others. Together, these changes make the game more flexible and inclusive for larger groups, but they also raise important design questions: does the new version remain fair? Do players of different skill levels still have a chance to compete? And do the extra options create deeper strategy or simply add confusion?

These are not just casual design questions. In game research, fairness means that no player has a built-in advantage from the rules alone; skill expression means that better strategies should clearly improve a player's results; and complexity refers to whether the decisions are rich and interesting without being overwhelming. These qualities are hard to judge by hand or with a few playtests, especially in a game that depends heavily on hidden information. This is why AI tools are useful.

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Research in card games like Bridge and *Oh Hell* has shown that AI can simulate many possible deals and strategies to test how different rules affect outcomes (Cowling et al., 2012; Ginsberg, 2001; Long et al., 2010). For example, AI can play thousands of simulated games to see whether score differences are fair across different player counts or whether some bidding strategies dominate others. Other studies on automated playtesting have shown that computer agents can mimic different play styles and give early feedback about whether a game feels balanced before it is tested by humans (Holmgård et al., 2018; Volz et al., 2016).

In this review, the author's patent is treated as a case lens—a concrete example of a real-world design challenge. The patented Estimation variant highlights the kinds of issues that arise when changing the structure of a trick-taking game. At the same time, the broader literature on AI in imperfect-information games provides tools and ideas for evaluating such changes. This review will therefore connect the two: it will explain how AI has been used to study and balance games, and then show how those methods can be applied to the patented Estimation variant to guide its development into a fair, strategic, and enjoyable game.

Methods

This paper is written as a narrative review. A narrative review is a type of article that summarises what has already been written in research, but it is not as strict or exhaustive as a systematic review. Its main goal is to give readers a clear overview, highlight important themes, and buweild an argument that links the literature together (Ferrari, 2015; Grant & Booth, 2009). To keep the review reliable, the author followed the SANRA guidelines, which give standards for writing high-quality narrative reviews. These guidelines ask authors to explain the aims, describe how the search was done, use proper references, and show a logical flow of ideas (Baethge et al., 2019).

The author looked for studies in ACM Digital Library, IEEE Xplore, Scopus, and Google Scholar. Searches were limited to work published in English between 2000 and September 2025. To make sure the author did not miss important work, the author also checked the reference lists of key papers, which is a recommended step in narrative reviews (Green, Johnson, & Adams, 2006).

The author used a mix of keywords such as *"trick-taking AI"*, *"imperfect information card games"*, *"Bridge bidding AI"*, *"Oh Hell AI"*, *"automated playtesting"*, and *"game balancing"*. These terms were combined using simple search operators (AND/OR) depending on the database.

The author included only peer-reviewed articles, conference papers, and important technical reports that dealt with one of three areas: (i) AI methods for decision-making under hidden information, (ii) AI used for game balancing and automated playtesting, and (iii) telemetry or analytics for evaluating games. The author excluded sources that were not peer-reviewed, such as blogs or company reports. The author's patent on Estimation was treated only as a case example to illustrate design challenges, not as a research source itself.

From each included study, the author noted the type of game studied, the AI method used, what it measured (for example, fairness, win rates, or bidding patterns), and the main lesson. After this, the author used a thematic approach to group the studies into three themes: (1) AI for decision-making in hidden information games, (2) AI for balancing and automated playtesting, and (3) telemetry and analytics for understanding fairness and difficulty (Braun & Clarke, 2006). Because this is a narrative review, it cannot claim to include every possible study. However, by following SANRA guidelines and using a clear search process, the review remains transparent and useful for readers.

AI for Decision-Making under Hidden Information in Trick-Taking Games

In card games like Estimation or Bridge, players cannot see all the cards. This makes decision-making hard because they must choose actions without full knowledge of what others hold. These games are

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

known as imperfect-information games. Artificial intelligence (AI) techniques have been created to assist computers to make decisions in these contexts and the techniques also give the means of testing how equitable or impartial a novel version of a game could be.

Monte Carlo Tree Search (MCTS) is one of the most common ones. This is achieved through simulation of the possible plays of a game and utilisation of the results to inform the next decision. It has performed quite well in most games including board games and card games (Browne et al., 2012). The merit of MCTS is that it is able to examine alternative strategy without perfect assessment of the ideal action beforehand.

