Proceedings of the Fifteenth International AAAI Conference on Web and Social Media (ICWSM 2021)

Experience-Driven Peer Effects: Evidence from a Large Natural Experiment

William Cai, Johan Ugander
Stanford University
willcai@stanford.edu, jugander@stanford.edu

Abstract

Social interactions between people are a central mechanism by which behavior spreads. Several field studies have shown how observing peer behavior affects ones own behavior in a wide range of domains including health, information diffusion, advertisement, and education. However, the role of observing peer experience or outcome—the reward for the observed peer behavior—has largely gone uninvestigated. Here we examine evidence from a large-scale online setting, game play records from League of Legends match-ups, where we are able to disentangle the effects of observing peer behavior from observing peer experience. We find that in addition to positive peer effects from observing behavior, the effect is accentuated by observed experience, with a large positive effect when observing a good outcome and a small (but still positive) effect when observing a bad outcome. We further find that this experience-driven peer effect is moderated by time, becoming more pronounced when less time passes between observing the positive outcome and making the decision. Finally, we find mixed evidence of heterogeneity by skill, finding some settings where the experience-driven peer effect is stronger among high-skill users and others where it is stronger among low-skill users. Our findings demonstrate the importance of the role of observed peer experience beyond peer behavior, and elucidate important heterogeneities in experience-driven peer effects. We anticipate this result to be of use to both practitioners and theoreticians of social influence. For example, online platforms may wish to broadcast the positive outcomes of peers, more than mere behaviors, when a user performs a behavior desirable to the platform. Furthermore, efforts aimed at maximizing the social spread of a product may benefit from modeling experience-driven peer effects as part of the spreading process.

Introduction

A wide variety of human behaviors spread via peer to peer social interactions. These peer effects have been shown to cause people to exercise (Zhang et al. 2015; Aral and Nicolaides 2017), interact with advertisements (Bakshy et al. 2012a), learn (Hoxby 2000), and share information (Bakshy et al. 2012b). In addition, a separate vein of both empirical and theoretical work explores how network structure affects the manner by which behaviors spread, including ideas such as complex contagion where multiple peers must adopt for a behavior to spread (Centola and Macy 2007). Furthermore, extensive investigation has gone into how to seed interventions as to maximize the spread of a behavior (Kempe, Kleinberg, and Tardos 2003; Centola 2010).

An understanding of peer effects is especially relevant to the design of web platforms, as their design inherently dictates how peer effects unfold and information spreads, influencing how users behave on the platform. Furthermore, because platforms can control what peer behavior a user sees, they offer the ability for researchers to randomize peer observation, allowing for accurate measurements of peer effects in various settings. The detailed trace data which comes along with these online observations also signals an opportunity to understand in detail the mechanisms by which peer influence occurs. Analogous to Anderson, Kleinberg, and Mullainathan (2017), who use a large scale chess database to closely examine what factors lead humans to make errors in decision making, we can study what factors, including observed experience, attenuate or amplify peer influence.

Peer influence could plausibly affect many separate parts of a decision making process. Consider a decision maker who is trying to choose a behavior from a set of options and observes a peer perform a certain behavior. The observation may influence the decision maker in the following ways:

- Information-Driven: If a decision maker does not know a behavior is possible, observing a peer do that behavior will modify the set of choices that the decision maker considers.
- Behavior-Driven: Observing a peer choose a particular behavior over other behaviors may increase the agent's posterior estimate of the result of the behavior, as they may infer the peer believes it is superior to alternative behaviors.
- Experience-Driven: If experiences are observable by the agent, they may update their posterior belief about the result of an observed behavior with the observed experience.

Further, in the dichotomy of informational and direct-benefit effects (Easley and Kleinberg 2010), all three of these peer effects should be considered to be informational, in the sense that they all focus on social learning rather than peer externalities. These effects are also generally distinct from and should not be confused with normative, interpersonal concepts such as conformity (Asch 1956; Wheeler 1966).

In this work we do not study information-driven effects because in our setting, League of Legends, users choose from a fixed set of displayed choices. We instead focus on disentangling behavior-driven and experience-driven peer effects. Prior empirical work on social influence has not typically focused on the role the observed *outcome* of a behavior plays, instead only investigating whether the *behavior* was observed. In reality, whether a behavior spreads may strongly depend on the outcome of peers and the observation of those outcomes. For a farmer considering whether to buy weather insurance, peer effects are plausibly (and found to be) stronger when their peers have secured insurance payouts versus not (Cole, Stein, and Tobacman 2014). One can easily imagine analogues in other domains: a patient in a medical setting may undertake a treatment more readily (or only) if a peer has had a good outcome, or a web user may re-share a news story posted by a peer more readily (or only) if it receives many likes. These experience-driven peer effects are distinct from a more behavior-driven effect, where only the behavior but not the outcome is seen. The idea that good payouts should be influential on adopting a behavior is consistent with a Bayesian updating model (Banerjee 1992; Gallagher 2014) where individuals use observed experiences of their peers to learn the results of various behaviors and choose the behavior which maximizes expected utility, and is related to the concept of social comparison, where an individual uses a peer's performance to judge one's own hypothetical performance (Martin, Suls, and Wheeler 2002). In this work, we seek to quantify the relative strength of these effects in a highly instrumented setting.

Beyond empirical work, theoretical models of social influence often touch upon one of experience-driven or behavior-driven effects, but not the other. For instance, much of the literature on social herding (Banerjee 1992) constructs models based on the experience-driven mechanism, updating posteriors from observed experience, but ignores that observing peer behavior independently of payout may affect decision making. On the other hand, many cascade models consider the behavior-driven mechanism, where nodes are "activated" into behaviors depending on the number or proportion of neighbors who are "activated", regardless of observed experiences. The theoretical literature on influence maximization allows for some heterogeneity in the strength of the tie/influence (Aral and Dhillon 2018), but largely uses simplifications and does not consider experience-driven and behavior-driven effects separately.

