The Wide, the Deep, and the Maverick: Types of Players in Team-based Online Games – Supplementary Information

JULIE JIANG, USC Information Sciences Institute, USA
DANAJA MALDENIYA, University of Michigan, USA
KRISTINA LERMAN, USC Information Sciences Institute, USA
EMILIO FERRARA, USC Information Sciences Institute, USA

A SUPPLEMENTARY INFORMATION FOR THE MAIN DATASET

A.1 Estimating position data

Riot's API does not provide the position that players selected during each match. Instead, they only provide the *role* and *lane* selections of each player in each match. Ideally, one could infer the position from the role and lane. But it is not a clear-cut, one-to-one mapping, with many possible combinations of the role and lane resulting in unknown positions[1].

To overcome this, we estimate positions for players in our main dataset by building a classifier that relies on different items and spells utilized by players during a match, reusing match data and methodology from Lee and Ramler [2]. Their dataset, which we refer to as the *L&R* dataset, contains nearly 5 million matches from the North American Server over a period of 8 months across multiple match queues including 112,301 ranked team matches and 2,044,068 ranked solo matches. Since we are interested in individual player behavior in matches and each match consists of 10 players (5 players in each team), each match corresponds to 10 units of analysis which we call *player match instances*. Additionally, for each player in every match, this dataset contains the recommended position for the employed champion based on the community metagame at the time. These positions had been scraped from the LoL metagame website https://champion.gg/. We randomly select 20,000 player-match instances from the ranked team queue data and train a one-vs-one SVM model to predict the positions the players had chosen. Following Lee and Ramler [2], we use the set of items and spells that players had acquired by the end of the game as our independent variables with the metagame positions as the ground truth. We also follow their choice to use the ranked team queue for training because people are more likely to employ strategies that align with the community metagame when queuing as a team, i.e. the metagame position for the champions is more likely to reflect ground truth. Next, we validate the performance of the model on the remaining 1,103,010 player-match instances from the ranked team queue ($\approx$ 85% accuracy) and then predict the positions for each instance of a player in a match in our own dataset. Further discussion on the validity of this approach can be found in Lee and Ramler [2].

Additionally, in a preliminary analysis, we have used role and lane to calculate player diversity and typicality, where we treat them as separate variables. The preliminary results are in line with our main results.

[1]See https://riot-api-libraries.readthedocs.io/en/latest/roleid.html

Authors' addresses: Julie Jiang, USC Information Sciences Institute, Marina del Rey, California, USA, juliej@isi.edu; Danaja Maldeniya, University of Michigan, Ann Arbor, Michigan, USA, dmal@umich.edu; Kristina Lerman, USC Information Sciences Institute, Marina del Rey, California, USA, lerman@isi.edu; Emilio Ferrara, USC Information Sciences Institute, Marina del Rey, California, USA, ferrarae@isi.edu.

A.2 Adaptation to balance patches

Tables S1 and S2 provides details and test results on the pick rate differences and mean KDA differences, respectively, for each player type under the context of each patch change.

Table S1. The champion pick rate differences in the before patch period t_1 vs. the control period t_0 (*"control"*) and the after patch period t_2 vs. before patch period t_1 patch (*"before and after patch"*). n is the total number of champions that had been buffed, nerfed or unaltered. We apply paired t-tests on the mean pick rate differences to analyze any significant difference.

			Control		Before and after patch	
		n	Δ Pick rate	T-test	Δ Pick rate	T-test
Buffs	Mavs.	135	0.0003	0.598($p = 0.551$)	0.0005	1.004(p=0.317)
	Gens.	135	−0.0001	−0.436($p = 0.664$)	0.0007	2.134*
	Spcs.	135	−0.0002	−0.484($p = 0.629$)	0.0007	1.065($p = 0.289$)
Nerfs	Mavs.	98	0.0005	0.800($p = 0.425$)	−0.0025	−4.065***
	Gens.	98	−0.0006	−0.868($p = 0.387$)	−0.0019	−2.986**
	Spcs.	98	−0.0007	−0.747($p = 0.457$)	−0.0019	−2.297∗
Unaltered	Mavs.	2154	−0.0000	−0.224($p = 0.823$)	0.0000	0.260($p = 0.795$)
	Gens.	2154	0.0000	0.275($p = 0.783$)	0.0000	0.163($p = 0.870$)
	Spcs.	2154	0.0000	0.230($p = 0.818$)	0.0000	0.124($p = 0.901$)

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table S2. The champion mean kDA differences before patch period t_1 vs. the control period t_0 (*"control"*) and the after patch period t_2 vs. the before patch period t_1 patch (*"before and after patch"*). n is the total number of champions who have been played by the correspondoing player type in the two consecutive periods. We apply Wilcoxon signed-rank tests on champions' mean KDAs between consecutive periods.

