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

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, USC Annenberg School of Communication, USA

Although player performance in online games has been widely studied, few studies have considered the behavioral preferences of players and how that impacts performance. In a competitive setting where players must cooperate with temporary teammates, it is even more crucial to understand how differences in playing style contribute to teamwork. Drawing on theories of individual behavior in teams, we describe a methodology to empirically profile players based on the diversity and conformity of their gameplay styles. Applying this approach to a League of Legends dataset, we find three distinct types of players that align with our theoretical framework: *generalists*, *specialists*, and *mavericks*. Importantly, the behavior of each player type remains stable despite players becoming more experienced. Additionally, we extensively investigate the benefits and drawbacks of each type of player by evaluating their individual performance, contribution to the team, and adaptation to changes in the game environment. We find that, overall, specialists tend to outperform others, while mavericks bear high risk but also potentially reap great rewards. Generalists are the most resilient to instability in the environment (game patches). We discuss the implications of these findings in terms of game design and community management, as well as team building in environments with varying levels of stability.

CCS Concepts: • **Human-centered computing → Empirical studies in collaborative and social computing**; **Empirical studies in HCI**; • **Information systems → Massively multiplayer online games**.

Additional Key Words and Phrases: diversity; conformity; adaptability; generalists; specialists; mavericks; user behavior modeling; MOBA games; metagame; performance; teams; League of Legends; e-sports; tensor factorization

ACM Reference Format:
Julie Jiang, Danaja Maldeniya, Kristina Lerman, and Emilio Ferrara. 2021. The Wide, the Deep, and the Maverick: Types of Players in Team-based Online Games. *Proc. ACM Hum.-Comput. Interact.* 5, CSCW1, Article 191 (April 2021), 26 pages. https://doi.org/10.1145/3449290

1 INTRODUCTION

In recent decades, collaborative teams, rather than solo endeavors, have been the driving force behind production and innovation in society. Facilitated by technological and societal advances, a new type of teamwork has emerged, where ad-hoc online teams are formed by strangers for short-lived purposes, such as open Github collaborations or online team-based gaming.

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, USC Annenberg School of Communication, Marina del Rey, California, USA, ferrarae@isi.edu.

2573-0142/2021/4-ART191 $15.00
https://doi.org/10.1145/3449290

In their ideal form, teams leverage the different skills and perspectives of their members while overcoming challenges in coordination. *Traditional* teams benefit from internal stability as the coordination costs decrease when members spend more time together. Less is known about *transient* teams, which are short-lived encounters in which members will not know each other in a meaningful way and have to rely on convention and stigmergy for coordination. While these features are seemingly limit effective team performance, these anonymous temporary teams have proven to be quite capable at specific tasks, such as collaboratively creating content, the most famous example being Wikipedia [54, 87], building complex open-source software [87], competing in challenging team-based games [53, 59] as well as solving a range of other problems in both real-world and virtual setting [86, 99, 111].

Prior research has revealed numerous effects on team performance from team diversity, in terms of individual experience and identity [62, 77], as well as team structures and conventions of coordination [80, 87]. However, they do not address another aspect, one that possibly plays a much more pivotal role in transient teams than traditional ones – the individuals' contributions to team performance based on their unique experiences, preferences, and work styles. Since communications between transient team members are limited and the teams themselves are short-lived, these teams are far less likely than traditional teams to converge on a team-wide modus operandi. In turn, this should allow transient team members to both express themselves more freely and also have greater influence on team outcomes compared to those of traditional teams [8, 44, 59].

In this study, we address the gap in the literature regarding the effects of different individual styles on transient teams in the context of League of Legends (LoL), a multiplayer online battle arena game (MOBA). By quantifying user gaming patterns on dimensions of *diversity* in the historical patterns of gaming choices and *conformity* to established gaming practices, we show that there are three distinct types of League of Legends players: (i) specialists who focus on a limited number of gameplay styles popular within the gaming community, (ii) generalists who explore a wide range of popular styles, and (iii) mavericks who consistently engage in a wide range of unconventional styles. On the whole, players maintain their playstyles over time, indicating relatively stable preferences despite their growing experience.

The three types of players correspond to varying degrees of success in performance and adaptation. On an individual level, under normal circumstances, specialists consistently outshine both generalists and mavericks. However, their comparative advantage is less clear when considering their contribution to team performance. By analyzing the differential effects of each player type on team performance, we show that mavericks have a greater effect on the outcome, increasing both the chance of winning when they are skilled and the chance of losing when they are unskilled. This suggests that the path of the maverick is one of high risk as well as high reward: They are more likely to propel a skilled team to the top but also render underperforming teams even worse. Considering adapting to perturbations in the environment through game patches, all three groups respond drastically to negative changes by departing from weakened playstyles. However, compared to the pigeon-holed specialists, and mavericks, generalists are more likely to embrace positive changes by adopting new playstyles that benefit from those changes. Finally, an analysis of performance in the aftermath of patches shows that specialists are the most vulnerable to negative changes. In contrast, generalists are the most resilient, with their performance remaining stable despite negative changes.

In sum, we, (i) identify different classes of team members within the League of Legends setting with stable preferences towards playstyles, (ii) show that the playstyle affects both individual and team level outcomes, and (iii) that this effect depends on both the skill level of the player and the stability of the environment. These findings provide insights on how more capable transient teams may be constructed by considering the work styles and preferences of individuals, especially in challenging and competitive environments. Further, they highlight the need to consider the stability of the environment as well as the associated risk/reward tradeoffs in determining the composition of these teams.

2 BACKGROUND

2.1 Overview of League of Legends

The decade-old League of Legends is now one of the most popular MOBA games worldwide, cementing its place as a behemoth in the billion-dollar e-sports industry [79]. An LoL game is a short match that lasts about 30 minutes between two teams of five players. In order to win, the teams compete to be the first to destroy their opponents' home base. The players each select a champion to play as, which, together with the starting position on the map they choose, largely determines their attack or defense strategies. During the game, players can earn points, gold, and upgrades by attacking the enemy. Next, we introduce several other integral components of the game.

Queue types. Queue types refer to different modes of the game, which dictate the way teams are formed. They can be broadly categorized as either ranked games or unranked games. Ranked game outcomes contribute to the players' seasonal ranking, and are restricted to only players with sufficient experience and skill. The ranking mechanism has been shown to mediate player practice and encourage players to take gaming more seriously [60]. Within ranked games, there are team-queue and solo-queue modes. Team-queue games are games in which players pre-form teams with friends or players they previously know. In contrast, players in solo-queue games individually join an *ad-hoc* team, likely teaming up with strangers with similar ranks [22]. Solo-queues are also far more popular than team-queue games [64], which is the queue type we focus on in this work. Much like other forms of team collaboration, effective teamwork is imperative to success in LoL. This is especially true for solo-queue games, wherein players must coordinate effectively with new and unfamiliar allies on the fly in a high-stress environment [67].

Balance patches. Online games are subject to frequent patch changes. LoL developers routinely release patches about every two weeks. Patches can fix bugs, introduce new features, and adjust the skills and abilities of champions, the latter type of patches are also known as balance patches. Champions that are perceived to be too strong and dominating can have their skills and abilities suppressed (i.e., they are *nerfed*) and likewise champions that are underperforming can have their skills and abilities boosted (i.e., they are *buffed*) [107]. Claypool et al. [17] showed that win rates increase for buffed champions and decrease for nerfed champions, providing evidence that the purpose of game balancing patches is to keep any champion's win rate close to 50%. Moreover, champion nerfs and buffs can influence users' preference for champions [107]. Overall, balance patches can draw mixed reactions from players. Balance patches can encourage players to pick a newly boosted champion, but may also drive away those who preferred the old setup. For some others, balance patches can keep the game new and exciting [17].

The biggest patches are generally introduced during *pre-seasons*, which happen prior to the commencement of annual competitive cycles of games called *seasons*. Player rankings are also partially reset at the beginning of each season to be artificially lower than what they achieved in the last season.

Metagame. A crucial component of online games is the *metagame*, loosely defined as anything that connects the game to elements beyond the game, such as utilizing the social context or external expertise to one's advantage [88]. Metagaming can also be considered as the awareness and knowledge of strategies, discussions, and changes of the game over time [23]. In his binary model of gaming expertise, Donaldson [23] argues that metagaming expertise is one of two types of gaming expertise, with the other being mechanical expertise, which is traditionally what constitutes skillful mastery of in-game mechanics. Importantly, the metagame is not static. Just as game patches are constantly released, the metagame is also a rapidly evolving guide for the current best gaming practices. One does not need to search far to find discussion forums, social media posts, or websites dedicated to debating the best team formulation based on positions or the best champions for each position.

2.2 Motivations and theoretical framework

Prior research, in domains such as sports psychology, science of science, and organizational studies, has studied how characteristics of individual members within teams affect both individual and team outcomes. Broadly, the majority of this literature considers three categories of human characteristics; (i) abstract psychological attributes such as motivations and personality [12, 13, 84, 101, 108], (ii) more concrete cognitive and phenomenological attributes such as working memory, collaborative capacity and experience [12, 29, 49, 52, 78], and (iii) anthropometric measurements such as height and muscle mass [15, 104].

In comparison to the major avenues of research, few studies have focused on psychosocial behavioral characteristics of individual diversity (experiential and/or skill) and conformity. The nature in which these characteristics manifest in individuals is influenced by their personality and motivations and they, in turn, influence their skill, experience and interactions with teams [1, 110]. The study of diversity and conformity of individuals in the context of teams can be useful, especially when those teams are temporary. In such settings, established knowledge and norms of the larger community (beyond the team) strongly shape behaviors. Therefore, individual diversity and conformity, as expressions of interactions between personalities and the community, are meaningful constructs for understanding individuals and their teams.

Further, many real-world team environments are characterized by a high degree of complexity and uncertainty. As these teams operate on a boundary between pre-existing knowledge and the unknown, individual diversity and conformity can have important effects on individual, team, and community outcomes. Next, we elaborate on the diversity and conformity of individuals in team settings to motivate our framework for characterizing individuals jointly on both dimensions to explore their implications.

2.2.1 Diversity. We operationalize diversity as the range of roles historically occupied by an individual. Belbin [10] defined roles in teams as “a tendency to behave, contribute and interrelate with others in a particular way.” We theorize that compared to traditional teams, transient teams of strangers allow individual members greater freedom to choose their roles, thereby allowing their personality and incentives to weigh more heavily in the diversification of roles. One of the Big Five personality traits that contribute to role diversification is openness to experience [31], which has also been found to be related to novelty-seeking tendencies, a fundamental personality temperament that motivates the pursuit of new experiences and sensations [21]. The behavior of a novelty-seeker will manifest through a wide-ranging exploration of strategies. At times, novelty-seeking tendencies can propel risky behavior [11], but they may also lead to the acquisition of unique complex skills because they actively seek out novel challenges [43]. In the context of gaming, novelty-seekers are immersion-oriented players who are motivated by the breadth and creativity of the game [110]. By contrast, people who score low in novelty-seeking may prefer to repeat familiar actions. Over time, they may become skilled at some narrow range of actions by virtue of prolonged repetition.