A recent exception to the neglect of the role of observed experience is a line of work examining the question of whether observed payouts affect behavior adoption beyond observed behavior in weather insurance adoption patterns. In rainfall index insurance, farmers receive a payout on their insurance policy if rainfall at a measured station is outside an acceptable range for growing crops, insuring them against catastrophic weather. A field experiment with farmers in rural India finds that farmers are more likely to adopt if policies purchased in the previous year by fellow villagers had high returns (Cole, Stein, and Tobacman 2014), suggesting experience-driven peer effects at an aggregate, village level. A separate field experiment with Chinese farmers (Cai, De Janvry, and Sadoulet 2015) finds that observing peer farmers purchase index insurance without knowing previous payoffs has no effect on whether farmers adopt, although they do find spillover effects from information diffusion, suggesting that there is not much behavior-driven peer influence. Together, these studies tentatively suggest that the main mechanisms by which farmers are influenced by peers into adopting rainfall insurance are information and experience-driven, and that there may be experience-driven peer influence even in settings where there is no behavior-based peer influence, further emphasizing the importance of disentangling the two effects.

One heterogeneity in peer influence that researchers have thought about are the disparate impacts of observing different types of peers. For example, Granovetter makes the point that strong ties are individually stronger than weak ties, but the latter are more numerous, and thus more likely to provide job opportunities in aggregate (Granovetter 1977). Similarly, observing two peers who are not socially connected adopt a behavior may be less (or more) influential than observing two peers from the same social circle do a behavior (Ugander et al. 2012; Su et al. 2020). Many of these heterogeneities can be thought of driven by the position of alters in the network, but independently from that, the social status of the observed peer may also play a role in the strength of the peer effect (Paluck, Shepherd, and Aronow 2016). While interesting, we do not study the role these structural factors play in this work.

Understanding the role of observed experience in behavior contagion informs many decisions made by practitioners in the space. For example, if observing positive experiences is the central mechanism mediating peer effects, the decision of whether a social intervention will work may be strongly informed by whether the intervention comes with a positive experience, and further whether peers can easily observe those experiences. It may also inform influence maximization efforts: if both observed experience and behavior are important, jointly modeling both experience and behavior may result in better performance compared to a contagion model based on only one mechanism. Instead of solely optimizing based on the network structure, this could result in seeding which more heavily prioritizes individuals who are predicted to have highly visible positive experiences. In particular, strong experience effects suggest that influence maximization may be more useful in settings where the network is strongly homophilous in expected (positive) experiences of the intervention.

In this work, we identify a large-scale online setting where we can investigate in detail the roles that both behavior and observed experience play in behavior adoption. We carry out our investigation on the decision of champion (character) selection in League of Legends (LoL), the most popular video game in the world, with over 100 million monthly and 8 million daily active users (Volk 2016; Goslin 2019) in 2019. Online settings are widely used to study social influence due to their large size and detailed trace data. Furthermore, games such as Second Life (Bakshy, Karrer, and

Adamic 2009) have been used to study social influence in settings where users organically make decisions, and LoL itself has been used to study collective intelligence within teams (Kim et al. 2017). We applied and were granted access to the official LoL API, which we used to examine the match history of a month of games for 20,000 players, observing which champions each player and their counterpart on the opposing team chose, giving us the observed *behavior*. We further recorded whether their counterpart won or lost the game, giving us the observed *experience*. We then ask: how does observing behavior and experience affect which champions users select? To answer this question, we define a test statistic measuring how often players choose the champion they just observed, compute its empirical value, and generate its distribution under a range of null hypotheses where we vary whether there are experience-driven and behavior-driven peer effects. We find that there are both behavior-driven and experience-driven peer effects in LoL champion selection, where players are more likely to choose a champion if they observe it in their previous game. Further, we find that the effect is diminished (but still positive) when the champion is observed losing, and augmented when the champion is observed winning. Investigating heterogeneity in the experience-driven peer effect, we find moderate evidence that it is moderated by time, becoming stronger when the positive experience is more recent. These results are confirmed by two robustness checks we perform, where we vary both the set of the users we consider as well as the experience metric.

League of Legends

There are three stages to playing a game of LoL: Matchmaking, Champion Selection, and Multiplayer Online Battle Arena. First, players enter the matchmaking queue either by themself or with a single friend. Each player chooses a primary and secondary position out of the five total positions. From all the players in the queue, the matchmaking algorithm attempts to choose 10 players with similar Elo rating, a system for rating player skill (Elo 1978), and assigns the chosen 10 players to two teams of five (Isto 2013). It also assigns the players to the 5 unique positions per team based on their expressed preferences. We take advantage of the randomization offered by matchmaking: controlling for time and skill, users are randomly assigned to games, and thus, peers to observe. Further, LoL's matchmaking system highly values balancing the two teams in a game, preferring to "wait a little longer in the queue to get a fairer match" (Isto 2013).

In the player selection phase, each player simultaneously is given the option to choose one champion to ban. Then, the players take turns choosing characters out of the pool of 131 remaining champions, of which each champion may be chosen at most once. This champion selection decision is our main behavior of interest. In particular, we focus on the champion chosen by the user's opponent (the player on the opposing team with the same position as the user), and see if our user adopts that champion by selecting it in the following game. We also focus our analysis on games played in one out of the five positions, top lane, which spends the most time during the game fighting one-on-one with the opposing counterpart, due to differential champion popularity per position[1].

In the Multiplayer Online Battle Arena stage, all 10 players control the character they chose in a shared Battle Arena. Simply put, members of each of the two teams work with their teammates and attempt to destroy the other team's base while preventing the other team from destroying theirs. Games typically last for around 30 minutes. This is the phase where the player of focus observes the experience resulting from their opponent's choice. In particular, we use the simplest notion of observed experience: whether their opponent wins the game. Later, we perform analyses to confirm that our results are robust to the choice of position and measure of experience.

