Is team gamer frustration justified?

The effect of team size on the convergence of Elo-rating

Rik Dolfing, Michiel Marcus, Rik van Toor

A common exclamation amongst gamers is that they cannot climb to their alleged true skill, because of the lack of competence amongst their own team mates. It is very easy to put the blame on someone other than yourself, but are these accusations really so far-fetched? To determine whether the expressed frustration is justified, the research conducted focused on finding a relation between team size and Elo convergence time. The results, for which Monte Carlo simulations were used, show a positive linear relation, which leads to the conclusion that team size indeed affects the time it takes a player to converge to their true Elo rating. However, the answer to the research question is no. No matter the time it takes to converge, a player will eventually converge to their true skill. This means that failing to climb the ladder for a long while can only be explained by insufficient proficiency.

1 Introduction

Elo (*1*) is a rating system that can be used to determine a player's relative skill. The system was originally created with *Chess* in mind, but gradually more and more implementations were found. Nowadays, a great number of video games make use of this system in some way, whether it is an unaltered version or a deeply adjusted one. Online, where players can compete with each other in so called PvP(Player vs. Player) games a lot of teams are formed. With team based games, people tend to say : "I do not belong at this Elo." (*2*) But does a team really have that much influence on an individual skill level or is it just a misconception? More importantly, does it matter how many teammates a player has, to converge to their true skill level within a certain time?

Nowadays, there are many existing alternatives to Elo. Popular choices are Mark Glickman's *Glicko* and *Glicko2* (*3*), as well as Microsoft's *TrueSkill*™ (*4*), which was based on some of the core concepts of *Glicko*. Both of these systems do not consider the effect of team sizes on a single players convergence, but are mainly focused on matchmaking and outcome. In this paper, the traditional Elo system is used to simulate the progress of a player's rating.

The Elo system describes a function to determine the chance that a player will beat another player.

Definition 1. E_A represents the chance that player A will win from player B and R_B and R_A are the Elo ratings of player A and B, respectively.

$$E_A = \frac{1}{1 + 10^{\frac{R_B - R_A}{400}}}$$

Additionally, the Elo system describes a function to adjust a players Elo rating after a game.

Definition 2. The updated Elo rating R' is determined based on the previous Elo rating R, a variable W, which equals 1 if the player has won the game and 0 if they lost, the chance the player would have won the game P, calculated using the evaluation function and a variable K.

$$R' = R + (W - P) \cdot K$$

Where

$$K = \begin{cases} 16 & R < 2000 \\ 12 & 2000 \leq R \leq 2200 \\ 10 & R > 2200 \end{cases}$$

2 Methods

To simulate the workings of this system, a program based on Monte Carlo simulations (*5*) was written. The program used is best described using a step-by-step plan with elaboration where necessary.

2.1 Assumptions

A players Elo Rating has converged if their Elo does not change significantly over a span of at least 5% of the number of games played so far. This is checked by taking the standard deviation of the player's Elo rating over the most recent games with regard to the players true skill. If this deviation is lower than a certain constant α, the Elo rating has converged. To get to the value of α, the following calculation was used:

$$R' = R + (W - P) \cdot K$$

$$\Delta R = (W - P) \cdot K$$

Since $(W - P)$ ranges from -1 to 1, the maximum value of ΔR equals the maximum value of K: 16. The minimum value of ΔR equals -16. Since the player will lose approximately the same number of games as they win, since they play against players of the same skill level, their Elo will fluctuate their true skill $\pm\Delta R$. To be on the safe side, the value of α was set to 20.
The Elo rating itself is clamped between 1200 and 2800, which means that a rating cannot go lower than 1200 nor higher than 2800. Every new player starts at 1500 Elo, whilst the team size can be any positive integer.

2.2 Matchmaking

> "[League of Legends] tries to match teams as fairly as possible: it computes the average of the player ranking values for each team, and then uses these averages to match teams similarly to one-versus-one fights." (*6*)

The proposed program uses the same matchmaking technique as the popular online game *League of Legends* does. Because of the MOBA/League of Legends based matchmaking, if a player's true skill exceeds the opponent's true skill by at least 100 points, the chance becomes 1 that this player wins.

2.3 Evaluation function

The members of both teams are subdivided into couples. This means that each player on a team is matched with one opponent. Each individual win earns a win chance of $\frac{1}{N}$ where N represents the team size. For example, in a situation with teams of size 3, all members of team A and team B will be paired up.

Example

> A_1 versus B_1, A_2 versus B_2, A_3 versus B_3. If A_1 beats B_1, A_2 beats B_2, but A_3 loses to B_3, the values will be $\frac{1}{3}$, $\frac{1}{3}$ and 0 respectively. Thus, the accumulated value of

this round equals $\frac{2}{3}$.

