

# Bad Apples Spoil the Fun: Quantifying Cheating Influence in Online Gaming

Xiang Zuo<sup>1</sup>, Clayton Gandy<sup>1</sup>, John Skvoretz<sup>2</sup>, Adriana Iamnitchi<sup>1</sup>

<sup>1</sup>Computer Science & Engineering, University of South Florida  
{xiangzuo, cgandy, anda}@mail.usf.edu

<sup>2</sup>Department of Sociology, University of South Florida  
jskvoretz@usf.edu

## Abstract

Understanding and quantifying the factors that lead to cheating in society is problematic, due to the inherent desire to hide socially unacceptable actions. While significant progress has been made in the understanding of unethical behavior via in-lab experiments, little was measured at scale, in the wild. By analyzing cheating in online games, this paper verifies at scale and in the wild a number of previous observations drew from controlled, in-lab experiments. We verify empirically that cheating behavior is contagious, and identify some of the factors that encourage cheating and some that limit it.

## Introduction

Unethical behavior raises significant issues in society, it is prevalent (Ariely and Jones 2012), yet it is difficult to quantify and accurately model. This difficulty is due to many factors, ranging from the difficulty to identify it to the difficulty to classify it: is taking a pen from the office more unethical than lying in a political campaign? However, understanding unethical behavior and identifying the factors that trigger or prevent it in various circumstances can have significant impact in many areas of life, from education to sports.

Cheating, as one type of unethical behavior, is defined as the act of breaking the rules in order to gain unfair advantage in a situation. Studying cheating online, where all interactions leave a digital mark, becomes a promising approach for understanding this behavior in real life. Gaming market was worth around \$91.5 billion in 2015 (Sinclair 2015). Gaming popularity transcends geography, age and gender: in 2011, the average age of American gamers was 37, 72% of households played games, and female players accounted for 42% of them (Economist 2011).

Understanding cheating in online games is in itself of interest for the gaming industry, but several other characteristics make it an ideal phenomenon to study. First, ties in this social network are supported by real (gaming) interactions (Blackburn and Iamnitchi 2013), which differentiates it from declared social networks like Facebook (Wilson et al. 2009). This translates into slightly different, and, we believe, more realistic social network structures.

Second, in-game behavior closely mirrors real-world social behavior (Zell and Thurner 2010). Studying a gaming

network is particularly interesting because of the *competitive* nature of many multiplayer games, a feature that has parallels in the offline world, for instance, corruption in cases such as Enron, where “internal [group] competition could set the stage for the diffusion of ‘widespread unethical behavior’ ” (Kulik, O’Fallon, and Salimath 2008). Under this assumption, results from real life experiments should be valid in this environment and vice-versa.

Third, cheating in online games is widespread (Pritchard 2000), which makes it easier to observe and measure over limited intervals of times.

In this work we collect and analyze a large-scale dataset that includes a significant number of instances of cheating behavior from the Steam community<sup>1</sup>, the world’s most popular online gaming platform. Drawing from the sociology and psychology literature, we tested several factors that motivate cheating but have remained untested outside of controlled laboratory experiments and small, survey-based studies.

Our empirical observations lead to the following contributions:

- We quantitatively characterize the spreading of cheating behavior in an interaction-based network of gamers. While we have shown cheating as a social contagion in a friendship network (Blackburn et al. 2014; Blackburn 2014), in this work we observe the contagion process by looking at the interaction network, a more likely platform for this contagion to happen.
- We verify that players’ cheating engagement can be significantly increased with exposure to neighbors who cheat and have not been punished.
- We confirm that in-group social contacts have larger influence in inspiring cheating than do outsiders.
- We measure the influence that the penalty for cheating has on adopting this behavior.
- We compare the adoption of cheating behavior across different classes of players.

Our findings lead to a better understanding of what are the factors that affect people’s engagement in unethical behavior, and pave the ways in which cheating detection/prevention mechanisms can be improved in real life.