Research Questions

We are interested in whether there are peer effects for champion adoption: does observing an opponent pick a champion increase the probability that a player chooses that champion in their subsequent game? In other words, are there behavior-driven peer effects? Further, we are interested in the role of observed experience: if the player observes their opponent having a good experience with their champion selection, are they more likely to adopt? Here we measure experience by observing if their opponent won the game, but later we find that our results are robust to the chosen experience measure. In addition to studying observed experience, our main construct of interest, we also examine how it interacts with both the sophistication of the user and the recency of which it was observed. With respect to sophistication, we want to understand if increasing sophistication with a system interacts with the role of peer effects, which occurs for other decision-making heuristics such as loss aversion (Haigh and List 2005). Additionally, studying users of varying rank allows us to check for heterogeneous effects of observed experience by sophistication. The Bayesian learning interpretation of peer effects suggests this may be the case because more sophisticated users may have stronger priors on expected experience resulting from choices, and their posterior beliefs may not update as much from a single instance of data compared to a less sophisticated user. The effect of recency, or the time gap between the user observing the peer behavior and making a decision, on the peer effect is an important question for the design of interventions: is it important to show the peer behavior and experience right before the decision, or is it more beneficial to have time between?

Data

We examine the match histories for one month of play (September 2018) for the players ranked 1-10,000 and 100,000-110,000 on the ranked ladder for the North Amer-

[1] League of Legends teams have 5 members, each playing a different position: top lane, mid lane, bottom lane, support, and jungle. Each position corresponds to different parts of the map and different champions are popular for different positions.

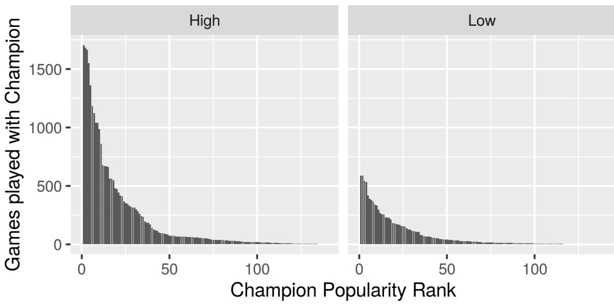

Figure 1: The number of games users played with each champion by skill level, with champions organized by popularity, showing the heterogeneity in champion popularity.

ican LoL server (at the start of the month)[2]. We use the LoL API to scrape their match history, obtaining records of 864,305 games by 16,629 players, excluding the $\sim$5% of games where the Riot API failed to record the position of the opponent. For each match, we observe the champion c the player chose, the champion o the opponent chose (again, focusing only on players and opponents in the "top lane" position), the champion s the player chose in their next game where they play the same position, k which denotes whether the player is among our high skilled (rank 1-10,000) or low skilled (100,000 - 110,000) sample, and the observed experience e_i (whether the opponent won or lost). Thus, each point of data is a tuple of form $(c_i, o_i, s_i, k_i, e_i)$. We call the set of all observed games $\mathcal{D} = \{(c_i, o_i, s_i, k_i, e_i)\}$. To avoid merely detecting when an opponent chooses the character that our player wanted to play, we filter our data to only contain games where the user picks first: that is, c is picked before o and thus there is exogenous variation in o among users who choose the same c. Table 1 and figure 1 summarize our filtered data, showing the number of scraped players, how many games they played in our month of data in the top lane where they chose first, and how often each champion was played[3].

Statistic	High Skill	Low Skill
Total Games	28817	10698
# Active Players	3705	2554
Mean GPP	7.78	4.19
SD GPP	10.84	6.39

Table 1: Summary statistics of scraped games where our scraped player chose first and played the "top lane" position, where GPP stands for games per player.

Analysis

After playing a game where they've observed the champion their opponents picked and the outcome of that game, we are interested in whether the players in the dataset choose, in their subsequent game, the same champion: in other words, whether $s_i = o_i$. To this end, we define a test statistic m, the proportion of games per player where $s_i = o_i$, averaged over all players, which will be central to the analysis:

$$m(\mathcal{D}) = \frac{1}{P}\sum_{p=1}^{P}\frac{1}{n_p}\sum_{i=1}^{n_p} 1_{s_{p,i}=o_{p,i}}.$$

Within the full data $\mathcal{D}$, we have P unique players and n_p games played by player p, where $s_{p,i}$ and $o_{p,i}$ are the champions that player p chose in their subsequent game and observed in their current game respectively. Informally we call m the *match proportion*.

Randomization Inference

In order to understand if we are truly observing any sort of peer effect in the observed data, we compare the observed value of m to its distribution under the null hypothesis that there are no effects of peer observation or observed experience on champion selection. To compute such a null distribution, we can take the observed data $\mathcal{D}$ and create a randomized dataset $\mathcal{D}'$. In $\mathcal{D}'$, we resample the champions and experiences that each player observes. That is, for each game $g_i = (c_i, o_i, s_i, k_i, e_i)$ in $\mathcal{D}$, we generate a corresponding tuple in $\mathcal{D}'$ as $g_i' = (c_i, o_i', s_i, k_i, e_i')$. Then, we compute the distribution of $m(\mathcal{D}')$ over many instances of $\mathcal{D}'$. In other words, we ask the question: what would the distribution of the test statistic be if there were no peer effects and observed experience did not affect behavior adoption?

The remaining technical question is how to sample o_i' and e_i'. First, we resample o_i' from the distribution of observed champions from games played at the same skill level in the same week. This design avoids selection effects caused by differences in the distribution of observed champions by skill or trending popularity of champions over the month of observations. Then, we resample e_i' from the empirical distribution of e_i in games played at the same skill level in the same week where the player observed champion o_i'. That is, we sample[4]

$$e_i' \sim \text{Bern}(\frac{\sum_{i=1}^{n} e_i 1_{o_i=o_i'}}{\sum_{i=1}^{n} 1_{o_i=o_i'}})$$

In addition to the *full null* hypothesis, described above, where neither the observed champion nor experience matters, we consider two weaker null hypotheses:

1. Behavior-driven null: The observed champion matters, but not the observed experience.
2. Change-driven null: The observed experience matters but not the observed champion.

