Chance and Skill in Games

Master Research Project

Eva van Agt, María Diez Pérez,
Max Knappe, Renske Nijssen (Group 14)

Supervisors: Dr. Cameron Browne and Dr. Matthew Stephenson

June 23, 2021

1 Introduction

Uncertainty plays a vital role in games. In fact, a game would not be a game if it was free of uncertainty: a game's outcome is per definition uncertain. Playing a game which you know for sure you will win might be somewhat fun, but if you know beforehand you are going to lose, why even bother starting it?

If a game's outcome is not known beforehand, what does it depend on? On the one hand, a game's outcome could depend on players' levels of skill. On the other hand, chance elements might play a role. 'Chance' can appear in several forms and will mostly be referred to as 'uncertainty' from this stage forward. Note that this refers to uncertainty *throughout* the course of a game; when discussing uncertainty in a game's outcome, it will be explicitly made clear.

How much skill and uncertainty contribute to the outcome of a game, varies per game. This project aims to quantify the contribution of both of these elements for several games in Ludii: a general game system designed to play, evaluate, and design many games, such as board games, card games, dice games, and mathematical games. It is designed and lead by Cameron Browne, and part of the Digital Ludeme Project hosted by Maastricht University (Maastricht University, 2021a). More precisely, the contribution of uncertainty and skill will be given in the form of a chance/skill ratio, which will be number between zero (pure chance) and one (pure skill).

Several methods have been developed to estimate the chance/skill ratio of games. The objective of this project is to adapt some of these methods and apply them to the Ludii game system, so that at least a couple of games will be provided with a chance/skill ratio. These methods can then later be used to provide all games with such a measure.

The following research questions will serve as a guide throughout the paper:

1. How do we define uncertainty in games and its relation to chance and skill?

2. How do we measure the balance of these aspects in general games?

3. How can our results be validated?

Question 1 will be answered in section 3.1. Section 4 will cover question 2. For simplification, only two-player games are considered in this report. The answer to question 3 can be found in section 6.1.

2 Motivation

2.1 Digital Archaeoludology

One of the aims of the Digital Ludeme Project is to increase the understanding of the development of traditional games throughout human history, using modern computational techniques. Browne has coined this field of study 'Digital

Archaeoludology' (Browne and Piette, 2019). One of the questions this area of research attempts to answer, is: *What aspects are important for a game to survive many generations?* This question can also be asked for different versions of within games. For example, the Roman game XII Scripta appears to be an ancestor of Backgammon (Maastricht University, 2021b). While XII Scripta is unknown to most people, Backgammon is very popular. Why did Backgammon stand the test of time, and XII Scripta did not? Perhaps this has to do with the difference in chance/skill ratios of both games. Although it is hard to say for sure which one of the two involves more skill. Since Backgammon only uses 2 dices while XII Scripta uses 3 dices, it seems that Backgammon has a higher skill component than XII Scripta. This agrees with the hypothesis of this project: games that have survived many generations generally have a high skill component.

2.2 Ludii General Game System

In addition, the Ludii general game system can benefit from chance/skill ratios assigned to its games. Players that use the game system have their own preferences: some players would like to play games that involve much chance, while others like to have more control over a game. When all games in Ludii have their own chance/skill ratios, players can pick games with the ratio they desire.

2.3 Gambling Legislation

Another reason why the chance/skill ratio of a game is important, is for gambling legislation. Most countries have a restriction for gambling, but what is gambling? How much chance must be involved in a game to qualify as a gambling game? Or put differently, when is the skill element in a particular game important enough so that the game is not subjected to the rules and regulations regarding gambling?

The answer to these questions has not been undisputed. For example, the Dutch government classifies poker as a game of chance, in which players have "no control over the outcome" (Rijksoverheid, 2021). In contrast, the Federal District Court in Brooklyn ruled in 2012 that poker is more a game of skill than a game of chance (Secret, 2012). Due to poker's popularity, much research has been done in estimating the chance/skill ratio of this game. Borm and van der Genugten (2016) found that Texas Hold'em, in all its practical variants, should be considered as a game of skill, rather than of chance.