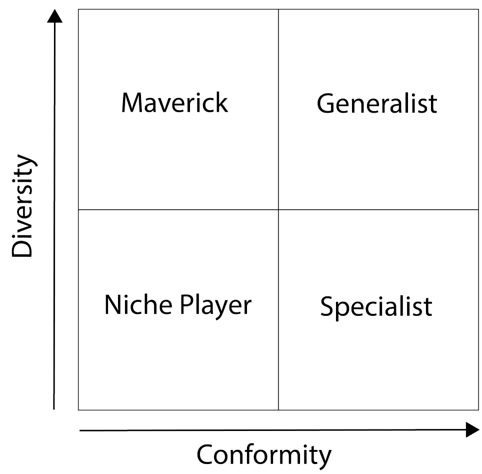

Fig. 1. Diversity-Conformity Classification Framework

2.2.2 Conformity. Humans are often gauged by how well they conform to norms and expectations. Yet not everyone chooses to conform. According to the theory of planned behavior, one's intention can determine one's behavior, the former of which combines attitude, subjective norm, and perceived behavioral control [1]. The subjective norm, in particular, refers to the degree to which subjective societal norms influence the behavior of an individual. In settings where established practices and social norms have a significant influence on the community (e.g., metagame), individuals generally display a level of conformity because their subjective norm pressures them to do so. Another widely recognized theory is frequency-dependent learning, which postulates that people tend to conform to behaviors that are the most common [6]. It can also be argued that people conform because they believe in the "wisdom of crowds" [98], the notion that the majority of the population holds the best, if not safest, strategy. Conformity can be readily seen, particularly, in team sports settings, where individual players tend to not deviate from the established norms of sports practices [82].

2.2.3 Diversity-Conformity Classification. By considering diversity and conformity jointly, we propose a framework (Fig 1) that classifies individuals into four distinct quadrants, described as follows:

(1) *Specialist.* An individual who repeatedly engages in a limited number of practices (low diversity) that are well-established and recognized within the community (high conformity). Specialists have been studied in numerous contexts such as team sports [24, 104], scientific discoveries [20, 35, 69, 73], online knowledge production [65, 105], and organizational teams [57, 81]. These studies frame functional specialization in teams as an adaptive mechanism in environments too complex for individuals to fully master [20], but also highlight that the division of labor in teams requires additional coordination work to integrate individual activities.
(2) *Generalist.* Similar to specialists, generalists are characterized by high conformity. But their range of practices is much more diverse (high diversity). Generalists are most often discussed in comparison to specialists in many of the same domains [40, 65, 69]. Literature suggests that generalists, given their greater breadth of experience, outperform specialists at making connections between different skills and domains of information as well as communicating

with others in a team [69, 73]. However, this comes at a cost – generalists are less knowledgeable and skilled in any one specific aspect of their work than a skilled specialist, limiting their effectiveness.

(3) *Maverick.* An individual who engages in a diverse range of practices (high diversity) that are rarely seen or contrary to ones that are well-established within the community (low conformity). Mavericks have been the most extensively considered in the context of scientific discovery and knowledge production [20, 30, 48, 71]. Mavericks may be driven by a constant intrinsic desire to differentiate from the rest of their community [71] or be motivated to seek greater impact by exploring previously uncharted territory [30]. In either case, the existing literature supports the argument that a maverick's strategy carries with it both high risk (i.e. most mavericks fail) and great potential for impact.

(4) *Niche Player.* An individual who is characterized by both low diversity and conformity. Niche players, as defined, have not been studied to any substantive degree in the context of teams, likely due to their rarity. Niche players require a certain level of stability in the environment to allow them to exploit a unique non-traditional strategy. While such settings can be found readily in business [28] and population ecology [51, 75] domains, most team environments may be too dynamic for successful behaviors to remain within a niche for long.

2.3 Research Questions

Based on our theoretical framework, we begin our investigation with the following question:

RQ1: Do LoL players, when characterized jointly in terms of diversity and conformity, form clusters that align with our classification?

2.3.1 Consistency in behavior. Players hone their skills and develop their expertise as they gain experience [93], and they may decide to change their gameplay styles as they acquire new knowledge of the game. Yet studies that evaluate longitudinal characteristics of user behavior have reiterated the reality that people tend to resist changes, either because they do not want to change or do not know how to change [83]. This tension has clear implications for the stability of our player classifications over time. Therefore, we raise the question:

RQ2: Do the gameplay styles of LoL players evolve over time?

2.3.2 Performance. It is in the game developer's interest to make a game desirable to all kinds of players [9, 74]. At the most basic level, this means ensuring that each player has reasonable odds of experiencing success – a fact that is reflected in LoL's goal of matching teams with roughly equivalent collective skill levels. However, each team in an LoL match is a unique combination of individual gameplay styles and some styles may systematically outperform others. Considering player styles defined based on our framework, we find that prior work has explored mechanisms that support the success of each player group. On the dimension of diversity, a targeted specialist can usually outperform a generalist with their expertise [95], but a generalist can bring in a diverse repertoire of knowledge to tackle complex situations [100]. In terms of conformity, a conformist excels by employing tried and tested strategies that are popular for the very reason that they have been historically successful, whereas a nonconformist has the element of surprise to their advantage [64, 102].

Additionally, in the context of MOBA games, success is contingent on not only individual competency but also cohesive collaboration among team members. Previous research shows that collaboration with strangers that takes place in LoL games is more complex and precarious, partly due to inadequate social interaction and communication [59]. Even collective intelligence ceases to be a good predictor of team performance in transient MOBA teams [53]. This complex interplay between attributes of individual playstyle and success motivates our next question:

RQ3: How do gameplay styles contribute to individual and team success in transient teams?

2.3.3 Adaptation. Drawing from the theory of mindsets [26], we theorize that people's tendency to diversify is a manifestation of their competing mindsets. The duality of mindset refers to a framework that describes people's perception and reaction to the ever-changing world [26]. People with a *fixed* mindset understand their ability as innate and constant, whereas those with a *growth* mindset believe that abilities are malleable and can change over time. When presented with challenging situations, a fixed mindset may foster helpless feelings and risk avoidance behaviors [26], yet a growth mindset is a predictor for perseverance and success due to their openness and willingness to try other approaches [26, 39]. One effect, though not necessarily direct or intended, of having a growth mindset is possessing a large repertoire of experiences [26]. Consequently, we argue that the growth mindset is correlated with the diversity of experience. As an example, Guay et al. [33] found that CEOs with a more diverse experience adapt better to industry shocks.

In MOBA games, frequent updates to the game made by the developers, i.e., balance patches, provide a perfect medium to explore the effect of a dynamically changing and challenging environment. As these historical updates are well documented, they enable us to investigate how different gameplay styles affect players' ability to adapt to exogenous shocks, which brings us to our final research question:

RQ4: How do gameplay styles affect players' adaptation to balance patches?

3 RELATED WORK

3.1 Multiplayer Online Battle Arena (MOBA) games

The wealth of large-scale MOBA game data has allowed researchers to investigate a range of game dynamics. Most of the focus has been on optimizing the win rate, such as predicting the best combination of hero (i.e., equivalent of champions in Dota 2) or champion lineups [37, 94] or the best teammates [90]. Other research aimed to improve game balance and matchmaking systems for a more engaging gaming experience [16]. While most matchmaking studies have considered only skills inferred from winning or losing, others have sought to improve matchmaking by incorporating players' historical position preferences [2, 72, 103]

Characterizing gameplay behavior can be approached from the individual level or the team level. At the individual level, some prior studies have used in-game statistics to profile users with unsupervised learning [76, 89]. Others have examined the effect of prolonged sessions [92] and live streaming [68] on individual performance. The effectiveness of teamwork in virtually and temporarily assembled teams has also received significant attention in the literature. Kim et al. [52] highlighted the proficiency-congruency dilemma, where players must choose between maximizing their own expertise or maximizing the teams' compatibility. Leavitt et al. [63] studied how non-verbal in-game communication affects team performance.

Although the metagame is an integral part of how the community of players interacts with the game, few have looked into how sticking to the metagame, or lack thereof, impacts performance. Players who follow the metagame likely have an advantage against those who do not, though this is not without exceptions. For instance, Lee and Ramler [64] found that LoL teams conforming to the metagame composition are more likely to succeed, but select non-meta team compositions can also perform exceptionally well. On an individual level, the literature on the influence of metagaming remains sparse.

3.2 Individual Diversity and Conformity in Teams

In more traditional domains beyond MOBA games, prior work has explored the interaction of individual knowledge diversification and heterodoxy with team performance in both production and innovation settings. Recent work on science and technology innovation shows that having team members with a broad range of experience can boost team performance and lead to greater impact, such as producing more publications [73] or more innovative patents [69, 106]. These studies validate theoretical arguments that suggest that generalists are likely to have higher absorptive capacities, i.e., they can integrate new knowledge with their existing background, a critical component of the innovation process [19]. In a similar, but distinct, scenario in a production setting, Huckman and Staats [44] showed when a temporary team of software developers works on a project, they are likely to complete work faster and produce higher quality software if some team members had experience working for different clients. In another example, from a domain more adjacent to the context of our study, Hornig et al. found that elite German footballers generally had more diverse experience in the form of unstructured play early in their careers than less successful players who have had comparatively more structured training [40]. However, the presence of generalist members does not always improve the performance of teams; production or innovation teams operating within an established domain where the knowledge space has already been thoroughly explored by the community, and surprises are rare, have no need for the unique talents of generalists. In fact, in established domains, where the work can be easily modularized, replacing generalists with specialists can improve performance [65, 69].

Compared to how generalists influence team performance, the effect of nonconforming team members on team dynamics is less well understood. Recent empirical works suggest that maverick team members represent a high risk and high reward gameplay style for a team: they can occasionally produce extraordinary team outcomes but more often lead to less than ideal results [3, 27, 63, 102]. For example, Foster et al. [27] and Alstott et al. [3] demonstrate that when scientists combine previously unrelated distant domains or problems in a new work, the recognition received from the scientific community is more varied than with conventional work, with some novel work being hailed as paradigm-shifting innovations but many receiving little to no attention.

LoL gameplay is a new setting to explore how individual behavior and team dynamics interact, providing unique advantages compared to domains employed in prior work. First, LoL players engage in direct zero-sum competitions, compared to previous settings where competition occurs indirectly at a community level (scientific community, industry, etc.). Therefore, these records provide direct and accurate indications of team-level performance compared to other settings. Further, it is difficult to isolate individual performance from team performance in production or innovation teams in industry or academia, as most measurements of performance rely on subjective peer recognition.

In contrast, individual performance metrics for LoL players are provided directly by the API. Finally, LoL developers make semi-frequent updates to the game that usually affect a small fraction of the game dynamics. As these historical updates are well documented and the times at which they were implemented are available publicly, they enable a far more precise investigation of individual contributions to team performance under uncertain conditions.