A simpler form of hidden card games, which is commonly used by researchers, is determinization or Perfect-Information Monte Carlo (PIMC). In this case, the program is guessing what the cards are by simply filling in the gaps at random and then they play the game as though they were visible. Repeating it numerous times it can determine what move is the best in the average. This approach can surprisingly be effective (Long et al., 2010). However, it can also make errors because it treats each guess as if it were the real game. For example, it might mix strategies that a human would never play because they depend on knowing information that should be hidden (Lisý et al., 2015).

To reduce these problems, researchers developed Information-Set MCTS (ISMCTS). Instead of pretending the hidden cards are known, ISMCTS searches directly within the information available to the player. This makes it a better fit for trick-taking games, where reasoning is based on what a player knows and what can be guessed from bids and previous plays (Cowling, Powley, & Whitehouse, 2012).

Strong computer play in Bridge shows how these ideas work in practice. The program GIB was an early example of an AI that could make strong decisions even when it did not know all the cards. It integrated search with ingenious methods of dealing with uncertainty and demonstrated that computers were competitive using sophisticated trick taking games (Ginsberg, 2001).

Other scholars have even taught AI systems to learn how to make more impressive guesses using previous game history. To illustrate the case, when player leads a given suit, or refuses to play a large card, this provides some indication about his/her hand. The study conducted by Solinas and their colleagues (2019) revealed that search combined with supervised learning on these histories can assist the AI to take stronger decisions. More recent work examined how to filter and simplify these histories so that the computer can focus only on the important parts, making the process more efficient (Solinas et al., 2023).

A newer approach is to plan directly in the observation space—that is, to work only with what the player has actually seen, not with imaginary full deals. Generative Observation MCTS (GO-MCTS) does this by combining MCTS with generative models (such as transformers) that predict what might happen next. This avoids some of the main weaknesses of PIMC and has been tested successfully in trick-taking games like Hearts, Skat, and The Crew (Rebstock et al., 2024).

So what does all this mean for Estimation-style games? These AI methods provide different levels of support for testing new rules:

- **PIMC** agents are easy to build and can give a quick sense of how often bids succeed or fail.
- **ISMCTS** agents provide more accurate results by respecting hidden information.
- **History-aware agents** make the simulations more realistic by using clues from bidding and play.
- **Observation-based methods** like GO-MCTS reduce bias and may give the clearest picture of balance.

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

By running thousands of simulated games with these agents, designers can collect data such as win-rate spreads, the stability of bidding, and whether some strategies are too dominant. This allows them to see if changes, like adding more players or new trump rules, make the game fairer, harder, or more enjoyable. Importantly, this can all be done before human playtests, saving time and effort while still giving useful insights.

In short, the literature shows a clear path: computers can already make good decisions in trick-taking games, and the same tools can be used not just to play the game but to evaluate and improve game design. For the author's patented Estimation variant, these methods provide a way to test fairness, skill, and difficulty in a structured way.

AI for Game Balancing and Ruleset Evaluation

Balancing a game means adjusting its rules so that players feel the experience is fair, challenging, and enjoyable. In traditional design, balance was usually tested by trial and error through small playtests. However, recent studies show that artificial intelligence can support this process by running thousands of simulated games and measuring whether the rules create equal chances and interesting outcomes. This allows researchers and designers to test changes more quickly and with more evidence. According to Yannakakis and Togelius (2018/2024), AI may serve in games to generate agents to play, as well as a design partner that assesses the fairness of the game, its difficulty, and experience of players in a systematic manner.

A good example of this is the research on the card game Top Trumps. Volz and colleagues (2016) demonstrated their ability to automate balancing by specifying a set of goals, including fairness in winning opportunities and excitement in play, and then to apply evolutionary algorithms to probability sampling through a large number of possible decks. By these criteria, the computer-generated decks were as good or in other cases better than human-designed decks. The significance of this study is that it proves that automatic balancing is not only a theory but also works in practice with actual card games.

The balance is also concerned with the experience of various kinds of players to a game. Procedural personas were designed by Holmggaard et al. (2018) and can be described as artificial intelligence agents that can simulate the play style of a cautious or an aggressive player. These personas play the game several thousand times and point at the inappropriateness of some strategies as being overpowering or having too few strengths. This allows one to determine whether a rule set is biased towards a certain type of player by mistake. They demonstrated with their work that even before human players enter the game, automated playtesting can provide useful information.