		Control			Before and after patch		
		n	Δ KDA	Wilcoxon	n	Δ KDA	Wilcoxon
Buffs	Mavs.	116	−0.072	3333 ($p = 0.869$)	122	0.100	3595 ($p = 0.927$)
	Gens	111	0.131	2813 ($p = 0.475$)	114	0.132	3022 ($p = 0.470$)
	Spcs.	43	−0.085	449 ($p = 0.772$)	55	−0.263	692 ($p = 0.664$)
Nerfs	Mavs.	97	−0.061	2114 ($p = 0.346$)	103	−0.217	1693**
	Gens	97	−0.027	2040 ($p = 0.227$)	101	−0.059	2117 ($p = 0.212$)
	Spcs.	70	0.181	1201 ($p = 0.808$)	73	−0.580	862**
Unaltered	Mavs.	1930	0.006	905123 ($p = 0.510$)	2035	0.007	1028575 ($p = 0.996$)
	Gens.	1846	0.037	807122 ($p = 0.153$)	1936	−0.065	861584*
	Spcs.	1202	0.065	355730 ($p = 0.817$)	1257	−0.019	377197 ($p = 0.430$)

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table S3. Odds ratios of a mixed effect model predicting if the team wins given the number of additional users they have in each user type (mavericks, generalists, or specialists) and the final Trueskill bracket. The random effect is the mode of the highest achieved season tier of all 10 players in a match.

Skill Level		**Mavericks**	**Generalists**	**Specialists**
	10%	1.082^{***}	1.063^{***}	1.071^{***}
Top	10%-20%	1.046^{***}	1.044^{***}	1.065^{***}
	20%-30%	1.033^{*}	1.034^{**}	1.051^{*}
	20% - 30%	0.976^{*}	0.972 ($p = 0.053$)	0.969^{*}
Bottom	10% - 20%	0.961^{***}	0.963^{**}	0.943^{**}
	10%	0.921^{***}	0.926^{***}	0.934^{**}

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

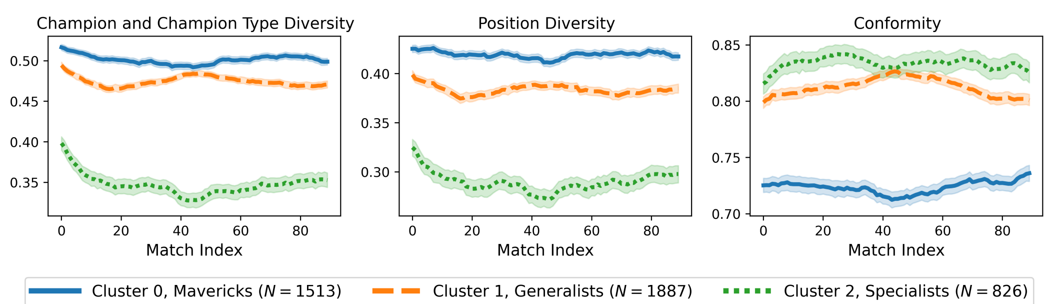

Fig. S1. Time series of behavior indices for the three clusters of user with bootstrapped 95% confidence interval (*L&R* dataset).

A.3 Effect of individuals on team performance given last season's tier

Most players begin the season with a tier that is determined by their previous season's ranking, which is used internally by Riot to match players and teams of similar skill levels together. Consequently, the effect of individuals on team performance can be confounded by player tiers. Though Riot does not release tier levels of players, it releases the highest achieved season tier of each player in the *previous* season. Borrowing from Kim et al. [1], we can use the highest achieved previous season tier as a proxy for each player's skill level.