4 DATA

4.1 League of Legends Dataset

We use an LoL dataset collected by Sapienza et al. [92] using the Riot API. The dataset was collected from mid-2014 to late 2016 from the North America server on all of the matches for a sample of the most frequent players. We consider only players who have played a sufficiently large number of ranked solo-queue 5x5 games ($\geq$ 100 matches) during the entire season of 2015. This choice is driven by three factors. First, a full season can help identify the development of user behavior and performance from the start of a new season. Therefore, we take the full season (2015) that was collected during the timeframe of the dataset. Second, ranked-solo queue games allow us to investigate the behavior of players independently, as opposed to under the influence of known teammates [47]. Finally, focusing only on players with at least 100 matches has statistical and methodological benefits. Doing so can mitigate bias from players who were not committed to the game. With a sufficiently large number of matches, we can also obtain a more confident estimate of each users' skill level.

There are 404 eligible users who played in 132,000 unique matches, totaling 148,000 unique player-match pairs. In the rest of the text, we will refer to this dataset as *ALL-MATCHES*, and it will be used for the patch-based analyses. Most of our analyses, which focus on playing styles and comparison across players with different playing styles, are based solely on the first 100 matches for every user (*100-MATCHES*). This subset of the data consists of 37,300 matches and 40,400 unique player-match pairs.

The matches are organized by the time at which they started, and the winning team is indicated in the dataset. For each player, the dataset records the champion they selected out of the 127 champions available during season 2015. We further match the champions to the 7 categories of champion types set by the game developers [97] – Fighters (26), Mages (22), Marksmen (20), Tanks (20), Slayers (18), Controllers (15), and Unique Playstyles (6). Each player also selects one of five major positions, widely accepted by the community – Top, Mid, Attack Damage Carry (ADC), Support, and Jungle. Since the positions are not officially recognized, Riot does not provide these natively as part of their API. We describe how we estimate the positions in §5.1.

Additionally, the dataset contains many in-game performance measures at the individual level. From among these, the most popular metrics are *kills*, *deaths*, and *assists* counts and by extension the *kills-deaths-assists* (KDA) ratio, defined as $(kills + assists)/(1 + deaths)$ [89, 92]. A player accumulates *kills* by killing enemy players and *assists* by helping a teammate kill an opponent. The *deaths* count is the number of times a player has died. A high KDA ratio is generally desirable and is indicative of strong individual performance, regardless of the outcome of the match.

Trueskill. LoL uses an internal ranking system called the *Matchmaking Rating* (MMR) to create match teams of players that are comparable in terms of skill [46]. While the exact design of this algorithm is not publicly available, we can approximate the MMR estimates for each player reasonably well using Trueskill, a Bayesian player ranking system developed by Herbrich et al. [38]. Trueskill has been successfully employed in prior work on MOBA games that required estimates of MMR [5, 72, 90]. A player's Trueskill is represented as a normal distribution with mean skill μ and uncertainty of the skill σ. The skill level of a player is updated after every subsequent match according to the outcome of the match through Bayesian inference. We employ an open-source Python package for Trueskill[1], using all default settings except for the draw probability, which is set to 0. All players receive an initial skill level of $\mu = 25$ and $\sigma = \frac{25}{3}$. Similar to Herbrich et al. [38], we form a leaderboard of the players at the end of all of their 100th match using $\mu - 3\sigma$, a conservative estimate of their Trueskill. For brevity, we refer to this point estimate as the Trueskills of players.

4.2 Secondary Dataset

Lee and Ramler [64] collected a larger LoL dataset during a similar time frame. It contains over 10 million matches, split roughly evenly between the North American Server and the West Europe

[1] TrueSkill: The Video Game Rating System: https://trueskill.org/

Server. However, this dataset is restricted to only the second half of the season 2015, limiting its usefulness for investigating our primary research questions. Instead, we use this dataset for two secondary purposes that support the main analyses. First, we utilize this dataset, which we call *L&R*, as a training dataset to predict player positions for matches in our dataset (see §5.1) by replicating the methodology [64]. Second, we use *L&R* to perform several robustness checks by replicating some of our analyses where possible in the Supplementary Information (S.I.).

4.3 Champion balance patch data

To analyze the effect of patches on players, we identify changes to champions, in terms of nerfs and buffs, in the 21 patches of season 2015, with the exact time of patches documented in Meraki Analytics [70]. LoL releases official patch notes detailing the exact changes made to each champion but avoids making a blanket statement on whether the champion was nerfed or buffed. In many cases, multiple changes were made to a champion at once, reducing some parts of their abilities while enhancing other parts, making it difficult to determine the net effect of the changes on one champion. Fortunately, the community of players often comes to a consensus view based on the impact of the patch. One such community is a popular LoL blog called *nerfplz*[2].

We use patch notes summarized and categorized on *nerfplz* as ground truth for determining the net effects of patches on champions. The website lists the champions that the author determines to be nerfed, buffed, or tweaked in every patch. Champions are deemed to have been tweaked if they were modified but not significantly nerfed nor buffed. Throughout the 21 patches in season 2015, there were 145 champion buffs, 110 nerfs, and 29 tweaks, for a total of 284 champion modifications. Given the small number of tweaks and their mixed consequences, we disregard tweaks in our analysis. On average, each patch buffs or nerfs 12 champions. Almost all champions (107 out of 127, or 84%) were modified at some point during season 2015, and most of them were modified only once or twice.

5 A FRAMEWORK FOR USER BEHAVIOR CHARACTERIZATION

5.1 Position detection

We estimate the positions played by each player during each match by building a classifier as proposed in Lee and Ramler [64]. Using their methodology, we train a one-vs-one SVM classifier on their dataset based on different items and spells acquired by the players during a match to predict which position they likely played on. We then apply this classifier, which achieved 85% accuracy on a held-out validation set, to our dataset to estimate which one of the five positions each player chose for each match.

A more detailed discussion of this process can be found in the S.I.

5.2 User behavior indices

We seek to understand user behavior in games from the choices they make. As such, we characterize user behavior along two dimensions: diversity and conformity. In our context, diversity refers to the diversification of champion and position selection. Following Nagle and Teodoridis [73], we present the diversity index as:

$$DiversityIndex = 1 - \sqrt{\sum_{k=1}^{K} CategoryPercentage_k^2} \tag{1}$$

[2]NERFPLZ - League of Legends Guides: https://www.nerfplz.com/

where K is the number of distinct categories and $CategoryPercentage_k$ is the percentage of times category k is selected out of all matches by a user. The diversity index is closer to 1 when the level of diversification is higher and is 0 when there is absolute specialization in one category. By definition, it is always strictly less than 1. We calculate the champion ($K_C = 127$), champion type ($K_{CT} = 7$), and position ($K_P = 5$) diversity indices separately to provide a multifaceted view of the diversification of each individual.

The conformity index of a user is a measure of how well an individual's playing style agrees with that of the crowd. We estimate conformity of playing style as the average of the conditional probabilities of positions given the champions used by a player over their career:

$$ConformityIndex = \frac{1}{n} \sum_{i=1}^{n} Pr(Position_i | Champion_i) \tag{2}$$

where n is the number of matches played by an individual during season 5. $Pr(Position|Champion)$ is computed empirically for each champion and position combination from all matches played throughout season 5.

The three diversity indices and the conformity index are collectively referred to as the *behavior indices*. As diversity and conformity may develop and evolve over the course of a player's gameplay progression, we do not assume that the diversity or conformity indices are stable over time. Rather, we examine each individual's diversity and conformity indices over a sliding window of 10 matches.

5.3 Non-negative tensor factorization

Tensor factorization is a powerful tool to discover latent factors of high-dimensional data. A tensor is a higher-order generalization of a matrix and it is a useful tool for many multi-dimensional and multi-modal datasets [56]. Prior work has applied tensor factorization methods across a number of domains such as human behavior mining [41, 89], temporal signal processing [25, 91], recommender systems [42, 50, 85], and topic detection [7, 34]. Our method uses non-negative tensor factorization (NTF), a variant of tensor factorization where all latent factors are restricted to be positive. Its ease of interpretability has been shown to be a useful trait in many domain applications [7, 41, 89, 91]. Next, we briefly discuss the application of this methodology to our dataset of player careers.

Let $\mathcal{X}$ be the *Users* $\times$ *Matches* $\times$ *Behavior Indices* for the 404 users with at least 100 matches. For every user, we organize their first 100 matches sequentially and calculate the four behavior indices over a sliding window of 10 consecutive matches, resulting in 90 sets of behavior indices for each user. The dimensions of $\mathcal{X}$ are thus $404 \times 90 \times 5$. We then fit the NTF model on $\mathcal{X}$, which learns to find a set of rank-one tensors to approximate $\mathcal{X}$, subject to non-negativity constraints:

$$\boldsymbol{\mathcal{X}} \approx \sum_{r=1}^{R} \mathbf{a}_r \circ \mathbf{b}_r \circ \mathbf{c}_r, \tag{3}$$

where R is the chosen number of components and $\circ$ denotes outer product. The objective is to minimize

$$\min \left\| x_{ijk} - \sum_{m=1}^{R} \sum_{n=1}^{R} \sum_{l=1}^{R} a_{im} b_{jn} c_{kl} \right\|_F^2, \text{ such that } a_{im}, b_{jn}, c_{kl} \geq 0 \tag{4}$$

where $|| \cdot ||_F$ is the Frobenius norm. The non-negativity constraint ensures that the factors are non-negative. We use the NTF inplementation in the open-sourced Python library Tensorly [58] for our work.

The number of components, R, is an important hyperparameter; Too many or too few components can lead to overfitting or underfitting, respectively. A useful method to help determine the suitable

R is the Core Consistency diagnostics test [14], implemented in MATLAB's N-way toolbox [4]. For every $R = 1, 2, \ldots 10$, we compute the core consistency value and study the slope of the core consistency curve for the best fit model, which is the component number that leads up to the biggest drop in the core consistency values. We find $R = 3$ to be the most appropriate value after applying this test. Every user is equipped with a latent representation in $\mathbb{R}^R$, which is retained and used for further analysis.

5.4 User Clustering

NTF provides us the latent factors for each user, upon which we can cluster users who are most similar to each other by fitting the K-means algorithm. We utilize silhouette analysis to select the optimal number of clusters by plotting silhouette scores for various numbers of clusters k using the elbow method, which is where a sharp decline is observed in the plot of silhouette scores. We also visually examine the silhouette plots to check if all the clusters are similar in silhouette sizes. The number of clusters K, same as the number of components R, is determined to be 3 (silhouette score = 0.445), signifying that there are three groups of users each primarily associated with one of the three components.

6 RESULTS

Using our user behavior characterization framework, we discover users who are most similar to one another along dimensions of diversity and conformity. In §6.1, we characterize each cluster in terms of these dimensions and then discuss what these observations reveal about the gameplay styles adopted by each user group. In §6.2, we study the evolution of the gameplay styles of members of each group during season 5. Next, we analyze the overall performance of each user group during the season in §6.3 and assess how effective each group is at adapting to a balance patch in §6.4. Finally, in §6.5, we discuss the robustness of our results specifically considering potential confounding variables and sampling biases in our dataset.