If every game has its own chance/skill ratio, it is clearer whether a particular game has to be considered as a gambling game, and thus whether it should comply with gambling legislation.

3 Background

3.1 Definition of Uncertainty

Uncertainty can appear in many different forms. Costikyan (2013) divides uncertainty into nine categories. Since his book focuses mostly on video games, not all sorts apply to the Ludii game system. For example, narrative anticipation, development anticipation, and schedule uncertainty are not relevant for any games in Ludii. Furthermore, performative uncertainty, solver's uncertainty, and analytic complexity are types of uncertainty that strongly correlate with the physical and mental skills a player possesses. In order to simplify this project, these types will be regarded as skill rather than chance. What remains are three types of uncertainty. Randomness is a fundamental form of uncertainty that can be regarded as chance. Examples of random events include dice rolls, coin flips, and the dealing of cards. Next, player unpredictability can appear in multiplayer games. This is the case when players do not know their opponents' strategies and depend on them to some degree. Closely related to this type of uncertainty is the notion of hidden information. In poker, for example, players try to deduce which cards their opponents hold.

Although Costikyan's categorization of uncertainty is powerful due to its completeness, this project will focus only on randomness, player unpredictability and hidden information. In fact, these types can be further combined into two different categories, following the approach by Dreef et al. (2004). Randomness is equivalent to external chance elements, while player unpredictability and hidden information together comprise internal chance elements. The latter are generated by players who use mixed strategies. The aim of the project is to quantify the role of external chance elements for a given game in Ludii.

3.2 Previous Work: Estimating Chance/Skill Ratios

Measuring skill and uncertainty in games is a well discussed topic. Previous work has proposed several approaches to methodically determine the chance/skill ratio of games.

3.2.1 Cheating player discovering external chance elements

Borm and van der Genugten propose a formula to calculate the amount of skill in any game with n players. The types of players involved consist of a beginner, a real advanced player, and a virtual advanced player. In contrast to the real advanced player, the virtual advanced player knows the outcome of all external chance elements included in the game, such as dice rolls, random actions, coin flips, etc. This player can be thought of as a 'cheating' player. The results all players achieve in the game are set into relation in the formula below, which results in a number between zero and one representing the chance/skill ratio in the given game.

$$skill = \frac{\text{result real advanced player} - \text{result beginner}}{\text{result virtual advanced player} - \text{result beginner}} \tag{1}$$

The numerator of this fraction represents the *learning effect*: the difference between an advanced player and a beginner. The denominator represents the sum of the learning effect and the *random effect*. The random effect directly relates to the amount of external chance elements involved in the game (Borm and Genugten, 2001). In summary:

$$skill = \frac{\text{learning effect}}{\text{learning effect} + \text{random effect}} \tag{2}$$

3.2.2 Cheating player discovering internal chance elements

Following the method by Borm and van der Genugten, a game such as rock-paper-scissors would be labeled as a pure skill game. This is because the random effect is zero: a virtual player, knowing the outcome of external chance elements, does not perform better than a non-cheating player. Intuitively, however, rock-paper-scissors seems to involve a significant chance element. One could of course argue that a player can somewhat increase their winning probability by trying to read its opponent's body language, but it seems incorrect to claim that rock-paper-scissors solely depends on skill.

Dreef, in cooperation with Borm and van der Genugten, addresses this issue three years later in an updated paper. Using a game-theoretic approach, the authors succeed in taking internal chance elements into account which occur when players use mixed strategies. (Dreef et al., 2004)

3.2.3 50%-chess

Duersch et al. (2017) propose a measure in their paper "Measuring Skill and Chance in Games" that provides a 50%-benchmark for the predominance of chance versus skill. They first observe that the more skill is involved in a game, the larger the standard deviation of players' ratings is. A player's rating in this context is a measure of a player's strength with respect to a particular game. To demonstrate their point, player ratings for chess follow a wider distribution than for poker.