We replicate the model set up in the main manuscript for predicting the winning team given the number of best or worst players of each player type (§6.3.3), with the exception of using a mixed-effect model instead of a regular logistic regression model, with the mode of the team members' highest achieved previous season tiers as the fixed effect. The results are presented in Table S3. We observe that all effect sizes are smaller following the addition of fixed effects, with all of them being closer to 1. This is not unexpected because many winning outcomes may be well predicted by the highest achieved previous season tiers alone, thus any effect by a single player is reduced. However, almost all mixed-effect model results are significant, indicating the robustness of our results. Moreover, all of our conclusions from the logistic regression model in the main text regarding the relative effect sizes among player types and among skill levels apply to this mixed-effect model.

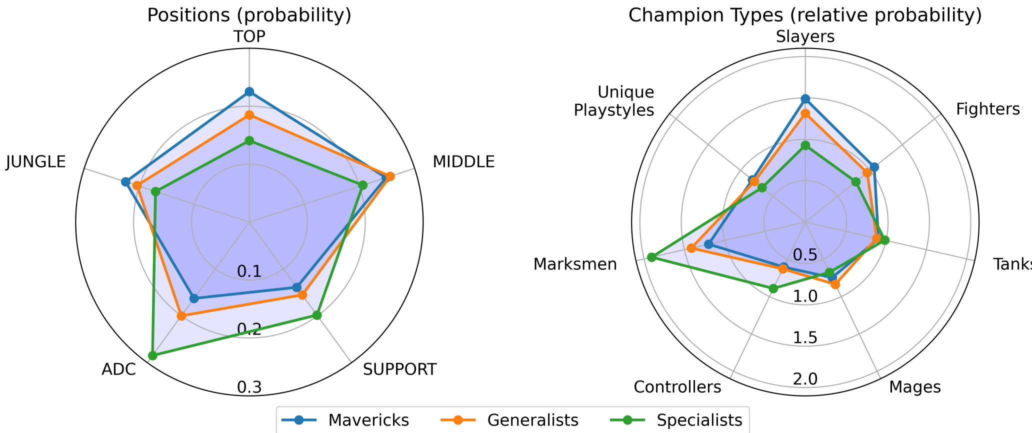

Fig. S2. Left: Proportion of matches where mavericks, generalists and specialists selected each position. Right: Relative probability of selecting a champion in each champion type compared to a uniformly random selection (*L&R* dataset).

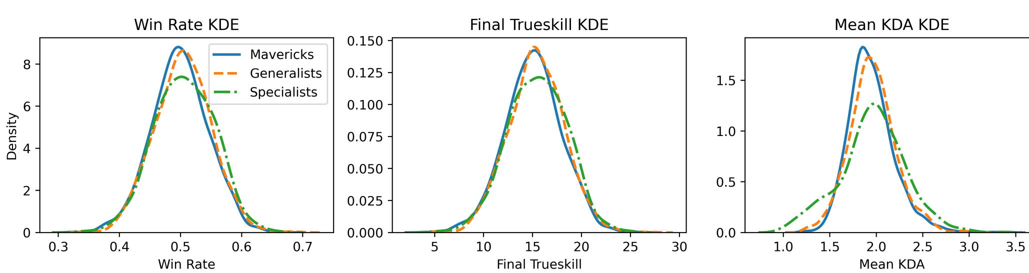

Fig. S3. KDE of win rate, final Trueskill, and mean KDA for each user's first 100 matches (*L&R* dataset).

B RESULTS ON THE SECONDARY DATASET

We use a secondary dataset collected by Lee and Ramler [2] (*L&R*). This dataset is a ranked solo-queue dataset collected from the EU server between June 2015 and November 2015. We note that a similar dataset was also collected from the NA server by Lee and Ramler [2], which is the same server as our main dataset. However, the timestamps of this dataset were erroneous making it impossible to observe the true behavioral progression of user behaviors. This dataset is much larger in size than our main dataset, with 4,226 eligible players (those with at least 100 matches), which is more than 10 times the number of eligible players than we have in the main dataset. However, this dataset only spans half of all the patches in season 2015. As such, we do not conduct analyses on patch effects. Instead, we only replicate our methods and analyses on this secondary dataset as a robustness check for non-patch related research questions (RQ1-RQ3). Considering only the first 100 matches of each of the 4,226 users, the EU dataset contains 390,000 unique matches for a total of 423,000 player-match pairs.