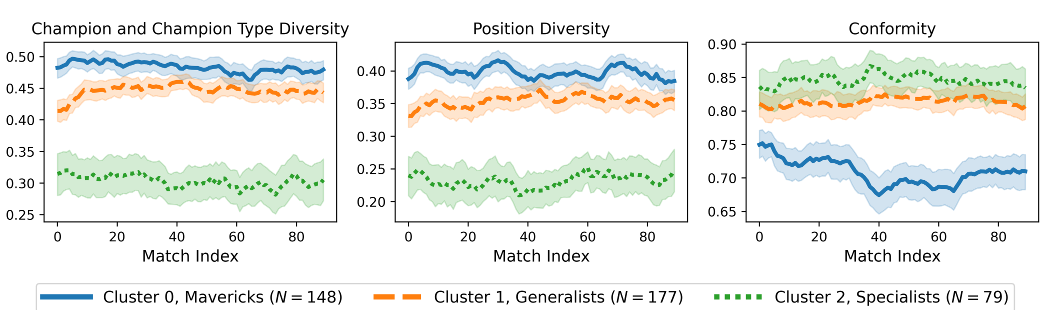

Fig. 2. Time series of behavior indices for the three clusters of user with bootstrapped 95% confidence interval. Champion and champion type diversity indices are averaged.

6.1 RQ1: Player Styles

To understand the characteristics of the player clusters, we plot the time series of the behavior indices for each of them (Fig. 2). Cluster 0 ($N = 148$), characterized by high champion diversity, high position diversity, and low conformity, is a group of nonconformists who are especially exploratory in terms of both champions and positions and have a consistent tendency to employ non-traditional combinations of them. Their behavior aligns best with *mavericks* in our framework. Cluster 2

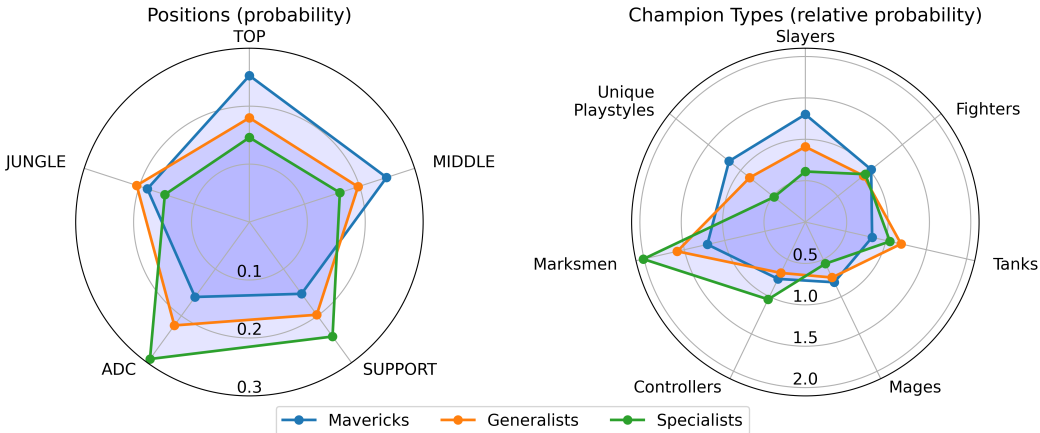

Fig. 3. Left: Proportion of matches where mavericks, generalists and specialists selected each position. Right: Relative probability of selecting a champion in each champion type compared to a uniformly random selection: $\frac{\text{count}(C_i,u)}{\sum_j \text{count}(C_j,u)} \cdot \frac{|C_i|}{K_C}$, where $\text{count}(C_i, u)$ is the number or matches played user group u with champion type C_i, $|C_i|$ is the number of champions in champion type C_i, and $K_C = 127$ is the total number of champions.

($N = 79$), characterized by low champion diversity, low position diversity, and high conformity, is a small group of conformists who exploit their skill with a limited set of champion and positions, which is the hallmark of *specialists* in our framework. Lastly, cluster 1 ($N = 177$) represents a group of players whose champion diversity, position diversity, and conformity all lie between the two polar opposites of clusters 0 and 1. However, as Figure 2 shows, the diversity of cluster 1 is nearly as high as that of cluster 0 (mavericks), while their conformity is nearly as high as that of cluster 2 (specialists). This combination of high diversity and high conformity best aligns with *generalists* in our classification. In our results, we do not observe a group of users that demonstrate characteristics of *niche players*. This may be due to, as we discussed in 2.2.3, the LoL community being too dynamic to allow niche formation in the long-term and also the relatively small number of players in our data. In the rest of the paper, we refer to the identified clusters as *mavericks* (cluster 0), *generalists* (cluster 1), and *specialists* (cluster 2), respectively.

6.1.1 Preferences in Positions and champions. Some positions and tactics may require specialization, whereas others may allow greater degrees of freedom. Indeed, there are differences in the preference of champions, champion types, and positions among the three clusters of players ($\mathcal{X}^2$ test, $p < 0.001$ for all three categories). In Figure 3, we visualize the probability that mavericks, generalists, and specialists select each position and champion type.

Compared to their peers, mavericks are most likely to select independent positions and unique champions. Their top positions are Top and Middle, followed by Jungle, all of which are conventionally solo positions. Comparatively, they favor champions who are Slayers, Fighters, and Mages. Importantly, champions branded as *Unique Playstyles* are most likely to be played by mavericks, a finding reflective of mavericks' inclination to not be bounded by rules or conventions, as *Unique Playstyles* champions are those who encompass such a diverse set of skills and abilities that the game developers could not place them in any one existing category.

Specialists most often play as ADC and Support, the two bottom lane positions who work together as a duo. This is consistent with their preferred champion types: specialists gravitate more towards Marksmen, which is almost exclusively used for ADC, and Controllers, a popular pick for

Support. There are several explanations for this correlation. For one, the bottom lane may require more specialization to be successful, or its lane design attracts a certain type of player. Alternatively, we theorize that this is a result of having to work closely with a stranger as a successful duo. Studies have shown that collaboration is difficult when team members do not have an established relationship, but having a clearly defined role for each team member can lead to better cooperation [32]. By both being conforming specialists, the pair of strangers can work most effectively together in the bottom lane.

Compared to specialists and mavericks, generalists have a more uniformly distributed selection of positions and champion types. By a small margin, they are more likely to pick Tanks, which are known as team-oriented, tough champions. Tanks help their teams succeed by disrupting enemies and absorbing the damage to themselves, but at the same time Tanks are too weak to engage in single combat. Overall, the results suggest that generalists have a wide variety of skills and expertise, and are most similar to what we typically regard as "team players".

6.2 RQ2: Evolution of gameplay styles

Figure 2 shows the temporal evolution of gameplay styles for each cluster, suggesting that players' styles are relatively stagnant over time. Even as players grow more and more familiar with the game, they are largely reluctant to change their gameplay styles. Although there are slight variations in the indices over time within each user cluster, their distinctive characteristics are stable across the clusters. The stability in user behavior is observed not only when plotted as a function of their first 100 matches but also as a function of absolute time, in terms of the match version, throughout the entire season of matches from *ALL-MATCHES*.

It should be noted that our framework characterizes a player's diversity over a sliding window. As such, it is possible for a slow but steady increase in diversity to go unnoticed. For example, a player may slowly branch out to more and more positions or champions, but their diversity indices within a short time frame remain stable. However, the average diversity indices over a sliding window of 10 for a player and the diversity index of all of their 100 matches have a Pearson's correlation of 0.93, 0.94, and 0.95 for the champion, champion type, and position diversity indices, respectively. The high correlations suggest that small window diversity indices are highly indicative of long-term diversity indices.

One follow-up observation is that players' gameplay styles remain unchanged despite the fact that, as we will see in the following sections, some styles can perform poorly under certain conditions. If players are in fact compensating for their shortcomings, they must be doing so in ways that do not fundamentally alter their gameplay styles. Another possible explanation is that players are not motivated entirely by winning or outperforming, but are driven by the pleasure of gameplay in whatever form they prefer. This is not an entirely foreign concept. For example, Kou et al. [61] finds that winning streaks can negatively impact a player's gaming experience, which demonstrates the disconnect between performance and enjoyment in certain settings. Finally, whether this persistence in behavior is conscious or not warrants further investigation. It cannot be conscious if the players are not even aware of their playing style. In other words, is their unvarying playing style a result of them not wanting to change, or them not being able to change?

6.3 RQ3: Performance

6.3.1 Overall performance. We evaluate the overall performance of all players using three metrics: win rate, Trueskill, and KDA (Figure 4). Our findings suggest that specialists have a distinct advantage over generalists and mavericks in terms of overall performance, while mavericks are, on average, the worst performers. Using a series of one-sided, one-sample t-tests against an expected win rate of 0.50, we find that generalists ($\mu = 0.504, t = 1.797, p = 0.037$) and specialists

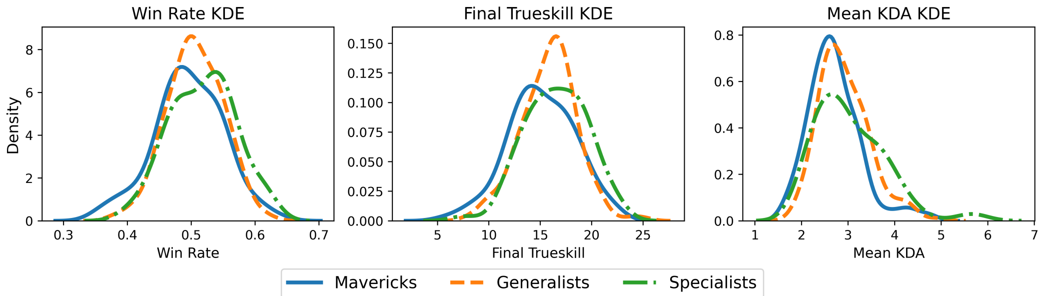

Fig. 4. KDE of win rate, final Trueskill, and mean KDA for each user's first 100 matches.

Table 1. 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.

	Win Rate CDF	**Final Trueskill CDF**	**Mean KDA CDF**
Mavericks < Generalists	$D- = 0.126, p = 0.069$	$D- = 0.176^{**}$	$D- = 0.205^{***}$
Generalists < Specialists	$D- = 0.149, p = 0.078$	$D- = 0.186^{*}$	$D- = 0.168^{*}$
Mavericks < Specialists	$D- = 0.189^{*}$	$D- = 0.211^{**}$	$D- = 0.241^{**}$

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

($\mu = 0.520, t = 3.206, p < 0.001$) are more likely to win more than 50% of the time. The win rates of Mavericks are not significantly different from 0.50 ($\mu = 0.497, t = -0.910, p = 0.182$). These differences, while quite small, are particularly significant considering that LoL's matchmaking system seeks to maintain overall win rates at 50% and the implications they have for substantially changing the odds of winning tournaments [64].

We further evaluate differences in the performance across specialists, generalists, and mavericks by comparing their performance distributions against each other using the Kolmogorov–Smirnov (K-S) test (Table 1). In terms of the win rate, specialists have a distinctly right-skewed distribution compared to mavericks, but it is otherwise not significantly different between mavericks and generalists or generalists and specialists. However, win rates in team-based games may not be the best representation of the skill of an individual player. Examining the distributions of Trueskills and mean KDA, which are more individualized measures of performance, we find that specialists collectively outperform generalists who in turn perform better than mavericks, indicating a clear hierarchy among the classes of players in terms of performance.