<sup>1</sup><http://steamcommunity.com/>

## Methodology

We collected data about online gamers from different sources and processed it to cover for some missing information, as described below.

### Datasets

We collected data from two different sources: the Steam platform where users buy games and maintain Steam Community profiles; and a third-party service that aggregates information about real-time playing status on many gaming servers around the world that host 26 highly popular games. Among these 26 games, many are notorious for having been infested with cheating: e.g., Counter-Strike:Global Offensive, Team Fortress 2, and Counter-Strike 1.6.

**Steam and Steam Community:** Steam is an online platform developed by Valve (Valve 2015) that provides a system for players to buy, install, and play games. It also provides an online social network, Steam Community, where players can create their profiles, befriend other players, join groups and chat with in-game partners. Each player has a Steam account but is not required to have a Steam Community profile (the design of the system itself). A Steam Community profile includes a nickname, groups the player joined, gameplay status for the past two weeks (achievement, badges, etc.), friend list, list of games owned, profile setting (private or public), and a cheating flag.

The cheating flag, called a “VAC-ban” after the Valve Anti-Cheat (VAC) service that detects players who cheat in games and marks their profiles, is permanent and publicly visible regardless of the profile’s privacy setting. More than 1.5 million Steam accounts have been banned by 2014 (Crookes 2014). A VAC-ban is associated with the game in which the player was caught cheating, but that association is not publicly visible. A VAC-banned player cannot play that game on “VAC-secured” servers, and most game servers are VAC-secured. The player can, however, play any other games on any appropriate server. The details of how VAC works are not made public to defend against the prosperous and active cheating industry (Maiberg 2014). What is known, however, is that VAC bans are not issued immediately upon cheat detection, but in delayed waves. More detailed information about cheating and anti-cheat in video games can be found in (Aboukhadijeh 2011).

**GameMe:** The contagion of cheating was shown in previous work on the Steam Community (Blackburn et al. 2014; Blackburn 2014), but it is clear that the social network is not the platform that induces cheating. Based on forum discussions (Huizinga 1950), the cheating flag is not seen as a badge of honor. Consequently, seeing it on a friend’s profile is not likely to inspire a player to cheat. The contagion, we believe, happens during play time, as retaliation when noticing that an opponent is cheating, or inspired by a teammate’s cheating-enabled performance.

That is why in our work we chose to look at records of in-game interactions. However, this is not trivial, as gamers play on various servers distributed around the world and owned by different individuals or groups. Various services

connect to these servers and report their status in real time: what game they host, who are the users engaged in playing, what the score is, etc. We collected playing information from one such service, GameMe<sup>2</sup>, whose main purpose is to provide statistics for tracking players’ in-game performance. These statistics include players’ in-game interactions, such as on which teams they are playing or with whom they are playing, the global ranking for each game, map performance overview, etc. For Steam players, the GameMe service connects their GameMe statistics with their Steam ID, which enables us to connect the two datasets.

**Data Collection:** For 32 days between March 20, 2015 and April 22, 2015, we recorded co-presence (co-match) information every 30 minutes for all 26 games supported by GameMe. Specifically, we collected evidence about players playing on 1,283 game servers located in 31 countries, where each server can support multiple matches simultaneously. As many of these games are built around (typically two) teams playing against each other, we also collected team information. Our choice of 30-minute periodic crawl was informed by the observation that most games can be finished within 30 minutes.

Some of the Steam users tracked by GameMe during our observation period did not have Steam profiles created, which prevented us from gathering more information about them, such as whether their accounts are VAC-banned. We discarded these players from our dataset. For the Steam Community users recorded by GameMe during this time we collected additional profile data using the Steam API. The data associated with each player thus contains a Steam ID, the player’s friend list, a timestamp for each friendship formation, a flag (VAC ban) that indicates whether the corresponding user has been detected cheating, and the time when the cheating flag was applied.