Their most interesting contribution is the invention of '50%-chess', which is used as a benchmark signifying exactly half pure chance and half pure skill. The game 50%-chess is constructed by replacing every other outcome of a chess game with a coin toss. Note that chess is regarded to be a game of pure skill, while a coin toss solely depends on chance. The standard variation in players' ratings for 50%-chess is then used as a comparison with the standard deviations of other games. If a game's standard deviation is higher than that of 50%-chess, it is more based on skill than on chance, and vice versa. This way, the authors found that for instance poker and backgammon predominantly depend on chance.

To measure a player's rating, the authors use the ELO rating which is well-established and can be applied to any game. Large datasets of various two-player games were acquired and used for analysis.

4 Methods

4.1 Computing Chance/Skill Ratios

4.1.1 Cheating player discovering external chance elements

Borm and van der Genugten calculate the chance/skill ratio using three types of players: a virtual advanced player, a real advanced player, and a beginner. These types can be modeled by agents available in Ludii. Initially, the idea was to let all three types of players play against an 'average' AI. However, the definition of an 'average' AI is not so straightforward, and differs per game. For example, for Chess, an average AI needs to be much smarter than for a game like Tic-Tac-Toe. Therefore, it was decided to only model the virtual advanced player and the real advanced player, and to let these two types play against each other. These players are modeled by agents using the Alpha-Beta search algorithm and the UCT algorithm (a special case of Monte Carlo Tree Search). The virtual advanced player has the additional knowledge of the outcome of all external chance elements, and has been implemented by Dennis Soemers, a Ludii developer.

The idea is that for games that involve external chance components, the virtual advanced player will perform better than the real advanced player. However, note that this is only the case if the virtual advanced player is able to influence its decisions based on the knowledge of the external chance elements. For example, while there are external chance elements in Pasa (dice), the virtual advanced player is not able to benefit from its knowledge about the outcome of its dice rolls, since there are no choices to be made by the agent. This is different in Royal Game of Ur: here, the agent is able to choose which piece to move.

Therefore, the usage of this method is limited to games that involve external chance elements and for which an agent can choose a variety of options while playing the game. This means that the method is only suitable for games that are not pure skill (pure skill games do not involve external chance elements), but also not pure chance (in these games agents do not have options to choose from).

4.1.2 Choosing random (opponent's) moves

Additionally, a method was suggested by Cameron Browne and Matthew Stephenson where an AI playing a game would not use the standard Alpha-Beta search algorithm, but instead choose a random move for its opponent in the search tree. This can be done at every stage in the search tree (resulting in a completely random agent), or only n% of the time. Note that if n=100%, then the agent is random, and if n=0%, the agent uses the original Alpha-Beta search algorithm.

If the result that a player obtains when choosing random moves for its opponent is the same (or better) as when applying the Alpha-Beta algorithm, then this is an indication that the game has a considerable chance element. If a player performs better when not choosing random moves for its opponent, then there must be an element of skill involved. Hence, the difference in the outcome for games played in the 'randomized' way and the standard Alpha-Beta algorithm should describe a game's reliance on chance elements.

Note that for some games, the Alpha-Beta search algorithm is not even defined, or its heuristics are bad. In addition, the optimal AI agent (the Ludii AI) uses the UCT algorithm in most games. Hence, experiments with agents that use the UCT algorithm have also been carried out. Rather than choosing random moves for its opponents, agents now play randomly n% of the time. Again, if n=100%, the agent is fully random, and if n=0%, the agent is equivalent to the original UCT agent.

4.1.3 Standard Deviations of ELO Ratings

The Elo (1978) Rating is a method to calculate the relative skill level of players in zero-sum games. The system was originally invented by Arpad Elo to calculate the skill level for chess, but is nowadays also used as a rating system for sports such as American football, basketball and baseball, and for board games such as Scrabble and Diplomacy.

In the ELO Rating system, if player A has an ELO Rating that is 400 points greater than player B, then player A should be 10 times more likely to win the game. More generally, if player A has rating R_A and player B has rating R_B, the expected score of player A is $E_A = \dfrac{1}{1 + 10^{(R_B - R_A)/400}}$.