6.3.2 The best and worst players. Our next finding reveals which type of player is more likely to become exceptional (top 10%) or worst (bottom 10%) of all players. Similar to our previous finding regarding overall performance, We show that specialists are most likely to be the best players while mavericks remain most likely to be the worst players. In Table 2, we rank players by their final Trueskills and win rates from their first 100 matches, and find the percentage of people in each user type in the top or bottom 10% of all players. Mavericks and generalists are comparably likely to be the top 10% of players, while specialists are almost twice as likely. Considering the worst players, we find that mavericks are the most likely to end up in the bottom 10%, while specialists are the least likely.

Table 2. The proportion of players in each user group who are in the top 10% and bottom 10% of all players by win rate and by final Trueskill.

	Top 10%		**Bottom 10%**	
	Trueskill	Win Rate	Trueskill	Win Rate
Mavericks	0.088	0.122	0.155	0.182
Generalists	0.079	0.136	0.079	0.130
Specialists	0.177	0.241	0.051	0.114

Table 3. Odds ratios of a logistic regression 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.

Skill Level		**Mavericks**	**Generalists**	**Specialists**
Top	10%	1.377***	1.280***	1.319***
	10% - 20%	1.199***	1.187***	1.288***
	20% - 30%	1.141*	1.141**	1.223*
Bottom	20% - 30%	0.907*	0.894	0.882*
	10% - 20%	0.851***	0.859**	0.791**
	10%	0.718***	0.733***	0.763**

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

6.3.3 Team Performance. So far, the win rate has been treated as an individual measure of success. We now evaluate the impact of a player's skill on the team's win rate. Using a logistic regression model, we measure the effect of having each type of player across a range of skill levels. We focus on users in 6 deciles of the final Trueskill distribution: 3 deciles in the top 30% and 3 in the bottom 30%. For each match in *100-MATCHES*, we randomly select one team as the reference team. We then calculate the number of *additional* mavericks, generalists, and specialists that the reference team has in comparison to their opponents in each of the 6 performance deciles, for a total of 18 categories. The odds ratios derived from the regression is shown in Table 3. Having a top 10% maverick increases a team's odds of winning the most, but having a specialist is the next best option in the top 10% - 30% of players. In some cases, having a lower tier (top 10%-30%) specialist could be even more advantageous than a higher-tier user of another type. In cases of having a low-ranked teammate, the odds of winning are most adversely affected by having a bottom 10% maverick, but otherwise a bottom 10%–20% or 20%–30% specialist is inferior to an equivalently ranked maverick or generalist.

These findings imply that having a maverick on the team can be either a substantial risk or substantial benefit depending on the level of play: The best mavericks are particularly beneficial to the team and the worst mavericks can most dramatically bring the team down. When we move away from the two extremes, it is the specialists who have the biggest positive or negative impact on winning.

6.4 RQ4: Adaptation

In this section, we take a look at a more transient aspect of gameplay performance. Balance patches can remarkably shake up or level the playing field, altering players' champion preferences [107] and transforming the metagame. To investigate how balance patches affect players, we gather the pick rates and average KDAs of all nerfed or buffed champions before and after patches separately for each player type. We use KDA as a proxy for performance changes rather than the win rate. Win rate is largely a team performance measure in that it can depend on the combined individual performances of all team members in complex ways. KDA, in contrast, is a measure of individual performance.

LoL is patched approximately every two weeks. Therefore, we characterize the effect of a patch by comparing each champion's gameplay statistics during the 7 day period before and after a patch takes effect, which we denote t_1 and t_2, respectively. To verify that any changes in gameplay are not due to chance, we also use the period, capped at 7 days, that is one week away from the target patch as the control period t_0. For example, patch 5.5 was released on March 12, 2015, followed by 5.6 on March 25, 2015, 13 days later. For target patch 5.6, we will use the 6-day period March 12 (inclusive) to March 18 (exclusive) as the control period (t_0), the 7 day period March 18 (inclusive) to March 25 (exclusive) as the before patch period (t_1), and the 7 day period March 25 (inclusive) to April 1 (exclusive) as the after patch period (t_2). Therefore, each target patch will have three associated periods. We repeat this for all patches in season 2015, excluding patch 5.1 because it was the first patch of the season.

The pick rate of champion c_i for user type u during period t for target patch p is given by

$$\mathit{Pick\text{-}Rate}(c_i|u,t,p) = \frac{\text{count}(c_i,u,t,p)}{\sum_j \text{count}(c_j,u,t,p)}, \tag{5}$$

where $\text{count}(c_i,u,t,p)$ is the number of times a type u user selected champion c_i during period t for patch p. We illustrate the changes in pick rates in figure 5.

Consistent with previous work, buffing champions raises the preference of those champions and vice versa for champions that were nerfed [107]. We include the pick rate changes for unaltered champions as a baseline, which shows virtually no changes in pick rate. We use paired t-tests to analyze the differences in pick rates between consecutive periods (Figure 5, details in S.I.).

Our first and immediate observation is that buffs have a limited effect on increasing champion preference, yet nerfs noticeably curtail champion popularity. Furthermore, different patches affect each player type differently. Generalists display a significant increase in the preference for buffed champions, but mavericks and specialists do not. Generalists are also sensitive to nerfs, suggesting that their gameplay preferences are most directly influenced by balance patches. While a significant decline in champion preference is observed in all users, the effect is smallest for specialists. Specialists remain the most loyal to the champions they specialize in even when these champions are nerfed. At the same time, specialists are not inclined to increase their champion preferences in spite of a buff patch. Their preferences remain the most stable compared to other user types. Finally, we find that mavericks' champion preferences, although not significantly impacted by buff patches, are the most dramatically reduced by nerf patches compared to other players.

Figure 6 illustrates the effect of patches on game performance, measured by the mean champion KDAs during a period. We apply Wilcoxon signed-rank tests to check for any significant differences in the mean KDAs of each champion between consecutive periods (details in S.I.). Overall, buff patches lead to larger changes in KDA compared to those of the control periods, but the effect is not significant for any user type. Moreover, generalists adapt rather well to both nerfs and buffs, showing no significant change in KDA. In contrast, nerf patches significantly decrease the KDA output for mavericks ($\Delta KDA = -0.217$) and specialists ($\Delta KDA = -0.580$).

All players are quick to react to nerfs by significantly turning away from the nerfed champions, yet those who stay with the nerfed champions show varying levels of success. Of all, generalists are the least impacted by the nerf patches, indicating they can quickly adjust to champions with diminished skills. Both specialists and mavericks are apparently less successful at adapting to nerfs. Specialists

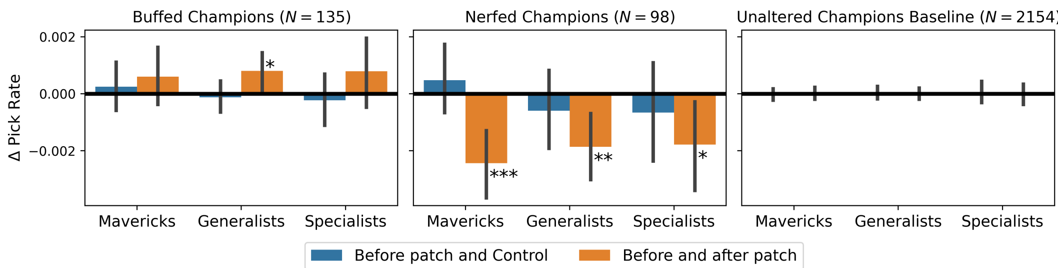

Fig. 5. The champion pick rate differences in the before patch period t_1 vs. the control period t_0 (*"Before patch and Control"*) and the after patch period t_2 vs. before patch period t_1 patch (*"Before and after patch"*). The total number of champion buffs, nerfs or unaltered champions are shown. We apply paired t-tests on the pick rates between consecutive periods and mark all significant differences ($^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$).

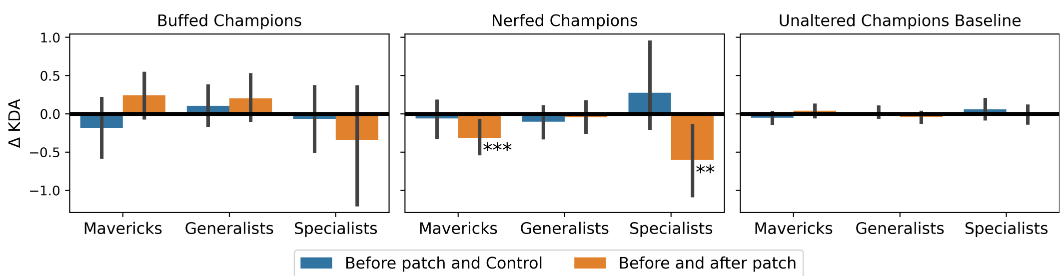

Fig. 6. The difference in each champion's mean KDA in the before patch period t_1 vs. the control period t_0 (*"Before patch and Control"*) and the after patch period t_2 vs. the before patch period t_1 patch (*"Before and after patch"*). We apply Wilcoxon signed-rank tests on champions' mean KDAs between consecutive periods and mark all significant differences ($^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$).

may adapt poorly because they lack wide-ranging skills to quickly master modified champions, and mavericks are likely to neglect widely accepted guides developed by the community in response to the patch. Generalists can avoid the pitfalls of both: they have a diverse set of experiences and also follow the established community conventions.

Although our results show that mavericks and specialists are the most negatively impacted by nerfs, no user type significantly benefits from buffs. This phenomenon could be partly explained by the learning curve associated with playing recently updated champions. Players need to learn how to best navigate a champion's new skills and abilities, even though the patch was meant to be a buff. Though it is difficult to observe the long-term effects of adaptation due to the nature of the rapidly changing gaming context in LoL, future work can evaluate players' long-term adaptation to more substantial changes in the environment.

6.5 Robustness checks and confounding factors

Player experience is a confounding variable that could potentially explain our results. For example, the individuals labeled as mavericks may simply be new to the game and are still learning to play the game prior to settling down on a gameplay style. However, we have restricted our attention to ranked games, which are only available to players who have reached a certain level of experience by which time we can reasonably assume that they would have settled down to a specific playing style. Additionally, to validate that experience is not determining our clusters, we look at matches played during late season 2014 and pre-season 2015. The proportion of users who participated during this time are nearly identical across the 3 user groups, being 90%, 88%, and 89% for mavericks, generalists, and specialists, respectively. Moreover, the high proportion of players with prior experience is evidence that most players in our analysis are experienced players. We also compute the number of games played by each individual in season 2014 and pre-season 2015. A Kruskal-Wallis H test indicates no significant difference in the number of prior games played by users in each group ($\chi^2 = 3.430, p = 0.180$), with the average number of prior games being 253 for mavericks, 241 for generalists, and 186 for specialists. Thus, these results do not support the idea that mavericks, or any particular player cluster, are significantly less (or more) experienced.