### Networks

Using the data we extracted from GameMe and Steam Community, we created two undirected networks, friendship and co-match. Table 1 gives the exact numbers.

- The *friendship* network is composed of edges between players who are declared friends on Steam Community. Each edge has a timestamp to indicate when the relationship was formed.
- The *co-match* network is a dynamic network where an edge exists between two players at time  $t$  if they played in the same match. Each edge has a timestamp to indicate when the match happened.

Note that players are not restricted to Steam IDs when they play games and they are allowed to use other accounts for play. However, we are only interested in Steam users, thus we only collected Steam users’ co-presence. We collected all monitored 26 games’ servers and Steam users in these games account for approximately 29.89% of the total players monitored by GameMe. The majority of Steam players reside in four popular games (Counter-Strike: Global

<sup>2</sup><http://www.gameme.com>

Offensive, Team Fortress 2, Counter-Strike 1.6 and Counter-Strike: Source) that are notorious for cheating (Dransfield 2014; Maiberg 2014; Fortress2 2013). In these four games, 91.2% of players use their Steam accounts to play. Thus, the co-match network is able to capture Steam players' in-game activities without excessive interference from non-Steam players.

In the co-match network, two players could have no declared friendship at the time of their match. Additionally, in most scenarios players chose game servers randomly. The servers allocate login allowances by considering users' locations, current bandwidth, number of players that are connected to server, etc., though some players may reserve the server purposefully. Therefore, during this short time period, most of the players only co-matched each other for a limited number of times. During our one-month observation time, 91.5% of the pairs of users only co-matched once.

### The Influence Timing Condition

Valve does not post the VAC-ban on a cheater's profile immediately after observing cheating, but with varying delays of days or even weeks (Valve 2015). To understand who cheats first and thus who influences and who is influenced, we need to know the time when cheating or communication about cheating occurred. To compensate for this missing piece of information, we estimate the time of cheating using a parameter  $w$  that limits the difference in the delays with which the cheating label was applied to any two gamers. We call it the influencing time.

Specifically, assuming player  $A$  cheated at time  $C_A$  and is labeled at time  $T_A$  ( $T_A > C_A$ , visible in our records as posted on the Steam Community profile) and player  $B$  was VAC-banned at time  $T_B$ , we consider  $B$  being potentially influenced (to cheat or to refrain from cheating) by  $A$ 's cheating if and only if:  $T_B < T_A \pm w$ . This condition provisions for the situation in which despite the fact that  $A$  cheated first, the VAC-ban on  $A$ 's profile was posted after the VAC-ban on  $B$ 's profile (thus,  $T_B < T_A$ ). Since the delays with which the labels are applied are limited to a difference  $w$ , then  $T_A > T_B - w$ .

### Cheating as a Complex Contagion

Cheating as a contagious process has been evidenced in real life. For example, research into academic cheating indicates the presence of a network effect: the acceptance of a single high school cheater into a United States military service academy has been shown to cause a statistically significant 0.37 to 0.47 additional students to cheat (Carrel, Malmstrom, and West 2008). A study of 158 private universities (Rettinger and Kramer 2009) shows that observing other undergraduate students cheat is strongly correlated with one's own cheating behavior.

Previous work (Blackburn et al. 2014; Blackburn 2014) shows cheating is a contagion process as evidenced by the friendship relations in Steam Community. The phenomenon was observed by hazard analysis with two datasets collected from Steam Community's friendship network in spring 2011 and August 2012. Here, we verify cheating in online gaming

environment is a contagion with a more accurate dataset that includes newly released VAC-ban labeling times. Moreover, we confirm this phenomenon on two networks that connect the same players. More importantly, this is the first time to study the cheating contagion in the co-match network, a likely more precise indication of how cheating spreads.