This accumulated value will be referred to as *factor 1* in future reference. This is a factor that is prominent within Multiplayer Online Battle Arena games such as League of Legends (*6*).The ability to outsmart or outplay an opponent in a 1 versus 1 setting can determine the outcome of an entire game. However, an equally important factor is the ability to play together. This is the core of most games involving teamwork. For that reason, the average of the true skill is taken for both teams. This average will be referred to as *factor 2* in future reference. Lastly, the average of *factor 1* and *factor 2* is inserted into the original Elo win chance function mentioned in the Introduction section.

3 The program

The program uses 4 main variables. These are the number of simulations(S), a team size(N), the evaluation function(e) and the maximum difference between players called the competence range(c). There are two phases in the program: the *initialisation phase* and the *game phase*. In the *initialisation phase*, a new player(p) is created which is monitored by the program, with a random true skill level t and a starting Elo of 1500. Starting the game phase, $N-1$ random allies(a) are added to ps team. N Random opponents(o) are added to the other team. For all generated opponents and allies, the Elo rating is calculated by

$$Elo_p \pm RandomMax(0, c)$$

When each a and o have been determined, the evaluation function(section 3.3) calculates which team wins and what the new rating of p will become. If p's Elo stays within $Elo_p \pm 20$ for 5% of the games played so far, the player's Elo rating has converged.

4 Results

Using the program described in the Methods section, the following data was generated. For each team size, a sample size of 10000 simulations was used. In order to identify the effect of team size, the simulations were run with different competence values. The confidence intervals have a probability of 0.95.

Table 1: Using competence range 50

Team Size	**Average**	**Standard Deviation**	**Confidence Interval**
1	193	11	193 - 193
2	229	66	228 - 230
3	280	126	278 - 282
4	341	187	337 - 345
5	404	245	399 - 409
6	474	306	468 - 480
7	537	365	530 - 544
8	595	414	587 - 603
9	659	466	650 - 668
10	712	516	702 - 722