In the absence of native time-evolving skill ratings for players from the Riot API, we have used the Trueskills algorithm to estimate ratings. However, this algorithm requires that we have the entire timeline of matches for an individual to derive an absolute ranking. Unfortunately, our dataset only contains partial timelines for the players starting with the second half of season 2014. Thus our Trueskill estimates, measured only based on season 2015 matches, are in fact more appropriately interpreted as changes in the absolute Trueskill during that season. This limits our ability to interpret our findings in terms of players being on a global skill hierarchy as any single player would have had their own unique unobserved skill rating at the start of our study period. However, this confounding effect is mitigated by LoL's matchmaking system, which only matches people in similar tiers together. As a result, while the Trueskill may not be an accurate representation of a player's skill level, it does represent how well a player has done since the beginning of the season. As an additional robustness check, we repeat our analysis in §6.3.3 with a mixed effect model, using their highest achieved tier in the previous season, which is an approximate measure of their skill at the start of the season, as a fixed effect. We report the details of this experiment in the S.I. This model confirms our findings regarding the relative contributions of skilled versus unskilled mavericks, generalists, and specialists.

As a final robustness check, we replicate our analyses, from discovering user types to analyzing individual and team-level performance, on the L&R dataset [64]. Since it was collected from a different server (EU), it helps to remove some cultural biases from our main dataset. We present the results of this additional analysis in the S.I. and confirm that our findings between the two datasets are consistent.

7 DISCUSSION

In this study, we explored the dynamics of preference and its interaction with performance in transient teams of strangers in competitive settings. Drawing on theories of individual behavior in teams, we established a framework for classifying individuals in teams based on the dimensions of diversity and conformity. Through computational analysis, we identified three prominent types of players in League of Legends that aligned with our framework; (i) *mavericks*, (ii) *generalists*, and (iii) *specialists*. We did not find a fourth type established by theory – *niche players* with low diversity and low conformity. As LoL is a very dynamically changing game, any niche behaviors may be difficult to observe in the long-term. Another possible explanation may be that there are fewer niche players, which prohibited us from discovering a cluster of such players. However, we reached the same conclusion about the existence of generalists, specialists, mavericks, and reinforced the lack of niche players, in our secondary dataset (*L&R*), which has over 10 times the number of players in our main dataset (see S.I.). Additional field studies can be conducted to validate if niche players do exist in such settings.

That LoL players demonstrate stable preferences in playing styles over time implies motivations go beyond winning games. Despite specialists being the most likely to perform best overall, most people remain exploratory and unconventional throughout their gaming careers. This phenomenon can be traced back to psychological theories of novelty-seeking [18] and risk-taking personality traits [55]. In a short-term, volatile setting, generalists are observed to adapt best both to beneficial and detrimental changes while specialists are found to be most susceptible to detrimental changes.

The observations in this study seemingly contradict similar recent research which looked at the performance of teams in scientific research and peer-production settings [65, 69, 73]. While these studies demonstrate the presence of team members with broad and atypical background experiences leads to greater impact in the scientific community, we show that – in the case of LoL – specialists perform substantially better than individuals who have a broad range of experience or try novel strategies. However, a more nuanced examination shows that these contradictions can be explained by differences in settings. League of Legends is a comparatively stable environment where developer updates only affect a few players at any given time. This has allowed the player community to construct a well-documented metagame that leaves very little uncertainty about how different strategies perform. In contrast, many domains of scientific inquiry contend with significant uncertainty and depend on exploration. In fact, Nagle and Teodoridis [73] show that even in the case of science, generalists and mavericks are superfluous or even an active hindrance in well-established domains with little uncertainty, in contrast to developing domains that benefit from exploration. This is further supported by our own finding that generalists and mavericks are more resilient than specialists to changes in the game environment that negatively affect them.

The presence of skilled mavericks who have an outsized impact on team performance indicates that even with a mature and established metagame, there are unorthodox strategies that offer distinct performance gains. Further investigation is necessary to find whether mavericks play a role in improving the quality and competitiveness of play by serving as trailblazers for the rest of the community.

Implications for game design and user-related research. For game developers and interested researchers, our paper confirms the need to consider users with different playstyles. Games need to be made competitive and appealing both to those who play-it-safe and those who, on the other end of the spectrum, derive joy from taking risks in the unknown. Gamers either *go deep* (exploit) or *go wide* (explore) in the game and rarely change their habits once they are formed. It is thus evident that there is no one-size-fits-all approach to game design. Importantly, our paper reveals how the heterogeneity of styles can impact players' performance and adaptation. This can benefit other game-related research such as improving the matchmaking system, reducing player churn (quitting the game), and balancing the game with patches. Developers should be aware of what attracts and deters different types of users, as well as who is more or less susceptible to changes in the game. Since empirical research on virtual worlds is invaluable for studying real-world behaviors [109], our framework can be extended to other domains such as academic research collaborations, corporate team building, and competitive sports. The constructs we employed to describe user diversity and conformity in the context of small teams can be utilized to understand phenomena in more general societal and organizational settings.

Implications for team coordination. Due to the nature of transient teams in LoL and its matchmaking algorithm, most teams are not only transient but also opportunistic, with each player primarily incentivized to improve their own rankings [53]. Recognizing that different player types have different preferences for positions and champions (§6.1.1), the flexibility with which players can play their favored positions or champions can crucially impact their performance. For example, specialists may only want to play in a certain position or in a particular way. If that is the case, it could be detrimental to the team and to the specialists themselves if this was not communicated properly to the team early on. Generalists may be suitable for a range of positions but will do poorly if a particular role, task, or champion calls for appropriate specialization. Mavericks' way of playing may be initially ill-received by their new teammates. One way to remedy these potential conflicts is to acknowledge or justify unconventional gameplay styles to avoid any negative impressions. Kou and Gui [59] finds that players can get emotional and aggressive when social interaction is lacking and when the team is performing poorly. Similar recommendations can be applied to other transient team settings, for instance, when a temporarily assembled medical team responds to a medical emergency, when a research project calls for multi-institution collaboration, or when corporate employees are arranged in regularly reshuffled fluid teams. It is integral for the team leader to balance both individual teammates' preferences and the team's collective congruency, and relay relevant information to the team.

Implications for individual team players. As our study highlights the advantages and disadvantages of different behavioral styles, it is worth considering if promoting behavioral changes in team members can be beneficial. If the end goal of the team is to maximize achievement, it may be useful to start by understanding what behaviors the players tend to exhibit. It is crucial that the playstyle asked of the player aligns with their intrinsic preferences if any changes are to bring out the full potential of the team and individual. A common form of miscommunication arises when players offer advice to others by only relaying what worked for themselves, without considering other's style or preferences. A maverick, for example, may find it difficult to heed advice from a specialist because they do not enjoy repetition.

That said, not all players are driven by achievement [110]. It is important to not criticize another player because they are playing the game in a different manner. As we have seen in this paper, no single gaming behavior is without its weakness. Even strategies that go against the established metagames have merit. To foster harmonious gameplay experiences and non-toxic social interactions, players must be aware that other preferences for gaming exist and to devise team-level strategies that combine the strength of individual members.

Implications for team sports. Unlike transient LoL teams, players in traditional sports teams have well-defined and heavily constricted roles [24, 104]. There exists a limited body of research suggesting individual diversity could be beneficial. In professional basketball, some diversity in playstyles is beneficial, but too much can present challenges [96]. In soccer, the cultural diversity of players can have a positive impact on team performance [45], while the diversity of experience, through unstructured play early in their careers, have been shown to correspond to greater eventual success [40].

As specialization is regarded as the norm in traditional team sports, it is worth questioning if generalists and mavericks feel unable to explore alternative positions or strategies, thereby hindering their enjoyment of the sport and development as players.

8 LIMITATIONS

A primary limitation of our study is the generalizability. Our study is based on data sampled from a single season of LoL from the North American server. Moreover, as our research questions require longitudinal analysis of individual players, we have limited our analyses to players who have participated in a large number of matches during this time frame. Consequently, we rely on a relatively small sample of individuals from a single region. We cannot remove any data biases or cultural biases that may arise from this, and thus cannot conclusively say that our results generalize to players with less experience or players from other regions. However, statistically significant results from our analyses align very well with similar work in other domains such as science and technology innovation and peer-production. Further, in the S.I., we show that our main findings hold in a much larger dataset collected from a different geographical region (the EU server). Future work could be directed at comparing and contrasting the gaming behavior of players hailing from various cultural backgrounds.

Our analysis of the interaction between individual playstyles and team-level outcomes does not employ causal methods and therefore we cannot make strong claims regarding how individual playstyles influence team outcomes. However, LoL's matchmaking algorithm, the objective of which is to give equal odds of winning to the two teams, takes skill and experience at the team level into account [36, 66]. Additionally, there's no publicly available information that suggests that matchmaking in ranked solo games considers playstyles, as we've defined in this paper. Therefore, we argue that our model coefficients provide strong circumstantial evidence regarding the relative influence of different individual playstyles on team outcomes. Since we are not privy to the exact matchmaking algorithm used in LoL, especially historical iterations of it that are relevant for our study, we assume that any deviations from the stated goal of applying the matchmaking algorithm are unbiased with respect to the variables of interest in the study.

Finally, our analyses rely on coarse-grained labels for players identified through clustering. A direct benefit of this method is the ability to gather sufficient data points to perform a holistic assessment of the performance and adaptation of all user groups. Future work can develop a more nuanced evaluation of performance and adaptation dynamics based on continuous measures of player diversity and conformity.

ACKNOWLEDGMENTS

The authors are grateful to DARPA for support under AIE Opportunity No. DARPA-PA-18-02-07. We thank Ivan Ramler for generously providing access to data from [64] and Daniel M. Romero for useful feedback.

REFERENCES

[1] Icek Ajzen et al. 1991. The theory of planned behavior. *Organizational behavior and human decision processes* 50, 2 (1991), 179–211. https://doi.org/10.1016/0749-5978(91)90020-T

[2] Josh Alman and Dylan McKay. 2017. Theoretical foundations of team matchmaking. In *Proc. AAMAS '17*. IFAAMAS, 1073–1081. https://doi.org/10.5555/3091125.3091277

[3] Jeff Alstott, Giorgio Triulzi, Bowen Yan, and Jianxi Luo. 2017. Inventors' explorations across technology domains. *Design Science* 3 (2017), e20. https://doi.org/10.1017/dsj.2017.21

[4] Claus A Andersson and Rasmus Bro. 2000. The N-way toolbox for MATLAB. *Chemometrics and Intelligent Laboratory Systems* 52, 1 (2000), 1–4.

[5] Aleo Aninda. 2019. *Meta discovery and role-based matchmaking (MEDIROMA)*. Master's thesis. Flinders University, College of Science and Engineering.

[6] Solomon E Asch. 1956. Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological monographs: General and applied* 70, 9 (1956), 1.

[7] Brett W Bader, Michael W Berry, and Murray Browne. 2008. Discussion tracking in Enron email using PARAFAC. In *Survey of Text Mining II: Clustering, Classification, and Retrieval*. Springer, 147–163. https://doi.org/10.1007/978-1-84800-046-9_8

[8] René M Bakker, Smaranda Boroş, Patrick Kenis, and Leon AG Oerlemans. 2013. It's only temporary: time frame and the dynamics of creative project teams. *British Journal of Management* 24, 3 (2013), 383–397. https://doi.org/10.1111/j.1467-8551.2012.00810.x

[9] Richard Bartle. 1996. Hearts, clubs, diamonds, spades: Players who suit MUDs. *Journal of MUD research* 1, 1 (1996), 19.