**Method:** We analyze the relationship between a user's adoption probability of cheating behavior and the number of the user's cheater neighbors. We follow the methodology in (Cosley et al. 2010; Romero, Meeder, and Kleinberg 2011; Hodas and Lerman 2014). We consider user  $A$  has one exposure to cheating every time one of its neighbors is VAC-banned. The probability of adopting the cheating behavior after  $k$  exposures,  $P(k)$ , is defined as the fraction of users who are labeled as cheaters after exactly  $k$  exposures to cheating. Specifically,  $P(k) = \frac{Adoption(k)}{Exposure(k)}$ , where  $Exposure(k)$  is the number of users who have exactly  $k$  cheaters in their neighborhood, and  $Adoption(k)$  is the number of users  $k$ -exposed to cheating who become cheaters before the  $(k + 1)$ -th exposure.

When we consider exposures in the friendship network, we assume that all of a player's neighbors observe when the player is VAC-banned. In reality, the change of status is not broadcasted, and it can only be noticed by looking at the Steam Community profile or when playing together in particular games. Thus, without timing information and no proof of observation or interaction, the exposure function ( $Exposure(k)$ ) simply counts the number of cheater neighbors of each player. The co-match network, on the other hand, records richer time-based interaction information among players. We use these times interactions as the vehicle for contagion and apply the time windows as explained above to estimate who influences whom.

**Results:** Figure 1 shows behavior adoption probability ( $P(k)$ ) as a function of  $k$  cheater neighbors in the two networks. We observe two key features from the  $P(k)$  curve shapes. First, both networks present an increased probability of cheating adoption as the number of cheater neighbors increases. These results suggest a complex contagion (Centola and Macy 2007): as unaccepted social behavior, cheating needs social affirmation from multiple sources and presents an increased likelihood of adoption with each additional exposure. That is, a player is more likely to engage in cheating if two of his neighbors cheat than when one of his neighbors cheats.

Second, the co-match network shows a different contagion process compared to the friendship network. Contagion arrives at its peak values for two exposures to cheating, and then decays. This shows an over-exposure trend—increases in exposure dramatically suppress contagion (Hodas and Lerman 2014). Similar over-exposure phenomena were also observed in the information diffusion process of other online social networks such as Twitter, Digg (Hodas and Lerman 2014) and Flickr (Cha, Mislove, and Gummadi 2009).

Table 1: Characteristics of Friendship and Co-Match networks used in the following experiments. CC: Clustering Coefficient, OT: Observation Time.

Network	Nodes	Edges	# Cheaters	Density	# Connected Component	CC	OT (days)
Friendship	3,148,289	44,725,277	223,527 (7.1%)	9.0247-06	1,685	0.1809	2,511
Co-Match	167,432	1,130,595	2,359 (1.4%)	8.0661e-05	2,879	0.5498	32