When a game is played, the ELO Rating of both players will change. The new rating of player A, denoted by R^*_A is given by the following formula, with the K_A-factor a positive constant that denotes the maximum possible rating adjustment. This factor varies based on the player's rating (a smaller K-factor is used for players with a higher rating):

$$R^*_A = \begin{cases} R_A + K_A(1 - E_A) & \text{if A wins} \\ R_A + K_A(1/2 - E_A) & \text{if A draws} \\ R_A + K_A(0 - E_A) & \text{if A loses} \end{cases}$$

Duersch et al. (2017) observed that the more skill is involved in a game, the larger the standard deviation of players' ELO ratings is. This approach requires a database with the ELO Ratings for a huge amount of players.

Board Games Arena is an online game platform with more than 6,5 million users from all over the world. For each game there is a ranking list that is based on the ELO Rating. There are about 23 games that are both available in Ludii and on this online game platform. Hence, for these games, the standard deviation can be calculated.

4.2 Test Games

4.2.1 Tic-Tac-Toe: 100% skill

Tic-Tac-Toe is a particularly useful game to analyze for this research due to its relatively small search tree and its modifiability into several variants. The original version is a pure skill game in the sense that there are no external chance elements. The drawback of this version is that the board is so small, that weaker agents can win or draw against stronger agents by accidentally playing optimal strategies. The chance of this happening is rather large. To fight this limitation, a version of Tic-Tac-Toe which uses a 9x9 board was designed. In this version, an agent needs to place six pieces in a row in order to win the game. Note that the chances of accidentally following an optimal strategy are now much smaller for weak agents, since it has nine times as many options to place its pieces.

4.2.2 Tic-Tac-Die: 0% skill

Tic-Tac-Die is another variant on Tic-Tac-Toe. Instead of being able to choose where to place their pieces, the agents now have to roll a die which decides where to put their pieces. If the place indicated by the die is already occupied by a piece, the player must re-roll until an empty square is rolled. Clearly, no element of skill is involved; all actions fully depend on the rolls of the die. Again, this game can be extended to a 9x9 grid, although this should not affect the distributions of wins and losses between the players.

4.2.3 Tic-Tac-Toe-Random: n% skill

Tic-Tac-Toe-Random is a variant devised by Matthew Stephenson, which combines Tic-Tac-Toe and Tic-Tac-Die. Each time a player is to move, with n% chance it is allowed to pick a square itself, and with 100-n% chance it must (re-)roll the die which decides where to place its piece on the board. Note that if n=100%, then the game is equivalent to Tic-Tac-Toe; if n=0%, then the game is equivalent to Tic-Tac-Die. As with the previously mentioned versions, the board of this game can also be enlarged to a 9x9 grid.

4.2.4 Chess: 100% skill

Chess does not contain any external chance elements. Since its search tree is so large, even strong agents might accidentally follow the 'wrong' paths when searching for the optimal strategies. In that sense, agents face some uncertainty which is of the form of analytic complexity. However, as described in section 3.1, this type of uncertainty is considered to be skill rather than chance. Therefore, the chance/skill ratio of chess is determined to be 0.

4.2.5 Pasa: 0% skill

Pasa is a dice game where the two players are trying to get hundred points. Players roll a seven-sided die, one side is worth 10 points, one side costs 10 points and for the other sides the players receive 1 till 5 points. Therefore the player who reach hundred points at first is the winner of the game. The players are completely dependent on the number on the die, so there is no skill involved.

4.2.6 Royal Game of Ur: x% skill

Royal Game of Ur contains external chance elements in the form of dice rolls, but players can choose which pieces to move. Hence, there is some degree of skill involved. As explained in section 4.1.1, this game is extremely suitable for testing the method involving the virtual advanced player.

4.2.7 Backgammon: x% skill

As mentioned earlier, from a historical standpoint it would be interesting to see whether Backgammon or XII Scripta has a higher skill component according to our methods. The hypothesis is that Backgammon requires more skill than XII Scripta, since it involves less external chance elements. When performing experiments with XII Scripta, it soon became clear that these experiments were not feasible: running a single playout took several hours. Therefore, only Backgammon has been used in the experiments.