B.1 RQ1 and RQ2: Gameplay strategies and their evolution

Our first observation is that the number of suitable clusters obtained from K-means clustering is 3, and the three types of players detected are consistent with the findings of our main dataset (Fig.

Table S4. K-S test results on pairwise comparisons of every user group using their win rate, final trueskill, and mean KDA CDFs of each user's first 100 matches (*L&R* dataset).

	Win Rate CDF	**Final Trueskill CDF**	**Mean KDA CDF**
Generalists < Specialists	$D- = 0.051^{*}$	$D- = 0.048, p = 0.064$	$D- = 0.093^{***}$
Mavericks < Generalists	$D- = 0.052^{**}$	$D- = 0.054^{**}$	$D- = 0.065^{***}$
Mavericks < Specialists	$D- = 0.084^{***}$	$D- = 0.086^{***}$	$D- = 0.119^{***}$

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table S5. Probability of being in the top 10% and bottom 10% of all players by win rate and by final Trueskill (*L&R* dataset).

	Top 10%		**Bottom 10%**	
	Trueskill	Win Rate	Trueskill	Win Rate
Mavericks	0.091	0.124	0.106	0.116
Generalists	0.097	0.131	0.096	0.108
Specialists	0.125	0.169	0.098	0.100

Table S6. Odds ratios of predicting if the team wins given the number of additional users they have in each user type (mavericks, generalists, or specialists) and final Trueskill bracket (*L&R* dataset).

Skill Level		**Mavericks**	**Generalists**	**Specialists**
Top	10%	1.417^{***}	1.405^{***}	1.405^{***}
	10%-20%	1.230^{***}	1.228^{***}	1.236^{***}
	20%-30%	1.146^{***}	1.126^{***}	1.150^{***}
Bottom	20%-30%	0.891^{***}	0.888^{***}	0.896^{***}
	10%-20%	0.832^{***}	0.829^{***}	0.836^{***}
	10%	0.717^{***}	0.724^{***}	0.730^{***}

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

S1), i.e. we observe mavericks, generalists, and specialists. The time series progression of diversity and typicality indices remain fairly stable across time (Fig. S1).

Using the χ^2 test of independence, we again find that preferences of the different player groups are significantly different for champion types and positions (Fig. S2). Consistent with the results from our main dataset, specialists favor the ADC role and Markmen champions, while Mavericks prefer solo roles.

B.2 RQ3: Long term performance

We replicate the key finding from our main dataset that specialists perform best overall, followed by generalists and then specialists. Figure S3 and Table S4 show the analysis of win rates and average KDA over their first 100 matches and final Trueskills at the end of the 100th match in terms of differences in kernel density and cumulative distribution respectively. Further, using one-sided, one-sample t-tests against an expected win rate of 50%, we are able to replicate the finding that generalists ($\mu = 0.504, t = 3.749, p = 0.000$) and specialists ($\mu = 0.506, t = 3.649, p = 0.000$) win more often than 50% of the time, but the win rate of mavericks do not significantly deviate than 50% ($\mu = 0.499, t = -0.598, p = 0.550$).

Table S5 supports our finding that specialists are most likely to be found in the top 10% of all players, whereas mavericks are most likely to be found in the bottom 10% of all players. Finally, we replicate our findings for predicting which team wins given the number of additional mavericks, generalists, or specialists each team has in Table S6. Similar to the model from our main dataset, we find that the top and the worst mavericks make the most impact on team performance. In the top 10%-30% bracket, specialists are the most valuable players. In conclusion, analysis of this secondary dataset validates our finding that specialists are overall the best players and mavericks are overall the worst players, although the best mavericks can be the most valuable assets for a team.

B.3 Concluding Remarks

We are able to successfully replicate most of our analyses on the secondary dataset L&R, which is larger than the dataset used for the analyses in the main document but spans a much shorter period of time.

REFERENCES

[1] Jooyeon Kim, Brian C. Keegan, Sungjoon Park, and Alice Oh. 2016. The proficiency-congruency dilemma: Virtual team design and performance in multiplayer online games. In *Proc. CHI '16*. ACM, 4351–4365. https://doi.org/10.1145/2858036.2858464

[2] Choong-Soo Lee and Ivan Ramler. 2017. Identifying and evaluating successful non-meta strategies in league of legends. In *Proc. FDG '17*. ACM, 1–6. https://doi.org/10.1145/3102071.3102081