The behavior-driven null corresponds to a purely *behavior-driven* peer effect, where peer effects are driven solely by

[2] Full analysis code available at https://github.com/wicai/experience_driven

[3] We note that champion winrates and popularity fluctuate over time and are loosely related.

[4] Sampling e_i' from a Bernoulli distribution parametrized by the mean observed experience can be thought of as a parametric bootstrap. We prefer this over shuffling because of the small number of games where rare champion are observed, where shuffling may not adequately express the true randomness of observed experience.

observing the behavior of peers (i.e., the users are influenced by what champions their peers chose, but not whether they observed the peers winning with them). The second hypothesis, the change-driven null, corresponds to a model where user behavior changes according to whether they win or lose, irrespective of which champion they observe. The purpose of this null is to disentangle two observed experience effects which are seemingly conjoined in this setting. Because one team wins and the other team loses in LoL, whenever a player observes the opposing champion win, that player must have lost the game. This latter null captures the possibility that players may simply get frustrated when they lose and pick a different champion than the one they played, but which champion they lose to does not affect their choice. We might expect m to be higher under losses under the change-driven null than under the full null hypothesis because champions are selected without replacement, so the champion the player observes cannot be the same as the one they were originally playing. Therefore, if players tend to change champions after a loss, even without a predilection for the champion they observed over other champions, they might still be more likely to choose the observed champion because it's not the one they were originally playing. Under this null, there is still a experience-driven peer effect – it's just that the impetus is to simply change one's behavior (thus the name *change-driven*), not to adopt the observed behavior.

To simulate m under these two hypotheses, we change how we resample o'_i and e'_i from the full null. Under the behavior-driven hypothesis, we simply let $o'_i = o_i$, as we want to preserve the original observed champion, but resample e_i in the same manner as before. On the other hand, for the change-driven hypothesis we resample o'_i as before, but do not resample e_i. In this manner, we preserve the original experience observed but modify the observed champion, in line with the hypothesis that the observed champion does not matter since the player is only being influenced by losing.

In addition to varying the null hypothesis under which we simulate the distribution of m, we also vary the set of games $\mathcal{D}$ we compute m over. For instance, we can separate games played by low skill and high skill players, $\mathcal{D}_l, \mathcal{D}_h \subset \mathcal{D}$, games where users observed a positive or negative experience (win or loss), $\mathcal{D}_1, \mathcal{D}_0 \subset \mathcal{D}$, or even $\mathcal{D}_c$ for games where a certain champion is observed. Importantly, the randomization is always done at the level of $\mathcal{D}$, the full dataset, and then $\mathcal{D}'$ is subsetted to match the subset of interest.

Consequentially, some nulls only affect the distribution of m over certain subsets. First, $m(\mathcal{D}_x) = m(\mathcal{D}'_x)$ under the behavior-driven null unless x includes a subset by observed experience. This is because under the behavior-driven null, $1_{s_{p,i}=o_{p,i}}$ remains the same for all games, so m only changes if the set of games it is computed over changes. Then, m will only vary if resampling the observed experience will change the games included in the subset of interest. For instance, whether a game goes into $\mathcal{D}'_1$ and $\mathcal{D}'_0$ depends on which experience is sampled for that game. Thus, we only consider the behavior-driven null when we subset the data by experience. Similarly, the change-driven null will not yield a different distribution of m compared to the full null unless we subset by experience: in both nulls we resample champions, and then as per above further resampling experience (as in the full null) only has an effect if we subset by experience.

Peer Effect

First we compute the test statistic m on the entirety of the dataset and its distribution under the null hypotheses. The observed $m(\mathcal{D}) = 0.0205$ and the 95% CI is $[0.01337, 0.01759]$ under the full and change-driven nulls (recall these yield the same distribution of m). We find that the observed m is higher than expected under these nulls, meaning that a user observing a peer play a champion increases the chance that in the following game the user chooses that champion.

Having confirmed that peer effects exist in this setting, we move on to disentangling behavior-driven and experience-driven peer effects. To this end, we compute m along with its null distributions for the data subsetted by experience (observed win vs observed loss), shown in figure 2. We find the observed m is higher than expected under both the change-driven and full null hypotheses, meaning that we observe peer influence for champions observed both winning and losing. This suggests that observed experience is not the only mechanism by which peer influence occurs in this setting, and that there is purely behavior-driven peer effect. However, we note that the difference is far larger for games where the user observes a win than in those where the user observes a loss: this difference is quantified in the "Observed Win vs Loss" comparison in table 2 and is statistically significant. We can further test this line of inquiry by comparing m to its distribution under the behavior-driven null hypothesis, where we only resample whether the user observed a win or a loss. The observed m for games where the user observed a champion win is higher than it is under the behavior-driven null and m for games with observed losses lower, meaning that the peer effect cannot solely be explained by which champion the user observed, thus confirming the existence of experience-driven peer effects, where users are more likely to adopt a champion if they observe a positive experience associated with it.

Heterogeneous Peer Effect: Skill Now, we take advantage of the scale of our data to investigate heterogeneities by skill and time in the experience-driven peer effect. First, we investigate player skill as a construct by itself, and compute m along with its null distributions for the data subsetted by skill (low skill vs. high skill), finding that $m(\mathcal{D}_h) = 0.02061$ with a null 95% CI of $[0.01264, 0.01737]$ under the full and change-driven nulls and $m(\mathcal{D}_l) = 0.02035$ with a null 95% CI of $[0.01277, 0.02064]$. Our results indicate that the observed m is higher than expected under the null for both skill levels, but only statistically significant at the $p < .05$ level for high skill players (although directionally the same for low skill players). The difference between m and its null mean is higher for high skill users, as shown in table 2 under the "High vs Low Skill" comparison, although the difference in differences is not statistically significant. Moving to the interaction between skill and experience, we compute m for each of the four conditions of (high skill, low skill)