The same approach has been applied to more complex games. For example, García-Sánchez et al. (2018) studied *Hearthstone*, a collectible card game with many possible deck combinations. They used evolutionary algorithms to automatically generate and test decks, identifying unbalanced or dominant strategies. This shows that AI-based playtesting is not limited to simple games—it can also help in large, complex systems where manual balancing would be slow and expensive.

These findings matter for Estimation-style games because the author's patented version introduces changes like more players and new trump rules. AI balancing methods can be used to test these changes. For example, simulations can measure whether win rates remain fair when the number of players increases or whether the "not-equal" bidding rule creates unexpected advantages. Personas can also test if aggressive bidding strategies or conservative styles are unfairly rewarded under new rules. By applying these techniques, designers can narrow down the most promising versions of the rules before inviting humans to test them.

At the same time, researchers caution that AI is not a replacement for human testing. Automated playtesting is best used as an early filter to identify problems and guide design decisions. Final evaluation must still involve real players to capture social dynamics and the human experience of fun.

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

Holmgård et al. (2018) and Yannakakis and Togelius (2018/2024) both stress that AI is most useful when combined with human feedback in a staged process. In this way, AI helps save time and effort while still leaving space for the creativity and judgment of designers and players.

Telemetry and Analytics with Minimal Human Trials

Telemetry is the process of collecting data during gameplay, such as moves, choices, timings, and results. In modern game research, telemetry is important because it allows designers to see how players actually interact with rules. Seif El-Nasr, Drachen, and Canossa (2013) explain that game analytics can turn these logs into clear insights for improving design, by showing patterns such as where players struggle or how strategies change during play. Yannakakis and Togelius (2024) also point out that AI can be combined with telemetry to model player behaviour and measure performance in ways that guide design decisions.

A key benefit of telemetry is that it can reduce the need for long or repeated human playtests at early stages. Instead, AI agents can act as test players. Holmgård et al. (2018) showed that procedural personas—computer-controlled players with different styles—can generate large amounts of play data. By studying this data, designers can detect problems such as strategies that are too strong or rule changes that make the game unfair, before running human tests. Other work also demonstrates that automated playtesting can create useful datasets for analyzing game balance in a faster and cheaper way (Mugrai et al., 2019).

Once data is collected, it needs to be analyzed. Drachen et al. (2012, 2014) used clustering techniques to group players depends on how good or bad they pla, using telemetry logs from real games. Their work shows that behavioural clustering can reveal different play styles and highlight how certain rules or settings lead to specific types of behaviour. This is valuable in trick-taking games, where patterns in bidding or trick play might show that one style dominates the game.

Telemetry is also useful for measuring skill. One of the most cited systems is TrueSkill, created by Microsoft for Xbox Live. TrueSkill uses the results of matches to estimate player skill, even in games with many players or teams (Herbrich, Minka, & Graepel, 2007). Later work improved this method by including extra information, such as player history and behaviour, to give more accurate ratings (Minka et al., 2018). Such methods demonstrate that detailed records of wins, losses, and behaviours are sufficient to make a sound judgment as to whether a game is skilled or a game of luck.

Telemetry may also be used in the case of Estimation and its new patented version. Recording the bids, tricks won and final scores of lots of simulated games allow designers to verify that the game is fair with five or seven players, or that the game does not advantage the player with the so-called not-equal-bidding rule. They also can analyze whether there is a consistent and stronger performance of the AI agents as compared to weaker agents, which would demonstrate that the game does not devalue skill. Based on the literature, the ideal practice is to take as a primary filter these telemetry results and subsequently validate potential promising rule sets by human playtests (Holmgård et al., 2018; Seif El-Nasr et al., 2013).

To conclude, telemetry and analytics allow one to test game balance with less human pre-testing. The fairness, skills expression, and behaviour patterns can be learned out of logs produced by automated agents. The available literature indicates that the method is successful in most types of games and it can be directly transferred to the Estimation-style games to inform a systematic and effective approach to design.

Synthesis: Linking AI, Balancing, and Telemetry for the Estimation Variant

The three themes explored in this review are AI for decision-making under hidden information, AI for balancing and ruleset evaluation, and telemetry with minimal human trials and these themes fit

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

together as parts of one design process. Each theme looks at the game from a different angle, but when combined they form a structured approach for improving and validating new trick-taking games like the author's patented Estimation variant.