5 Results

5.1 Cheating player discovering external chance elements

The results of the experiments performed using the method described in section 4.1.1 can be found in tables 1 and 2, using agents following the Alpha-Beta search algorithm and the UCT algorithm, respectively. Several observations can be made.

First, note that for games that are supposed to be 100% skill (Chess, Tic-Tac-Toe 3x3 and 9x9), the virtual advanced player wins approximately just as much as the real advanced player. This is to be expected: in 100% skill games, there are no external chance elements from which the virtual advanced player can benefit.

Second, the same holds for games that are supposed to be 0% skill (Pasa, Tic-Tac-Die 3x3 and 9x9). While these games contain external chance elements, the agents have no free choice in moving their pieces; the virtual advanced player can look into the future, but cannot influence the outcome of the game. Therefore, the virtual advanced player has no advantage over the real advanced player.

Third, the results for Royal Game of Ur are just as expected: the virtual advanced player (almost) always beats the real advanced player. This is due to

Game	% Skill	#Wins Virtual	#Draws	#Wins Real
Royal Game of Ur		20	0	0
Pasa	0	1028	0	972
Backgammon		27	0	23
Chess	100	8	1	11
Tic-Tac-Toe 3x3	100	0	20	0
Tic-Tac-Toe-Random 3x3	75	11	2	7
Tic-Tac-Toe-Random 3x3	50	13	4	3
Tic-Tac-Toe-Random 3x3	25	12	4	4
Tic-Tac-Toe-Die 3x3	0	11	2	7
Tic-Tac-Toe 9x9	100	0	19	1
Tic-Tac-Toe-Random 9x9	75	8	0	2
Tic-Tac-Toe-Random 9x9	50	9	0	1
Tic-Tac-Toe-Random 9x9	25	9	0	1
Tic-Tac-Toe-Die 9x9	0	5	0	5

Table 1: Results for playing the virtual advanced player (having knowledge of the outcome of external chance elements) versus the real advanced player, both using the Alpha-Beta search algorithm

the fact that there are external chance elements, and the agent can influence the course of the game by choosing which piece to move.

Fourth, playing Tic-Tac-Toe on a 9x9 grid versus a 3x3 grid, gives the virtual advanced player more chance to win (if it is not a 0 or 100% skill game). This makes sense: the virtual advanced player has an advantage in the game, but for smaller grids, the real advanced player can 'accidentally' still perform very well, diminishing the benefits of this advantage for the virtual advanced player. When the grid is expanded, there are many more options, and the advantage becomes clearer.

Fifth, the advantage of the virtual advanced player is not necessarily stronger when the skill percentage of Tic-Tac-Toe-Random goes down (i.e. moves more towards Tic-Tac-Die), or vice versa. This could have to do with the earlier described trade-off: the more chance external chance elements there are, the more extra knowledge the virtual advanced player has compared to its opponent; however, this often goes hand in hand with a smaller degree of freedom of choices the virtual advanced player has, meaning that influencing its moves is made harder. Perhaps, the optimal skill percentage for the virtual advanced player is 50%, but insufficient experiments have been performed to determine this exactly.

Sixth, the virtual advanced player performs much better when using the UCT algorithm than when using the Alpha-Beta search algorithm when playing Backgammon. An explanation for this could be that the heuristics for the Alpha-Beta search algorithm are not optimal, resulting in distorted results.

Game	% Skill	#Wins Virtual	#Draws	#Wins Real
Royal Game of Ur		19	0	1
Pasa	0	1000	0	1000
Backgammon		47	0	3
Chess	100	7	7	6
Tic-Tac-Toe 3x3	100	6	8	6
Tic-Tac-Toe-Random 3x3	75	16	3	1
Tic-Tac-Toe-Random 3x3	50	20	0	0
Tic-Tac-Toe-Random 3x3	25	17	2	1
Tic-Tac-Toe-Die 3x3	0	9	2	9
Tic-Tac-Toe 9x9	100	7	6	7
Tic-Tac-Toe-Random 9x9	75	8	2	0
Tic-Tac-Toe-Random 9x9	50	7	3	0
Tic-Tac-Toe-Random 9x9	25	9	1	0
Tic-Tac-Toe-Die 9x9	0	4	0	6