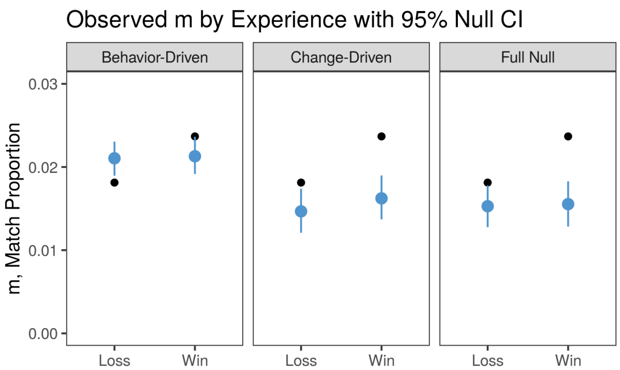

Figure 2: The match proportion m by observed experience, with 95% null confidence intervals in blue. The point estimates are above the CIs for the change-driven (middle) and full null (right) hypotheses, indicating that peer effects are present in all skill $\times$ experience conditions and that observed experience is not the only mechanism at work here, although the higher difference in the observed win conditions suggests that observing a positive experience increases the strength. This is further confirmed by examining the relation of the point estimates to the null CIs from the behavior driven hypothesis (left), which shows that the point estimate for m is higher under wins and lower under losses than it would be if observed experience did not affect m.

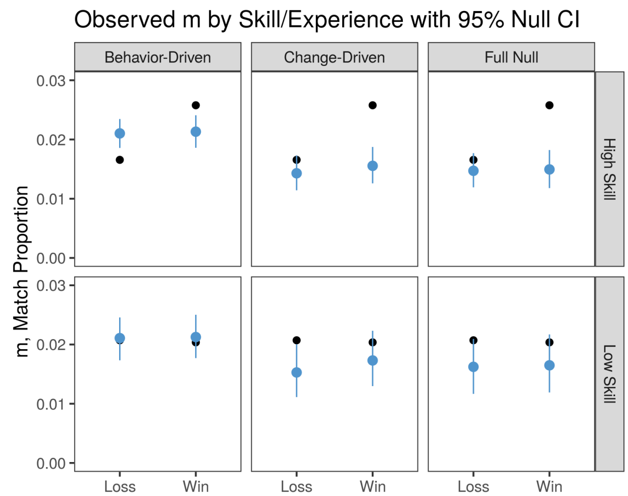

Figure 3: The match proportion m for each of the 4 skill $\times$ observed experience conditions, with 95% null confidence intervals in blue. First considering the change-driven and full null hypotheses, we find that directionally all observed m are above that expected under the null, although we lack the power to reach $p < .05$ for all interaction effects. We find that the match proportion m exceeds its expectations under the null the most when high skill players observe a win. We find that high skill players have higher m after observed wins and lower m after observed losses than expected under a behavior-driven null but low players do not, suggesting that in this setting experience-driven effects are mainly observed in sophisticated users.

$\times$ (observe opponent win, observe opponent loss). We compare these values with the distribution of m under all three null hypotheses. Figure 3 shows the results of this analysis. We first consider the change-driven and full nulls. We find that directionally, all point estimates for m are higher than their distributions under these null, although not all differences for games with observed losses are statistically significant at the $p < .05$ level. This suggests a certain baseline of behavior-driven peer effect is occurring at all skill levels for all observed experiences. Further, we indeed find significant evidence of heterogeneity. In fact, the entirety of the experience-driven effect in this setting is actually driven by high-skill users, where m is not significantly different for low skill users between losses and wins (again quantified in table 2). Finally, we can again examine the behavior-driven null to look for evidence of experience-driven peer influence. We confirm that experience-driven peer effects only exist in the high skill setting, where the observed m is higher in observed wins and lower in observed losses than expected under the null, whereas in the low skill setting it is squarely in the distribution expected under the null.

Heterogeneous Peer Effect: Time In addition to sophistication, we also investigate the effect of how much time has passed between the user observing an experience and making the decision of which champion to play in their next game, which we call the timegap. Again, we first consider the timegap by itself. We compute m along with its null distributions for the data subsetted by timegap (short ($<$ 1hr) vs. long ($>$ 1hr)), finding that after a long break $m(D_{\text{long}}) = 0.01961$ and the 95% null CI is $[0.0132, 0.01831]$ under the full and change-driven nulls. After a short break, $m(\mathcal{D}_{\text{short}}) = 0.02325$ and the 95% null CI is $[0.01278, 0.01786]$ under the full and change-driven nulls. Regardless of the length of break, we find that the observed m exceeds its expectation under the null, but we find that the difference is higher when the user only takes a short break: the difference is statistically significant at the $p < .05$ level and quantified in table 2 under Short vs Long Timegap. Moving on to the interaction between the timegap and the experience-driven peer effect, in figure 4 we show the observed m for each of the four conditions of (short timegap, long timegap) $\times$ (observe opponent win, observe opponent loss) along with its null distribution under our three null regimes. We find that the experience-driven peer effect is larger and only statistically significant at the $p < .05$ level when the user plays the next game within an hour of the previous one, although the difference in the experience-driven peer effect between the two timegap settings is not statistically significant. Those differences are quantified in table 2, under Observed Win vs Loss, (Long/Short) Timegap.