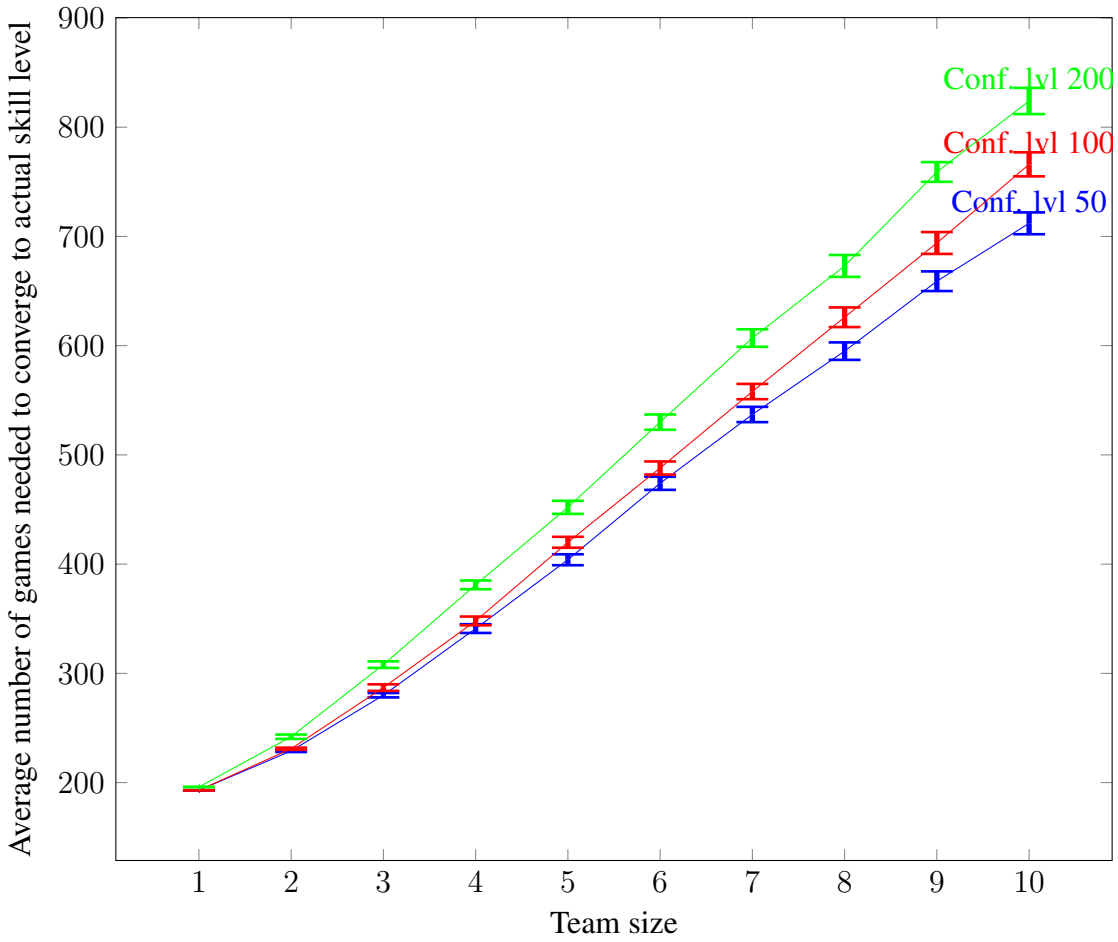
Table 2: Using competence range 100

Team Size	**Average**	**Standard Deviation**	**Confidence Interval**
1	193	13	193 - 193
2	231	70	230 - 232
3	287	133	284 - 290
4	348	192	344 - 352
5	420	259	415 - 425
6	488	320	482 - 494
7	558	381	551 - 565
8	626	451	617 - 635
9	694	511	684 - 704
10	766	580	755 - 777

Table 3: Using competence range 200

Team Size	Average	Standard Deviation	Confidence Interval
1	196	18	196 - 196
2	242	83	240 - 244
3	308	153	305 - 311
4	381	221	377 - 385
5	452	282	446 - 458
6	530	355	523 - 537
7	607	430	599 - 615
8	673	494	663 - 683
9	759	564	748 - 770
10	824	636	812 - 836

Fig. 1: Games needed for team of given size



5 Discussion

The figures show that there is a linear relation between the amount of games that are needed to converge a player's Elo rating towards their real skill level. The confidence intervals are small, while the average number of games can become high numbers for bigger teams. Compared to TrueSkill™, which has been Researched by Microsoft, the proposed simulation needs more games to converge a player's Elo to their true skill level. However, while TrueSkill™is based on some of the fundamentals of Elo, it uses a Bayesian rating system. (*4*) For a game with 2 teams of 4 players each, TrueSkill™needs approximately 46 games (*7*), whereas the traditional Elo system needs roughly 350 games, which is around 8 times as many games as TrueSkill™needs. If a team consists of 8 players, TrueSkill™needs 96 games, while the simulation indicates that 595 games are needed to converge.

While TrueSkill™does not particularly take the team sizes into account (just flexible teams), the same positive linear relation can be found here. It can be noted that the program used in this paper will not be of any use in the real (gaming) world.

> "However, there is another factor taken into consideration which is MVP (Most Valuable Player). MVP is given to a player which has done the most work for their team. Please note that a player with 5 kill can still lose the MVP to a person who defuses or successfully detonated the bomb. The reason for this is because, CSGO is a team-based game, a player can kill the whole enemy team but still lose the round because he may not have the defuser to defuse the bomb in time." (*8*)

The program does not take into account anything specific about the current game, like MVP awards or the main weapon of choice of the player, which will increase a player's skill for that game and might get a higher win percentage for the current game. Last, the program does not take into account something like a draw, which suggests a good match (*7*). Further research

is needed to show if the mentioned factors like MVP and draws do really affect a player's convergence rate in relation to the size of the team.

6 Conclusion

The hypothesis posed in this paper stated that the size of a team in a particular game would influence the time it takes a player's rating to converge to the correct number. The results show that a positive linear relation can be found between team size and convergence time. This leads to the conclusion that the answer is yes; the size of a team impacts the overall convergence of the respective player's rating. However, this does not yet answer the question posed at the start of this paper: Is gamer frustration justified? To justify the frustration expressed by gamers, one needs to prove that at some point there is nothing a player can do to climb to their allegedly higher rating. For this to be true, an example is required where it takes excessively long for a player to climb to their own level. In contrast, though there have been examples in the simulations that did take long, it never occurred that the rating did not converge. Consequently, a sound conclusion would be that when a player is not able to climb for a long time, they have converged to their skill level and failure to climb further is to be blamed on themselves and themselves alone.

References and Notes

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

2. Example of claim: http://forums.na.leagueoflegends.com/board/showthread.php?t=1992064.

3. Dr. M.E. Glickman, *The Glicko system* (2016).

4. R. Herbrich et al., *TrueSkill™: A Bayesian Skill Rating System* (2006).

5. S.J. Mason et al., *Introduction to Monte Carlo simulation* (2008).

6. MOBA League matchmaking.

7. TrueSkill^{TM}Convergence Rates https://www.microsoft.com/en-us/research/project/trueskill-ranking-system/.

8. CounterStrike MVP https://steamcommunity.com/sharedfiles/filedetails/?l=dutch&id=312582297.