Table 2: Results for playing the virtual advanced player (having knowledge of the outcome of external chance elements) versus the real advanced player, both using the UCT algorithm

5.2 Choosing random (opponent's) moves

The results of the experiments based on 4.1.2 are displayed in table 3.
For games that contain 100% skill, the Random AI should lose (almost) all the games. The more chance is involved, the more the Random AI will win. For games that are 0% skill, the numbers of wins should be the same for the random AI and the Alpha-Beta AI. This corresponds with the results of the experiments. The Alpha-Beta AI for example wins almost all chess games, a pure skill game. On the other hand, the Random AI and the Alpha-Beta AI share their wins for Pasa, a pure chance game. Note that Tic-Tac-Toe causes many draws: Random AIs can 'accidentally' draw against Alpha-Beta AIs if they pick the right squares by chance.

Unfortunately, the results for Alpha-Beta AIs choosing n% of random moves for their opponents were meaningless. Whether this is due to a bug in our code, a shortcoming of the Alpha-Beta search heuristics, or a limitation of the method, remains unclear. Since the results for the agents using the UCT algorithm were also not promising, it is likely that the problem is a bug in our code or a limitation of the method.

Game	#Wins Random AI	#Draws	#Wins AlphaBeta
Tic-Tac-Chess	0	0	200
Go	0	2	198
Chess	3	0	197
3D Tic-Tac-Toe	0	7	193
Ultimate Tic-Tac-Toe	0	8	192
Reversi	7	1	192
Kalah	18	1	186
Checkers	51	0	149
Tic-Tac-Toe	0	60	140
Gomoku	61	0	139
Tic-Tac-Toe Misere	0	71	129
Backgammon	76	0	124
Pasa	104	0	96

Table 3: Results for a player that chooses Random opponents moves 100% of the time vs a standard Alpha-Beta player

5.3 Standard Deviation ELO Ratings

There are 22 games which are both available in Ludii and on Board Games Arena. The standard deviation of the ELO Ratings for all those games have been calculated.

Game	St. Dev	#Players	Game	St. Dev	#Players
Hex	123.68	902	Lines of Action	88.97	170
Chess	116.15	10084	Quantik	88.93	936
Connect Four	108.40	20905	Tablut	88.65	229
Quoridor	104.76	5877	Quarto	88.00	4884
Reversi	102.67	5900	Backgammon	86.81	10058
Kalah	99.58	2128	Nine Men's Morris	85.93	2000
Gomoku	96.54	4826	Xianghi	77.26	952
Pylos	95.92	726	Squadro	76.90	931
Connect6	92.79	978	Senet	74.86	693
Santorini	91.05	6574	Yahtzee	63.38	35290
Checkers	90.00	5415	Gopher	44.57	2709
Go	89.92	1144			

Table 4: Standard deviation ELO Rating for games that are both in Ludii and on the online platform

For some games, there were only a limited amount of players. In figure 1, all standard deviations of games with more than 2,000 players are shown. There were 4 games with more than 10,000 players which are displayed in green; the games in blue are the games played by 2,000 to 10,000 players.