By Champion

In addition to aggregating the strength of peer effects over all champions, we can also do the analysis on individual champions. In particular, we are curious whether peer effects vary in some manner depending on the rarity of the

	Comparison	Delta	SE	P-value	Significance
1	High vs Low Skill	0.0020	0.00297	0.251	
2	Observed Win vs Loss	0.0053	0.00248	0.016	*
3	Observed Win vs Loss, High Skill Players	0.0090	0.00288	0.001	***
4	Observed Win vs Loss, Low Skill Players	-0.0006	0.00447	0.552	
5	Short vs Long Timegap	0.0042	0.00251	0.048	*
6	Observed Win vs Loss, Long Timegap	0.0031	0.00307	0.158	
7	Observged Win vs Loss, Short Timegap	0.0079	0.00317	0.006	**

Table 2: Comparisons of the difference (Delta) between the observed value of m and its mean under the full null distribution for a variety of scenarios, where the number of * corresponds to p $<$ .05, .01, .001. We find that the difference between m and its null expectation is not significantly different for high skill vs low skill users (row 1), but is significantly higher in games where the user observes a win rather than a loss (row 2), suggesting experience-driven peer effects. Further, we find that this difference is mainly driven by high skill users (rows 3, 4). Considering heterogeneity by time, we find that m exceeds its null more when the user chooses their next champion within an hour of observing the experience (short timegap) than waiting more than an hour (row 5). Furthermore, we find that the experience-driven peer effect is only significant in the short timegap regime (rows 6, 7), although directionally the same in both timegap settings.

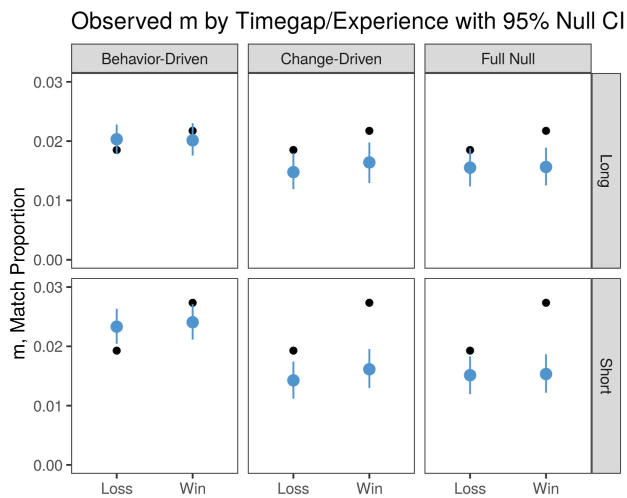

Figure 4: The match proportion m by whether the user took a long break (>1 hr) before the next game or not, and whether they observed a loss or a win, along with its null distribution under each of our three nulls. We find that under the change-driven and full nulls, the amount by which m exceeds its expectation under the null is higher when the user does not take a long break after observing the positive experience, suggesting that the effect of observing the champion win diminishes in influence over time. Examining the behavior-driven null, we find that the observed m is higher than the null after observed wins and lower after observed losses, suggesting experience-driven effects in both timegap settings, but m is further from its expectation under the behavior-driven null when the user does not take a long break, suggesting that experience-driven peer effects are stronger in that regime.

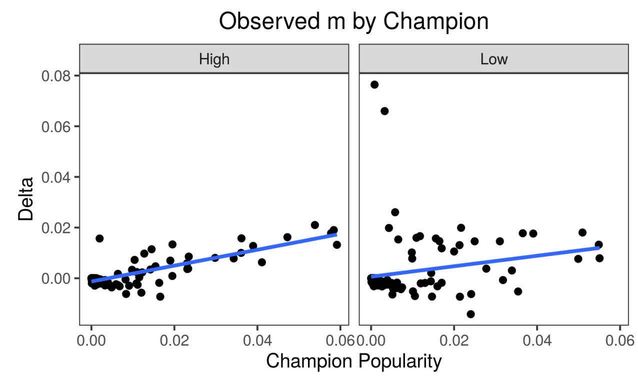

Figure 5: Difference between observed m and its mean (labeled Delta) under the full null, by champion popularity. We find a positive correlation between champion popularity and Delta, suggesting that the most popular champions benefit the most from the peer effect. The outliers and steadily decreasing points for popularity range 0.00 to 0.02 are artifacts resulting from unpopular champions being observed infrequently and thus having small denominators in inference: for example, one outlier has an observed m of $\frac{1}{13}$. Correspondingly, many observed m are 0. However, under the null, champions have monotonically increasing null m (and thus, decreasing Delta) as popularity increases. Thus, below popularity 0.02, many champions have empirical m of 0 but a steadily decreasing null m, corresponding to steadily decreasing points. Inferences on individual unpopular champions have low power but are not heavily weighted in our aggregate analyses since unpopular champions are rarely observed.

champion. In figure 5 we plot the difference between the observed m and its mean under the full null distribution for all champions. We find that in general, more popular champions have larger boosts in match proportion compared to less popular champions - that is, the difference between the observed m and the m under the full null for games where a certain champion was observed was higher for already popular champions. Taking the point of view of a random utility model, it could be that these champions were already close to the threshold for being chosen. As popular champions, we might expect many people to already have high utilities for them, which were pushed beyond the choice threshold by the peer observation. This phenomenon suggests a "rich get richer" model for the role of peer effects in this setting, where champions who are already more popular and thus observed more also get a larger boost from peer effect per observation because there are more users on the margin for being converted.

Robustness

Here we assess the robustness of our analysis to two choices we made earlier: (1) using observed winning/losing as the chosen metric of experience, and (2) limiting our analysis the position of "top lane", the position which spends the most time alone fighting their counterpart, out of the 5 positions in League of Legends. Varying both the experience and position considered, we find that our results largely still hold, suggesting greater generalizability of the experience-driven peer effect.

Another Experience Metric

Whether the user observed a win or a lose is a very intuitive measure of experience, but also depends on the actions taken by the opposing team as a whole instead of just the user's counterpart. Here we consider a measure of experience which is more tied to how the user performed relative to their peer counterpart. In particular, we investigate the Kills/Deaths/Assists (KDA) ratio, given by $\frac{\text{kills}+\text{assists}}{\text{deaths}}$, a measure of how many takedowns (kills + assists) of enemy champions a player helped with compared to how many times they were taken down by the enemy team (deaths). We repeat the parts of our main analysis related to observed experience with respect to this new experience metric. Our results are quantified in table 3. We find that there is indeed an experience-driven peer effect, where the difference between the observed m and its expectation under the null is higher when a user observes a peer get a higher KDA than when observing a peer get a lower KDA (row 1). We confirm our results regarding heterogeneity in skill, again finding that the experience-driven peer effect is largely driven by high skill players (rows 2, 3). Finally, we also confirm our results on heterogeneity by the time between games: m exceeds its null by a significant amount only when subsetting to observations with a short timegap, although the difference is directionally the same for both timegaps (rows 4, 5).