The first theme showed how AI can make choices in games where players cannot see all information. Methods such as Perfect-Information Monte Carlo (PIMC) offer quick but imperfect estimates, while Information-Set MCTS (ISMCTS) and newer observation-space planning methods provide more faithful models of uncertainty (Cowling et al., 2012; Rebstock et al., 2024). These approaches mean that instead of guessing how a ruleset will behave, designers can let AI agents play thousands of games and observe patterns in bidding success or trick outcomes. Even though these agents are not perfect models of human thought, they create a baseline that is faster and more systematic than relying only on small playtests (Browne et al., 2012; Long et al., 2010).

The second theme introduced the concept of systematizing rules. Balancing does not just pertain to fairness but also to making play exciting and not to use one strategy to dominate the other. Previous studies demonstrated that cards games can be equilibrated by means of the evolutionary algorithms that seek rule or deck configurations that both maximize fairness and pleasure (Volz et al., 2016). Procedural personas were used by other players AI players who followed a particular play style to determine whether a rule set benefits one style over the other (Holmgård et al., 2018). More complex games like Hearthstone have automated deck gen and testing to discover concealed imbalances (García-Sánchez et al., 2018). Those studies imply that the Estimation version might also be experimented using a combination of simulated players and optimization techniques to identify rule settings that are stable and enjoyable in varying circumstances.

The third theme was the fact that telemetry and analytics transform huge amounts of gameplay logs into design information. Recording bids, scores, and outcomes of simulations, designers can cluster based on those to see that there are recurring patterns in play(Drachen et al., 2012, 2014) and they can use rating systems such as TrueSkill to determine whether stronger agents win at higher rates (Herbrich et al., 2007; Minka et al., 2018). According to research of game analytics, even a simple event logs can be used to create useful dashboards that inform design decisions (Seif El-Nasr, Drachen, and Canossa, 2013). This implies that in the Estimation variant, the concept of fairness and skill can not just be evaluated based on raw outcomes but also on more intricate criteria like the frequency of close games, the variety of strategies in bidding and whether minor modifications to the rules can be expected to yield similar outcomes.

Collectively, these themes imply that there could be a step-by-step playbook of the testing of the Estimation variant. To begin with, establish the goals of the new rules: equitable opportunities to all five to seven players, further expression of skills and intriguing scoring. Second, run simulations with search-based agents, starting simple and moving to more advanced methods that better capture hidden information. Third, add balancing methods by using personas and optimization to compare rule sets. Fourth, log telemetry from these games and analyze it for fairness, skill separation, and stability. Finally, confirm the most promising variants through small, focused human playtests.

This combination has clear strengths. It reduces reliance on intuition, gives measurable indicators of design quality, and saves time by filtering out weak rule sets before human testing. It also connects directly to published research, showing that the approach is grounded in methods already used in other card and strategy games. At the same time, the literature cautions about limitations. Automated agents cannot fully model human creativity or the social aspects of multiplayer games, which means human playtests remain essential at the final stage (Holmgård et al., 2018; Yannakakis & Togelius, 2024). Another challenge is deciding which metrics best represent fun and fairness; while win rates and bid accuracy are helpful, they do not capture everything about player experience.

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

For future research, this framework could be extended by studying online play telemetry once the Estimation variant is released to real players. Data from live sessions could test whether the variant scales well in casual and competitive settings, and whether the "not-equal" bidding rule adds excitement without frustration. Combining simulation-based results with real-world analytics would create a complete cycle of design, testing, and refinement.

In summary, the three themes link into one pathway: AI decision-making gives a foundation for large-scale simulation, balancing methods provide systematic ways to compare rules, and telemetry translates results into clear insights. Applied to the author's Estimation patent, this process makes it possible to show that the game is not just novel but also tested and fair, making it strong enough for journal publication.

Ethical/IP

This review uses the author's patent as an example, but only the information that is already public in the patent is discussed. Nothing confidential or unpublished should be included. The paper should also have a short note explaining that the author is the inventor of the patent. If the journal uses blind review, the patent can be cited as "Author's Patent (year)" during review and then replaced with the full reference later.

Since this review is based on published studies and computer simulations, no ethics approval is needed at this stage. But if the work later includes human playtests, then proper ethics approval, informed consent, and data privacy rules will be required. If telemetry from real players is collected, it should be explained clearly to players, stored safely, and anonymised before analysis.

Finally, any third-party materials used (for example, datasets, code, or images) should follow their licenses. If code or synthetic data is shared, it should not reveal anything beyond what is already in the patent. This way, the work stays transparent while protecting intellectual property and respecting player privacy.