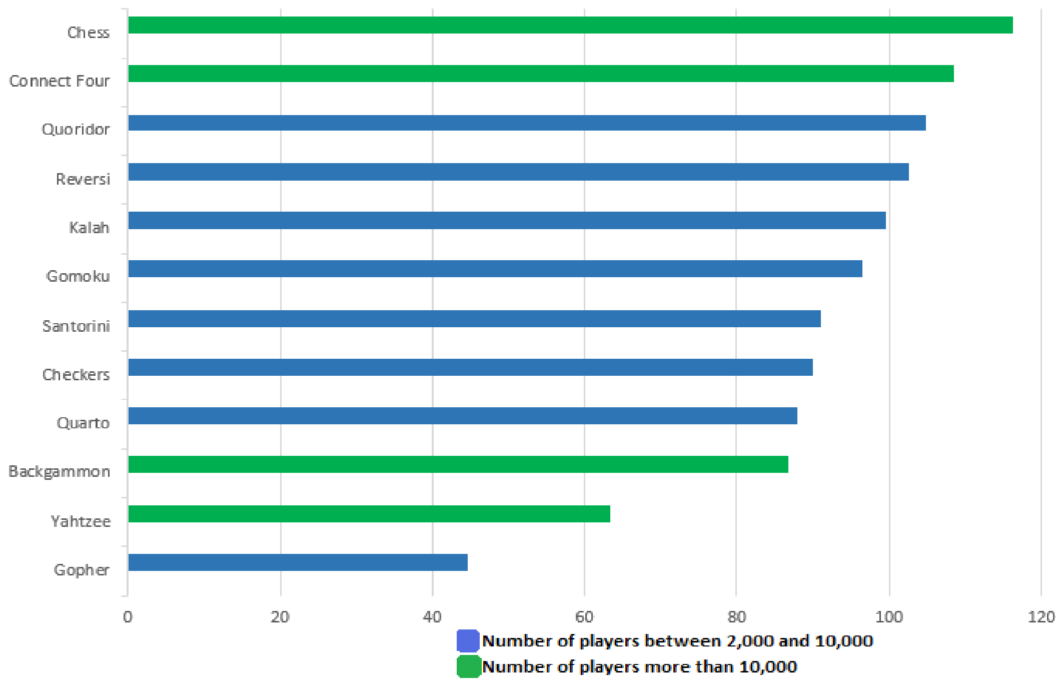

Figure 1: Standard deviation ELO Rating for games with at least 2000 players

Chess, as a pure skill game, has the highest standard deviation, while Yahtzee, a game where a lot of luck is involved, has a low standard deviation. So it seems that the standard deviation of the ELO Rating can be a good measure of the chance/skill ratio. However, for some games, like Gopher, the standard deviation does not seem reliable. This is probably caused by the fact that the amount of players is limited.

6 Discussion

6.1 Validation

The most challenging aspect of this project is that the translation of the chance/skill ratio of a game to a number between zero and one is somewhat arbitrary. For the extreme cases it is obvious: chess is 100% skill and a dice roll is 100% chance. For all other percentages, however, it is not so clear. The exact number depends on the method used: there is no guarantee or even a reason that this number should be equal for a particular game.

Therefore, we rely on an intuitive ordering of games with respect to their chance/skill ratios. In addition, we can compare the orderings determined by several methods. Note that the experiments for the method using the virtual advanced player are not sufficient in establishing a proper ordering. Many games tested for this method are games that are either 100% skill or 100% chance. The remaining games (Royal Game of Ur, Backgammon, and all Tic-Tac-Toe variants which are not pure skill and not pure chance) give very similar results in terms of wins/losses.

Therefore, the comparison of the method considering choosing random opponent's moves with the method using ELO-ratings remains. The results can be found in table 6.1.

Game	Win Random AI	Draw	Win AlphaBeta	Game	St. Dev
Chess	3	0	197	Chess	116.15
Reversi	7	1	192	Reversi	102.67
Kalah	18	1	186	Kalah	99.58
Checkers	51	0	149	Gomoku	96.54
Gomoku	61	0	139	Checkers	90.00
Backgammon	76	0	24	Backgammon	86.81

Table 5: Ordering of games based on the random AI vs the Alpha-Beta AI (left) vs the ordering of the games based on the ELO Ratings (right)

As can be seen, there is only one difference in the ordering of both methods: checkers and Gomoku have been swapped. Since the orderings of the two methods mostly correspond, this is an assurance that both methods work as desired. In addition, the orderings are in line with our intuitive estimations.

6.2 Limitations

6.2.1 Exact quantification of the chance/skill ratio

While the methods described in this report properly order games by their chance/skill components, exactly quantifying this chance/skill ratio has been unsuccessful, given the limited number of experiments done. Each experiment required much more time than initially expected. Therefore, establishing an ordering of a subset of games in Ludii is the best this project could achieve.