Another Position

In addition to varying the considered metric, we also vary the subset of users we consider by looking at users who play the "mid lane" position. Repeating our analysis (quantified in table 3), we again find evidence of our main experience-driven effect, where m exceeds its null expectation more when observing a win than observing a loss (row 7). We again search for heterogeneity by skill and timegap, finding that for the overall peer effect, neither is a strong moderator (rows 6, 10). The finding for skill echoes that of our main analysis, but the finding for the timegap suggests that it is not as strong a factor as we found in our main analysis. Finally, examining heterogeneity of the experience-driven effect, we find that the effect is more heavily driven by low skill users in this setting (rows 8, 9), suggesting weaker conclusions regarding user sophistication than we might have drawn from our main analysis. On the other hand, we find that the experience-driven peer effect is still stronger following a short timegap (rows 11, 12), confirming our previous result.

Conclusion

In this work we investigate in-depth the roles that observed experience, user skill, and the timegap play in peer effects in the choice of champion selection in League of Legends. We find positive evidence of peer effects in this setting, and that the peer effect is stronger when observing a win than a loss, although still positive in both instances. We rule out the possibility that the effects are caused by losing the game versus observing a win by comparing the observed test statistic to that of the change-driven null, instead attributing them to observed experience. We find evidence the average strength of the overall peer effect remains roughly the same between users of varying skill level, and that it may be stronger if there is only a short duration between observing the experience and choosing a champion (the overall effect is larger in the top lane but not mid lane). We further find weak evidence for heterogeneity of the experience-driven effect by skill, where for top lane the experience-driven peer effect was stronger in high skill users but in mid lane stronger in low skill users. Finally, we find that the experience-driven effect is higher when the user takes only a short (< 1 hr) time between observing the experience and choosing a champion, finding directionally the same result in both our main analysis and our two robustness checks.

Returning to the breakdown of the possible effects observing a peer could have, the information, behavior, and experience-driven hypotheses, we find strong evidence of experience-driven peer effects which are responsible for a large portion of the overall peer effect. The remainder could be attributed to either the information or behavior-driven hypotheses, but the data leans towards the behavior-driven hypothesis. Evidence of this comes from observing the peer effect by the rarity of champion, where we find that for rarely played champion there are weaker peer effects than for more popular champions. Under the information-driven hypothesis, we would expect to see higher peer effects for rare champions, as those are called into attention less frequently. The finding that observing a popular behavior is more likely to marginally tip an individual into performing the behavior than observing an unpopular behavior, combined with the tautology that more popular behaviors are observed more, suggest a "rich get richer" model of peer effects, where peer influence mainly serves to make already-popular behaviors get more popular.

We believe that this work informs efforts in influence maximization and social interventions. The importance of observing positive experiences suggests efforts to maximize influence of interventions might be well served by seeding users who will have strongly visible positive experiences, especially ones connected to other such individuals. For example, a public health intervention might be best seeded in a

	Comparison	Delta	SE	P-value	Significance
1	Observed Higher vs Lower KDA	0.0049	0.00181	0.003	**
2	Observed Higher vs Lower KDA, High Skill Players	0.0093	0.00218	0.000	***
3	Observed Higher vs Lower KDA, Low Skill Players	-0.0022	0.00331	0.748	
4	Observed Higher vs Lower KDA, Long Timegap	0.0021	0.00314	0.247	
5	Observed Higher vs Lower KDA, Short Timegap	0.0095	0.00324	0.002	**
6	High vs Low Skill, Mid Lane	-0.0020	0.00266	0.771	
7	Observed Win vs Loss, Mid Lane	0.0049	0.00217	0.011	*
8	Observed Win vs Loss, High Skill Players, Mid Lane	0.0039	0.00239	0.052	
9	Observed Win vs Loss, Low Skill Players, Mid Lane	0.0067	0.00359	0.031	*
10	Short vs Long Timegap, Mid Lane	0.0009	0.00217	0.341	
11	Observed Win vs Loss, Long Timegap, Mid Lane	0.0022	0.00275	0.210	
12	Observed Win vs Loss, Short Timegap, Mid Lane	0.0090	0.00281	0.001	***

Table 3: Comparisons in our two robustness checks, where we consider a different experience (observing a higher KDA) and then a different position in the game (mid lane). We find our main effect of interest holds, where the increase in m over the null (Delta) is significantly higher when observing a positive experience than a negative experience (rows 1, 7), meaning that we still detect experience-driven peer effects even if we vary the measure of experience or the position considered. Revisiting heterogeneity, we again find no evidence of difference in the overall peer effect by skill (row 6) in the mid lane, and contrary to the top lane we find no evidence of the timegap mattering for the overall peer effect (row 10). Examining the heterogeneity of the experience-driven peer effect, we find that although the experience-driven effects in our main analysis were driven by high skill users, which is echoed in our KDA robustness check (rows 2, 3), for the mid lane analysis the effect is actually greater in the lower skill players (rows 8, 9). Finally, examining heterogeneity of the experience-driven effect by timegap, we find the effect is mainly driven by games with a short timegap (rows 4, 5, 11, 12), which is directionally the same as our main analysis.

cluster of individuals who will likely exhibit positive health outcomes to their peers. Paradoxically, this may also suggest that beginning an intervention by seeding in a vulnerable population may actually inhibit the spread of the intervention if some of the seed users exhibit negative experiences to others, even if on average the intervention is more helpful to those in the vulnerable population than others. Of course, experience-driven peer effects are not the only factor which should determine seeding. There can be other good reasons to begin an intervention within a vulnerable population, and they may be important enough to override the concern of observing negative peer effects. Furthermore, observed experiences might belie the actual usefulness of the intervention: if the intervention was in fact useless but the seed individuals were going to be healthy no matter what, their peers might mistakenly believe that the intervention would help them, leading to bad outcomes. This suggests that seeding an intervention in users who will have strongly visible positive experiences should only be done when a policymaker has strong prior knowledge that the intervention will be helpful to those who might observe the intervention and see users' positive experiences with it.