Limitations

This review is narrative, not systematic. That means it gives a wide overview but may not cover every single study. The simulation methods discussed are also conceptual—they describe how AI agents could be used, but actual experiments were not run here. Another limit is that computer agents cannot fully capture how real people play. Social interaction, creativity, and enjoyment are all important in games but are hard to model with AI.

The measures suggested—fairness, skill expression, and bid success—are useful, but they cannot capture the full idea of "fun." Also, the focus on trick-taking games makes the review less general for other types of games. Using the patent as the case example narrows the scope further. For these reasons, the paper is best seen as a design guide and starting point, rather than as final proof of how the new rules work.

Future Directions

The next step is to build the simulation framework described here. Start with simple AI agents, then add more advanced agents that handle hidden information more realistically. Making a small package of rules, code, and data would also help others repeat and test the work.

After that, run small human playtests. These could compare two or three of the most promising rule sets from the simulations. Feedback should focus on fairness, clarity of rules, and enjoyment. If the game is played online, collecting telemetry data from real players could give insights into how the variant works over time.

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

Methodologically, researchers could refine fairness and excitement measures, test how sensitive results are to small rule changes, and explore whether different play styles (such as cautious vs. risk-taking) remain balanced. Comparing the patented Estimation variant with other bidding games could also show where it adds something new. In the long run, building a shared benchmark for trick-taking games would allow other researchers to test their methods and agents on the same tasks.

Conclusion

This review reframes the author's Estimation patent as a case study for how AI and analytics can support game design. It connects three themes: AI methods for decision-making in hidden-information games, automated balancing and playtesting, and the use of telemetry to read results. Together, these methods create a structured path from design idea to tested, publishable game rules.

The suggested process is simple but effective: define goals, simulate large numbers of games with different agents, log outcomes, and analyze fairness, skill, and stability. The strongest variants can then be confirmed with targeted human tests. This staged approach saves time, protects intellectual property, and produces clear evidence that can be shared with the research community.

Although simulations cannot replace the creativity and social elements of human play, they provide a valuable starting point. With further testing, the Estimation variant can be presented as not only novel but also tested, balanced, and ready for wider use.

References

[1] Baethge, C., Goldbeck-Wood, S., & Mertens, S. (2019). SANRA—a scale for the quality assessment of narrative review articles. *Research Integrity and Peer Review, 4*(1), 5. https://doi.org/10.1186/s41073-019-0064-8

[2] Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://doi.org/10.1191/1478088706qp063oa

[3] Browne, C. B., Powley, E., Whitehouse, D., Lucas, S. M., Cowling, P. I., Rohlfshagen, P., Tavener, S., Pérez, D., Samothrakis, S., & Colton, S. (2012). A survey of Monte Carlo Tree Search methods. *IEEE Transactions on Computational Intelligence and AI in Games, 4*(1), 1–43. https://repository.essex.ac.uk/4117/1/MCTS-Survey.pdf

[4] Cowling, P. I., Powley, E. J., & Whitehouse, D. (2012). Information Set Monte Carlo Tree Search. *IEEE Transactions on Computational Intelligence and AI in Games, 4*(2), 120–143. https://eprints.whiterose.ac.uk/75048/1/CowlingPowleyWhitehouse2012.pdf

[5] Drachen, A., Sifa, R., Bauckhage, C., & Thurau, C. (2012). Guns, swords and data: Clustering of player behavior in computer games in the wild. *2012 IEEE Conference on Computational Intelligence and Games (CIG)*, 163–170. https://doi.org/10.1109/CIG.2012.6374152

[6] Drachen, A., Thurau, C., Sifa, R., & Bauckhage, C. (2014). A comparison of methods for player clustering via behavioral telemetry. *arXiv preprint*. https://arxiv.org/abs/1407.3950

[7] Ferrari, R. (2015). Writing narrative style literature reviews. *Medical Writing, 24*(4), 230–235. https://doi.org/10.1179/2047480615Z.000000000329

[8] García-Sánchez, P., Tonda, A., Mora, A. M., Squillero, G., & Merelo, J. J. (2018). Automated playtesting in collectible card games using evolutionary algorithms: A case study in Hearthstone. *Knowledge-Based Systems, 153*, 133–146. https://doi.org/10.1016/j.knosys.2018.04.021