[10] R Meredith Belbin. 2011. Management teams: Why they succeed or fail. *Human Resource Management International Digest* (2011).

[11] Rick A Bevins. 2001. Novelty seeking and reward: Implications for the study of high-risk behaviors. *Current Directions in Psychological Science* 10, 6 (2001), 189–193.

[12] Justin W Bonny, Mike Scanlon, and Lisa M Castaneda. 2020. Variations in psychological factors and experience-dependent changes in team-based video game performance. *Intelligence* 80 (2020), 101450.

[13] John H Bradley and Frederic J Hebert. 1997. The effect of personality type on team performance. *Journal of Management Development* (1997).

[14] Rasmus Bro and Henk AL Kiers. 2003. A new efficient method for determining the number of components in PARAFAC models. *Journal of Chemometrics* 17, 5 (2003), 274–286.

[15] Anis Chaouachi, Matt Brughelli, Gregory Levin, Nahla Ben Brahim Boudhina, John Cronin, and Karim Chamari. 2009. Anthropometric, physiological and performance characteristics of elite team-handball players. *Journal of sports sciences* 27, 2 (2009), 151–157.
[16] Mark Claypool, Jonathan Decelle, Gabriel Hall, and Lindsay O'Donnell. 2015. Surrender at 20? Matchmaking in league of legends. In *IEEE GEM '15*. IEEE, 1–4. https://doi.org/10.1109/GEM.2015.7377234
[17] Mark Claypool, Artian Kica, Andrew La Manna, Lindsay O'Donnell, and Tom Paolillo. 2017. On the impact of software patching on gameplay for the League of Legends computer game. *The Computer Games Journal* 6, 1-2 (2017), 33–61.
[18] C. Robert Cloninger, Thomas R. Przybeck, Dragan M. Svrakic, and Richard D. Wetzel. 1994. *The Temperament and Character Inventory (TCI): A guide to its development and use.* Center for Psychobiology of Personality, Washington University, St. Louis, MO.
[19] Wesley M Cohen and Daniel A Levinthal. 1990. Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly* 35, 1 (1990), 128–152.
[20] Adrian Currie. 2019. Existential risk, creativity & well-adapted science. *Studies in History and Philosophy of Science Part A* 76 (2019), 39–48.
[21] Filip De Fruyt, Lieve Van de Wiele, and Cornelis Van Heeringen. 2000. Cloninger's psychobiological model of temperament and character and the five-factor model of personality. *Personality and Individual Differences* 29, 3 (2000), 441–452.
[22] Tanner Dedmon. 2018. Riot reveals most popular 'League of Legends' game modes by region. Retrieved October 14, 2020 from https://comicbook.com/gaming/news/league-of-legends-most-popular-game-modes-by-region/
[23] Scott Donaldson. 2017. Mechanics and metagame: Exploring binary expertise in League of Legends. *Games and Culture* 12, 5 (2017), 426–444. https://doi.org/10.1177/1555412015590063
[24] Ricardo Duarte, Duarte Araújo, Vanda Correia, and Keith Davids. 2012. Sports teams as superorganisms. *Sports medicine* 42, 8 (2012), 633–642.
[25] Daniel M. Dunlavy, Tamara G. Kolda, and Evrim Acar. 2011. Temporal link prediction using matrix and tensor factorizations. *ACM TKDD* 5, 2, Article 10 (Feb. 2011), 27 pages. https://doi.org/10.1145/1921632.1921636
[26] Carol S Dweck. 2008. *Mindset: The new psychology of success.* Random House, New York.
[27] Jacob G Foster, Andrey Rzhetsky, and James A Evans. 2015. Tradition and innovation in scientists' research strategies. *American Sociological Review* 80, 5 (2015), 875–908.
[28] Elena Galateanu and Silvia Avasilcai. 2016. Framing the Competitive Behaviors of Niche Players: The Electric Vehicle Business Ecosystem Perspective. *Procedia-Social and Behavioral Sciences* 221 (2016), 342–351.
[29] Shuli Gao, Yanli Guo, Jianbin Chen, and Lin Li. 2016. Factors affecting the performance of knowledge collaboration in virtual team based on capital appreciation. *Information Technology and Management* 17, 2 (2016), 119–131.
[30] Elliroma Gardiner and Chris J Jackson. 2012. Workplace mavericks: How personality and risk-taking propensity predicts maverickism. *British journal of psychology* 103, 4 (2012), 497–519.
[31] Samuel D Gosling, Peter J Rentfrow, and William B Swann Jr. 2003. A very brief measure of the Big-Five personality domains. *Journal of Research in Personality* 37, 6 (2003), 504–528.
[32] Lynda Gratton and Tammy Erickson. 2007. Eight ways to build collaborative teams. Retrieved October 14, 2020 from https://hbr.org/2007/11/eight-ways-to-build-collaborative-teams
[33] Wayne R Guay, Daniel J Taylor, and Jason J Xiao. 2015. Adapt or perish: Evidence of CEO adaptability to industry shocks. *Available at SSRN 2234886* (2015). https://doi.org/10.2139/ssrn.2234886
[34] Xin Guo, Yang Xiang, Qian Chen, Zhenhua Huang, and Yongtao Hao. 2013. LDA-based online topic detection using tensor factorization. *Journal of Information Science* 39, 4 (2013), 459–469. https://doi.org/10.1177/0165551512473066
[35] Carolin Haeussler and Henry Sauermann. 2020. Division of labor in collaborative knowledge production: The role of team size and interdisciplinarity. *Research Policy* 49, 6 (2020), 103987.
[36] Gabriel Glass Hall, Jonathan M Decelle, and Lindsay R O'Donnell. 2015. *The importance of matchmaking in League of Legends and its effects on Users.* Technical Report. Worcester Polytechnic Institute. https://digitalcommons.wpi.edu/iqp-all/765/
[37] Lucas Hanke and Luiz Chaimowicz. 2017. A recommender system for hero line-ups in MOBA games. In *Proc. AIIDE '17*. AAAI, 43–49.
[38] Ralf Herbrich, Tom Minka, and Thore Graepel. 2007. TrueSkill™: A Bayesian skill rating system. In *Proc. NIPS '07*. MIT Press, 569–576.
[39] Aaron Hochanadel and Dora Finamore. 2015. Fixed and growth mindset in education and how grit helps students persist in the face of adversity. *Journal of International Education Research* 11, 1 (2015), 47–50.
[40] Manuel Hornig, Friedhelm Aust, and Arne Güllich. 2016. Practice and play in the development of German top-level professional football players. *European Journal of Sport Science* 16, 1 (2016), 96–105.
[41] Homa Hosseinmardi, Hsien-Te Kao, Kristina Lerman, and Emilio Ferrara. 2019. Discovering hidden structure in high dimensional human behavioral data via tensor factorization. (2019). arXiv:1905.08846

[42] Yang Hu, Xi Yi, and Larry S. Davis. 2015. Collaborative fashion recommendation: A functional tensor factorization approach. In *Proc. MM '15*. ACM, 129–138. https://doi.org/10.1145/2733373.2806239

[43] Han-Chung Huang, Li-Shia Huang, Yu-Jen Chou, and Ching-I Teng. 2017. Influence of temperament and character on online gamer loyalty: Perspectives from personality and flow theories. *Computers in Human Behavior* 70 (2017), 398 – 406. https://doi.org/10.1016/j.chb.2017.01.009

[44] Robert S Huckman and Bradley R Staats. 2011. Fluid tasks and fluid teams: The impact of diversity in experience and team familiarity on team performance. *Manufacturing & Service Operations Management* 13, 3 (2011), 310–328.

[45] Keith Ingersoll, Edmund J Malesky, and Sebastian M Saiegh. 2013. Heterogeneity and Group Performance: Evaluating the Effect of Cultural Diversity in the World's Top Soccer League. *Available at SSRN 2333289* (2013).

[46] Isto. 2017. Matchmaking guide. Retrieved October 14, 2020 from https://support-leagueoflegends.riotgames.com/hc/en-us/articles/201752954-Matchmaking-Guide

[47] Daniel Johnson, Lennart E. Nacke, and Peta Wyeth. 2015. All about that base: Differing player experiences in video game genres and the unique case of MOBA Games. In *Proc. CHI '15*. ACM, 2265–2274. https://doi.org/10.1145/2702123.2702447

[48] Candace Jones, Silviya Svejenova, Jesper Strandgaard Pedersen, and Barbara Townley. 2016. Misfits, mavericks and mainstreams: Drivers of innovation in the creative industries.

[49] Adam S Kahn and Dmitri Williams. 2016. We're all in this (game) together: Transactive memory systems, social presence, and team structure in multiplayer online battle arenas. *Communication Research* 43, 4 (2016), 487–517.

[50] Alexandros Karatzoglou, Xavier Amatriain, Linas Baltrunas, and Nuria Oliver. 2010. Multiverse recommendation: N-dimensional tensor factorization for context-aware collaborative filtering. In *Proc. RecSys '10*. ACM, 79–86. https://doi.org/10.1145/1864708.1864727

[51] Michael Kearney, Stephen J Simpson, David Raubenheimer, and Brian Helmuth. 2010. Modelling the ecological niche from functional traits. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365, 1557 (2010), 3469–3483.

[52] 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

[53] Young Ji Kim, David Engel, Anita Williams Woolley, Jeffrey Yu-Ting Lin, Naomi McArthur, and Thomas W Malone. 2017. What makes a strong team? Using collective intelligence to predict team performance in League of Legends. In *Proc. CSCW '17*. ACM, 2316–2329. https://doi.org/10.1145/2998181.2998185

[54] Aniket Kittur, Bryant Lee, and Robert E Kraut. 2009. Coordination in collective intelligence: the role of team structure and task interdependence. In *Proc. CHI '09*. ACM, 1495–1504. https://doi.org/10.1145/1518701.1518928

[55] Nathan Kogan and Michael A Wallach. 1964. *Risk taking: A study in cognition and personality.* Holt, Rinehart & Winston, Oxford, England.

[56] Tamara G Kolda and Brett W Bader. 2009. Tensor decompositions and applications. *SIAM review* 51, 3 (2009), 455–500.

[57] Dorien Kooij, Jaap Paauwe, Karin Sanders, Gráinne Kelly, Michele Mastroeni, Edel Conway, Kathy Monks, Katie Truss, Patrick Flood, and Enda Hannon. 2011. Combining diverse knowledge: knowledge workers' experience of specialist and generalist roles. *Personnel Review* (2011).

[58] Jean Kossaifi, Yannis Panagakis, Anima Anandkumar, and Maja Pantic. 2019. TensorLy: Tensor Learning in Python. *The Journal of Machine Learning Research* 20, 1 (Jan. 2019), 925–930.