6.2.2 Dependence on Alpha-Beta Heuristics

As described several times throughout the report, the quality of the methods using Alpha-Beta AIs is heavily dependent on the quality of their search heuristics. If these are not well defined, the methods cannot perform well. Instead, agents using the UCT algorithm can be used to mitigate this problem.

6.2.3 Dependence of ELO-ratings method on large player bases

For the method calculating standard deviations of ELO-ratings, large player bases are needed for each game. Since Ludii is set up to also investigate ancient games (which are unpopular these days), for many games there will be no (reliable) ELO-ratings available.

6.2.4 'Drawish' games

For simple games (that have low analytic complexity), purely random play might result in many draws. An example of such a game is Tic-Tac-Toe. To mitigate this limitation, a larger grid was experimented with. Indeed, the number of draws went down significantly.

6.2.5 Method: cheating player discovering internal chance elements - future work

The method in which the virtual advanced player has knowledge of external chance elements does not take internal chance elements (hidden information and player unpredictability) into account. Following the method by Borm and van der Genugten, a game such as rock-paper-scissors would be labeled as a pure skill game. This is because the random effect is zero: a virtual player, knowing the outcome of external chance elements, does not perform better than a non-cheating player. Intuitively, however, rock-paper-scissors seems to involve a significant chance element. One could of course argue that a player can somewhat increase their winning probability by trying to read its opponent's body language, but it seems incorrect to claim that rock-paper-scissors solely depends on skill.

Dreef, in cooperation with Borm and van der Genugten, addresses this issue three years later in an updated paper. Using a game-theoretic approach, the authors succeed in taking internal chance elements into account which occur when players use mixed strategies. (Dreef et al., 2004). Just like the virtual advanced player, an agent could be developed which has knowledge of all hidden information and/or player strategies.

6.2.6 Two-player games - future work

This project only considers two-player games. However, there are many games that involve a single player, or more than 2 players. Single-player games are often puzzles. One way to test the chance/skill ratio of these games, is to let a strong AI and a weak AI play with the same configuration, and check whether the puzzle can be solved by both of them. If only the strong AI can solve it, then there must be a considerable skill component. Clearly, different strengths of AIs can be experimented with. For multi-player games, an experiment can be carried out in which one strong AI plays against weak AIs. Assuming each player tries to maximize their own reward, and no coalitions are formed, the strong AI should mostly win if there is a skill component. Another option is to let one weak AI play against several strong AI; in this case, the weak AI should mostly lose if there is considerable skill component.

References

Borm, P. and Genugten, B. (2001). On a relative measure of skill for games with chance elements. *TOP: An Official Journal of the Spanish Society of Statistics and Operations Research*, 9:91–114.

Borm, P. and van der Genugten, B. (2016). Texas hold'em, a game of skill. *International Game Theory Review*, 18(3).

Browne, C. and Piette, E. (2019). *Digital Archaeoludology*.

Costikyan, G. (2013). *Uncertainty in Games*. The MIT Press.

Dreef, M., Borm, P., and Genugten, B. (2004). A new relative skill measure for games with chance elements. *Managerial and Decision Economics*, 25:255–264.

Duersch, P., Lambrecht, M., and Oechssler, J. (2017). Measuring skill and chance in games. Discussion Paper Series 643, University of Heidelberg, Department of Economics, Heidelberg. urn:nbn:de:bsz:16-heidok-238671.

Elo, A. (1978). *The Rating of Chessplayers, Past and Present*. Arco Pub.

Maastricht University (2021a). Ludii portal. `https://ludii.games/`.

Maastricht University (2021b). Ludii portal, xii scripta. `https://ludii.games/details.php?keyword=XII%20Scripta`.

Rijksoverheid (2021). Rules for games of chance. `https://www.government.nl/topics/games-of-chance/rules-for-games-of-chance/`.

Secret, M. (2012). Poker is more a game of skill than of chance, a judge rules. https://www.nytimes.com/2012/08/22/nyregion/poker-is-more-a-game-of-skill-than-of-chance-a-judge-rules.html.