[9] Ginsberg, M. L. (2001). GIB: Imperfect information in a computationally challenging game. *Journal of Artificial Intelligence Research, 14*, 303–358. https://jair.org/index.php/jair/article/view/10279/24508

[10] Grant, M. J., & Booth, A. (2009). A typology of reviews: An analysis of 14 review types and associated methodologies. *Health Information & Libraries Journal, 26*(2), 91–108. https://doi.org/10.1111/j.1471-1842.2009.00848.x
[11] Green, B. N., Johnson, C. D., & Adams, A. (2006). Writing narrative literature reviews for peer-reviewed journals: Secrets of the trade. *Journal of Chiropractic Medicine, 5*(3), 101–117. https://doi.org/10.1016/S0899-3467(07)60142-6
[12] Herbrich, R., Minka, T., & Graepel, T. (2007). TrueSkill™: A Bayesian skill rating system. *Advances in Neural Information Processing Systems*, 20, 569–576. http://papers.nips.cc/paper/3079-trueskilltm-a-bayesian-skill-rating-system.pdf
[13] Holmgård, C., Green, M. C., Liapis, A., & Togelius, J. (2018). Automated playtesting with procedural personas through MCTS with evolved heuristics. *IEEE Transactions on Games, 11*(4), 352–362. https://doi.org/10.1109/TG.2018.2876205
[14] Lisý, V., Lanctot, M., & Bowling, M. (2015). Online Monte Carlo counterfactual regret minimization for search in imperfect information games. *Proceedings of AAMAS 2015*, 27–36. https://mlanctot.info/files/papers/aamas15-iioos.pdf
[15] Long, J., Sturtevant, N., Buro, M., & Furtak, T. (2010). Understanding the success of Perfect Information Monte Carlo sampling in game tree search. *Proceedings of AAAI*. https://webdocs.cs.ualberta.ca/~nathanst/papers/pimc.pdf
[16] Minka, T., Cleven, R., Zaykov, Y., Dang, K. D., & Chen, W.-T. (2018). TrueSkill 2: An improved Bayesian skill rating system. *Microsoft Research Technical Report*. https://www.microsoft.com/en-us/research/wp-content/uploads/2018/03/trueskill2.pdf
[17] Mugrai, L., Szubert, M., & Johanson, M. (2019). Automated playtesting of matching tile games. *2019 IEEE Conference on Games (CoG)*, 1–8. https://doi.org/10.1109/CIG.2019.8848085
[18] Pagat. (2025, July 22). *Oh Hell! – Card game rules*. https://www.pagat.com/exact/ohhell.html
[19] Rebstock, D., et al. (2024). Generative Monte Carlo Tree Search for imperfect-information environments. *arXiv preprint*. https://arxiv.org/pdf/2404.13150
[20] Rebstock, D., Solinas, C., Sturtevant, N. R., & Buro, M. (2024). Transformer-based planning in the observation space with applications to trick-taking card games (GO-MCTS). *arXiv preprint*. https://arxiv.org/abs/2404.13150
[21] Seif El-Nasr, M., Drachen, A., & Canossa, A. (Eds.). (2013). *Game Analytics: Maximizing the Value of Player Data*. Springer. https://doi.org/10.1007/978-1-4471-4769-5
[22] Solinas, C., et al. (2019). Improving search with supervised learning in trick-based card games. *AAAI Conference on Artificial Intelligence*. https://cdn.aaai.org/ojs/3909/3909-13-6968-1-10-20190702.pdf
[23] Solinas, C., Rebstock, D., Sturtevant, N. R., & Buro, M. (2023). History filtering in imperfect information games: Algorithms and complexity. *NeurIPS 2023*. https://openreview.net/forum?id=inIONNg8Sq
[24] Volz, V., Rudolph, G., & Naujoks, B. (2016). Demonstrating the feasibility of automatic game balancing. *Proceedings of the Genetic and Evolutionary Computation Conference (GECCO '16)*, 269–276. https://doi.org/10.1145/2908812.2908913
[25] Yannakakis, G. N., & Togelius, J. (2018/2024). *Artificial Intelligence and Games* (1st & 2nd eds.). Springer. https://gameaibook.org/wp-content/uploads/2024/08/book2.pdf
[26] Yannakakis, G. N., & Togelius, J. (2024). *Artificial Intelligence and Games* (2nd ed.). Springer. https://gameaibook.org/wp-content/uploads/2024/08/book2.pdf