Finally, understanding the role observed experience plays allows us to understand when social interventions may or may not be expected to work. If experiences are not easily observed (for example, in single-instance decisions like whether to apply for college, where most social ties are between individuals who are deciding at the same time), social interventions may be less successful because they can only appeal to behavior-driven effects, which may be weaker than experience-driven effects. We hope that this work inspires further work in identifying what factors, including status or foreignness of the behavior, correspond to individuals adopting peer behavior, and how they interact with observed experience.

References

Anderson, A.; Kleinberg, J.; and Mullainathan, S. 2017. Assessing human error against a benchmark of perfection. *ACM Transactions on Knowledge Discovery from Data (TKDD)* 11(4): 45.

Aral, S.; and Dhillon, P. S. 2018. Social influence maximization under empirical influence models. *Nature Human Behaviour* 2(6): 375.

Aral, S.; and Nicolaides, C. 2017. Exercise contagion in a global social network. *Nature Communications* 8: 14753.

Asch, S. E. 1956. Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychological Monographs: General and Applied* 70(9): 1.

Bakshy, E.; Eckles, D.; Yan, R.; and Rosenn, I. 2012a. Social influence in social advertising: evidence from field experiments. In *Proceedings of the 13th ACM Conference on Electronic Commerce*, 146–161. ACM.

Bakshy, E.; Karrer, B.; and Adamic, L. A. 2009. Social influence and the diffusion of user-created content. In *Proceedings of the 10th ACM conference on Electronic commerce*, 325–334. ACM.

Bakshy, E.; Rosenn, I.; Marlow, C.; and Adamic, L. 2012b. The role of social networks in information diffusion. In *Proceedings of the 21st International Conference on World Wide Web*, 519–528. ACM.

Banerjee, A. V. 1992. A simple model of herd behavior. *The Quarterly Journal of Economics* 107(3): 797–817.

Cai, J.; De Janvry, A.; and Sadoulet, E. 2015. Social networks and the decision to insure. *American Economic Journal: Applied Economics* 7(2): 81–108.

Centola, D. 2010. The spread of behavior in an online social network experiment. *Science* 329(5996): 1194–1197.

Centola, D.; and Macy, M. 2007. Complex contagions and the weakness of long ties. *American Journal of Sociology* 113(3): 702–734.

Cole, S.; Stein, D.; and Tobacman, J. 2014. Dynamics of demand for index insurance: Evidence from a long-run field experiment. *American Economic Review* 104(5): 284–90.

Easley, D.; and Kleinberg, J. 2010. *Networks, crowds, and markets*, volume 8. Cambridge University Press.

Elo, A. E. 1978. *The rating of chessplayers, past and present*. Arco Pub.

Gallagher, J. 2014. Learning about an infrequent event: evidence from flood insurance take-up in the United States. *American Economic Journal: Applied Economics* 206–233.

Goslin, A. 2019. League of Legends has nearly 8 million peak daily concurrent players globally, says Riot. https://www.riftherald.com/2019/9/17/20870382/league-of-legends-player-numbers-active-peak-concurrent. [Online; accessed 8-October-2019].

Granovetter, M. S. 1977. The strength of weak ties. In *Social networks*, 347–367. Elsevier.

Haigh, M. S.; and List, J. A. 2005. Do professional traders exhibit myopic loss aversion? An experimental analysis. *The Journal of Finance* 60(1): 523–534.

Hoxby, C. 2000. Peer effects in the classroom: Learning from gender and race variation. Technical report, National Bureau of Economic Research.

Isto. 2013. Matchmaking Guide. https://support.riotgames.com/hc/en-us/articles/201752954-Matchmaking-Guide. [Online; accessed 8-October-2019].

Kempe, D.; Kleinberg, J.; and Tardos, É. 2003. Maximizing the spread of influence through a social network. In *Proceedings of the ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 137–146. ACM.

Kim, Y. J.; Engel, D.; Woolley, A. W.; Lin, J. Y.-T.; McArthur, N.; and Malone, T. W. 2017. What makes a strong team?: Using collective intelligence to predict team performance in League of Legends. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, 2316–2329. ACM.

Martin, R.; Suls, J.; and Wheeler, L. 2002. Ability evaluation by proxy: Role of maximal performance and related attributes in social comparison. *Journal of Personality and Social Psychology* 82(5): 781.

Paluck, E. L.; Shepherd, H.; and Aronow, P. M. 2016. Changing climates of conflict: A social network experiment in 56 schools. *Proceedings of the National Academy of Sciences* 113(3): 566–571.

Su, J.; Kamath, K.; Sharma, A.; Ugander, J.; and Goel, S. 2020. An Experimental Study of Structural Diversity in Social Networks. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, 661–670.

Ugander, J.; Backstrom, L.; Marlow, C.; and Kleinberg, J. 2012. Structural diversity in social contagion. *Proceedings of the National Academy of Sciences* 109(16): 5962–5966.

Volk, P. 2016. League of Legends now boasts over 100 million monthly active players worldwide. https://www.riftherald.com/2016/9/13/12865314/monthly-lol-players-2016-active-worldwide. [Online; accessed 8-October-2019].

Wheeler, L. 1966. Toward a theory of behavioral contagion. *Psychological Review* 73(2): 179.

Zhang, J.; Brackbill, D.; Yang, S.; and Centola, D. 2015. Efficacy and causal mechanism of an online social media intervention to increase physical activity: results of a randomized controlled trial. *Preventive Medicine Reports* 2: 651–657.