[59] Yubo Kou and Xinning Gui. 2014. Playing with strangers: Understanding temporary teams in League of Legends. In *Proc. CHI PLAY '14*. ACM, 161–169. https://doi.org/10.1145/2658537.2658538

[60] Yubo Kou, Xinning Gui, and Yong Ming Kow. 2016. Ranking practices and distinction in League of Legends. In *Proc. CHI PLAY '16*. ACM, 4–9. https://doi.org/10.1145/2967934.2968078

[61] Yubo Kou, Yao Li, Xinning Gui, and Eli Suzuki-Gill. 2018. Playing with streakiness in online games: How players perceive and react to winning and losing streaks in League of Legends. In *Proc. CHI '18*. ACM, 1–14. https://doi.org/10.1145/3173574.3174152

[62] Steve WJ Kozlowski and Daniel R Ilgen. 2006. Enhancing the effectiveness of work groups and teams. *Psychological Science in the Public Interest* 7, 3 (2006), 77–124.

[63] Alex Leavitt, Brian C. Keegan, and Joshua Clark. 2016. Ping to win? Non-verbal communication and team performance in competitive online multiplayer games. In *Proc. CHI '16*. ACM, 4337–4350. https://doi.org/10.1145/2858036.2858132

[64] 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

[65] Jürgen Lerner and Alessandro Lomi. 2018. Diverse teams tend to do good work in Wikipedia (but jacks of all trades don't). In *Proc. ASONAM '18*. IEEE, 214–221. https://doi.org/10.1109/ASONAM.2018.8508597

[66] Jeffrey Lin and Paul Sottosanti. 2017. Systems and methods that enable player matching for multi-player online games. US Patent 9,795,887.

[67] Viktoriya Lipovaya, Yuri Lima, Pedro Grillo, Carlos Eduardo Barbosa, Jano Moreira de Souza, and Francisco José de Castro Moura Duarte. 2018. Coordination, communication, and competition in eSports: A comparative analysis of teams in two action games. In *Proc. ECSCW '18*. EUSSET, Europe. https://doi.org/10.18420/ecscw2018_11
[68] Akira Matsui, Anna Sapienza, and Emilio Ferrara. 2020. Does streaming esports affect players' behavior and performance? *Games and Culture* 15, 1 (2020), 9–31.
[69] Eduardo Melero and Neus Palomeras. 2015. The Renaissance Man is not dead! The role of generalists in teams of inventors. *Research Policy* 44, 1 (2015), 154–167.
[70] Meraki Analytics. 2020. League of Legends Static Data. https://github.com/meraki-analytics/lolstaticdata
[71] Ryan Muldoon. 2017. Diversity, rationality and the division of cognitive labor. *Scientific collaboration and collective knowledge: New essays* (2017), 78–91.
[72] Mateusz Myślak and Dominik Deja. 2015. Developing game-structure sensitive matchmaking system for massive-multiplayer online games. In *Social Informatics*. Springer International Publishing, 200–208.
[73] Frank Nagle and Florenta Teodoridis. 2020. Jack of all trades and master of knowledge: The role of diversification in new distant knowledge integration. *Strategic Management Journal* 41, 1 (2020), 55–85. https://doi.org/10.1002/smj.3091 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/smj.3091
[74] Jeannie Novak. 2011. *Game development essentials: an introduction*. Cengage Learning, Boston, MA.
[75] Eli Noy. 2010. Niche strategy: Merging economic and marketing theories with population ecology arguments. *Journal of Strategic Marketing* 18, 1 (2010), 77–86.
[76] Hao Yi Ong, Sunil Deolalikar, and Mark Peng. 2015. Player behavior and optimal team composition for online multiplayer games. (2015). arXiv:1503.02230
[77] Scott E Page. 2019. *The diversity bonus: How great teams pay off in the knowledge economy*. Princeton University Press, Princeton, NJ.
[78] Ismael Pedraza-Ramirez, Lisa Musculus, Markus Raab, and Sylvain Laborde. 2020. Setting the scientific stage for esports psychology: A systematic review. *International Review of Sport and Exercise Psychology* (2020), 1–34.
[79] Annie Pei. 2019. League of Legends is now 10 years old but may be tops in esports. Here's why. Retrieved October 14, 2020 from https://www.cnbc.com/2019/10/15/league-of-legends-may-be-tops-in-esports-heres-why.html
[80] Edward Platt and Daniel Romero. 2018. Network structure, efficiency, and performance in WikiProjects. In *Proc. ICWSM '18*. AAAI, 251–260.
[81] Steven Postrel. 2002. Islands of shared knowledge: Specialization and mutual understanding in problem-solving teams. *Organization science* 13, 3 (2002), 303–320.
[82] Harry Prapavessis and Albert V Carron. 1997. Sacrifice, cohesion, and conformity to norms in sport teams. *Group Dynamics: Theory, Research, and Practice* 1, 3 (1997), 231.
[83] James O Prochaska and Janice M Prochaska. 1999. Why don't continents move? Why don't people change? *Journal of Psychotherapy Integration* 9, 1 (1999), 83–102.
[84] Richard R Reilly, Gary S Lynn, and Zvi H Aronson. 2002. The role of personality in new product development team performance. *Journal of Engineering and Technology Management* 19, 1 (2002), 39–58.
[85] Steffen Rendle and Lars Schmidt-Thieme. 2010. Pairwise interaction tensor factorization for personalized tag recommendation. In *Proc. WSDM '10*. ACM, 81–90. https://doi.org/10.1145/1718487.1718498
[86] Daniela Retelny, Sébastien Robaszkiewicz, Alexandra To, Walter S Lasecki, Jay Patel, Negar Rahmati, Tulsee Doshi, Melissa Valentine, and Michael S Bernstein. 2014. Expert crowdsourcing with flash teams. In *Proc. UIST '14*. ACM, 75–85. https://doi.org/10.1145/2642918.2647409
[87] Daniel M. Romero, Dan Huttenlocher, and Jon Kleinberg. 2015. Coordination and efficiency in decentralized collaboration. In *Proc. ICWSM '15*. AAAI, 367–376.
[88] Katie Tekinbaş Salen and Eric Zimmerman. 2004. *Rules of play: Game design fundamentals*. MIT press, Cambridge, MA.
[89] Anna Sapienza, Alessandro Bessi, and Emilio Ferrara. 2018. Non-negative tensor factorization for human behavioral pattern mining in online games. *Information* 9, 3 (2018), 66.
[90] Anna Sapienza, Palash Goyal, and Emilio Ferrara. 2019. Deep neural networks for optimal team composition. *Frontiers in Big Data* 2 (2019), 14.
[91] Anna Sapienza, André Panisson, Joseph Wu, Laetitia Gauvin, and Ciro Cattuto. 2015. Detecting anomalies in time-varying networks using tensor decomposition. In *Proc. ICDMW '15*. IEEE, 516–523. https://doi.org/10.1109/ICDMW.2015.128
[92] Anna Sapienza, Yilei Zeng, Alessandro Bessi, Kristina Lerman, and Emilio Ferrara. 2018. Individual performance in team-based online games. *Royal Society Open Science* 5, 6 (2018), 180329.
[93] Peter G Schrader and Michael McCreery. 2008. The acquisition of skill and expertise in massively multiplayer online games. *Educational Technology Research and Development* 56, 5-6 (2008), 557–574.

[94] Aleksandr Semenov, Peter Romov, Sergey Korolev, Daniil Yashkov, and Kirill Neklyudov. 2016. Performance of machine learning algorithms in predicting game outcome from drafts in Dota 2. In *Proc. AIST '16*. Springer, 26–37. https://doi.org/10.1007/978-3-319-52920-2_3

[95] Hermann Simon. 1996. You don't have to be German to be a "Hidden Champion". *Business Strategy Review* 7, 2 (1996), 1–13. https://doi.org/10.1111/j.1467-8616.1996.tb00118.x

[96] Edward Bishop Smith and Yuan Hou. 2015. Redundant heterogeneity and group performance. *Organization Science* 26, 1 (2015), 37–51.

[97] Statikk. 2016. Champion subclass list. Retrieved October 14, 2020 from https://web.archive.org/web/20160604025426/http://boards.na.leagueoflegends.com/en/c/developer-corner/3A5uuBw7-champion-subclass-list

[98] James Surowiecki. 2005. *The wisdom of crowds.* Anchor.

[99] John C Tang, Manuel Cebrian, Nicklaus A Giacobe, Hyun-Woo Kim, Taemie Kim, and Douglas "Beaker" Wickert. 2011. Reflecting on the DARPA red balloon challenge. *Commun. ACM* 54, 4 (2011), 78–85.

[100] Florenta Teodoridis, Michaël Bikard, and Keyvan Vakili. 2019. Creativity at the knowledge frontier: The impact of specialization in fast- and slow-paced domains. *Administrative Science Quarterly* 64, 4 (2019), 894–927. https://doi.org/10.1177/0001839218793384

[101] Jonathan K Trower and Kris K Moore. 1996. A study of the impact of individual goals and team composition variables on team performance. In *Proc. ACM SIGCPR/SIGMIS-CPR '96*. 206–213.

[102] Brian Uzzi, Satyam Mukherjee, Michael Stringer, and Ben Jones. 2013. Atypical combinations and scientific impact. *Science* 342, 6157 (2013), 468–472. https://doi.org/10.1126/science.1240474

[103] Maxime Véron, Olivier Marin, and Sébastien Monnet. 2014. Matchmaking in multi-Player on-line games: Studying user traces to improve the user experience. In *PROC. NOSSDAVW '14*. ACM, 7–12. https://doi.org/10.1145/2578260.2578265

[104] Herbert Wagner, Thomas Finkenzeller, Sabine Würth, and Serge P Von Duvillard. 2014. Individual and team performance in team-handball: A review. *Journal of Sports Science & Medicine* 13, 4 (2014), 808.

[105] Isaac Waller and Ashton Anderson. 2019. Generalists and specialists: Using community embeddings to quantify activity diversity in online platforms. In *The World Wide Web Conference*. 1954–1964.

[106] Pengfei Wang, Vareska Van De Vrande, and Justin g JP Jansen. 2017. Balancing exploration and exploitation in inventions: Quality of inventions and team composition. *Research Policy* 46, 10 (2017), 1836–1850.

[107] Qi Wang, Yi Yang, Zhengren Li, Na Liu, and Xiaohang Zhang. 2020. Research on the influence of balance patch on players' character preference. *Internet Research* 30, 2 (2020), 995–1018. https://doi.org/10.1108/INTR-04-2019-0148

[108] Warren E Watson, Tracey Minzenmayer, and Matt Bowler. 2006. Type A personality characteristics and the effect on individual and team academic performance. *Journal of Applied Social Psychology* 36, 5 (2006), 1110–1128.

[109] Dmitri Williams. 2010. The mapping principle, and a research framework for virtual worlds. *Communication Theory* 20, 4 (10 2010), 451–470. https://doi.org/10.1111/j.1468-2885.2010.01371.x

[110] Nick Yee. 2006. Motivations for play in online games. *Cyberpsychology & Behavior* 9, 6 (2006), 772–775.

[111] Mengxiao Zhu, Yun Huang, and Noshir S Contractor. 2013. Motivations for self-assembling into project teams. *Social Networks* 35, 2 (2013), 251–264.

Received October 2020; revised January 2021; accepted January 2021