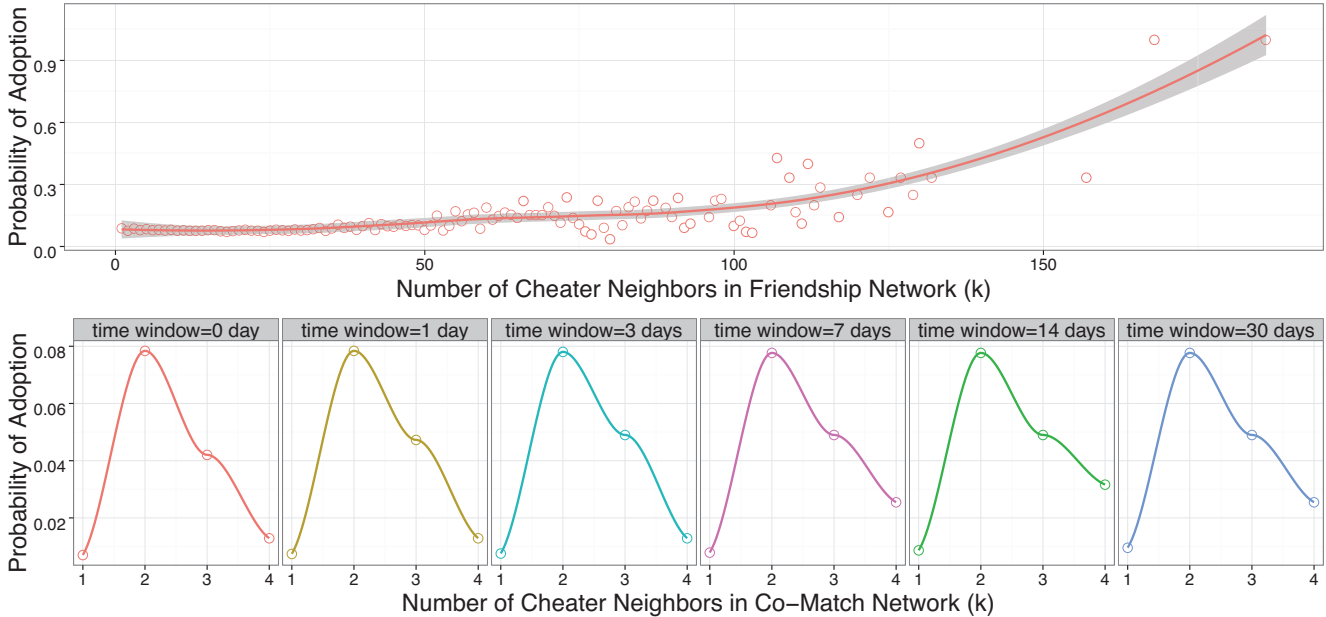


Figure 1: The exposure response function of players who adopted the cheating behavior after  $k$ -exposure to cheater neighbors with 6 time windows in the friendship and co-match network respectively. The adoption probability is the fraction of users who adopt cheating behavior directly after their  $k^{th}$  observation of it. The shaded area represents the 95% confidence interval.

## Factors That Influence Cheating Engagement

According to the results just presented, cheating in Steam Community spreads as a complex contagion and is associated with the effect of exposure on individuals. In this section we investigate under what conditions, given the exposure to cheating, a player tends to cheat or tends to refrain from cheating. To this end, we rely on results from sociology and psychology to formulate a number of hypotheses that we test empirically, using the co-match network. Our dataset includes all crawled 26 games thus the analysis results can generalize to other online games.

### Factor I: Observing Unpunished Cheaters Aggravates Cheating

**Hypothesis:** Observing unpunished cheaters in action increases the likelihood of cheating.

This hypothesis is supported by Gino et al. (Gino, Ayal, and Ariely 2009) who showed with in-lab, controlled experiments with human subjects that after observing people who cheated and were not punished, the subjects were more likely to cheat. The controlled experiment involved 141 subjects who were asked to solve 20 matrix problems in 5 minutes and were invited to take money as payment, accord-

ing to how many problems they (claimed to have) solved, at the rate of \$0.5 for each correct solution. The whole process was self-reported and nobody checked the solutions. However, it was impossible to solve all the problems in the time allocated. Only 1 minute in the experiment, a hired actor claimed to have finished all problems correctly and took the maximum possible payment, \$10. In this scenario, all the other participants observed that the actor cheated without any punishment.

**Assumption & Analysis:** For testing this hypothesis, we processed the dataset in a way that replicates the experiment just described. The assumptions are the following: (i) if VAC-banned during our 32-day observation period, we assume a player cheats in all matches played before he is VAC-banned; (2) we assume all players with whom the cheater played before being VAC-banned noticed he was cheating (and not punished); (iii) we assume all players who have not been VAC-banned by the end of our observation time never cheated (and thus were not seen cheating by their co-players); (iv) we consider all players who get VAC-banned in an interval that satisfies the influence timing condition and played with the recently VAC-banned player were influenced by him. These assumptions overestimate a cheater’s influence, thus the influence is an upper bound.

Table 2: Chi-square tests in four different time windows. FD: Freedom Degree

FD	$\chi^2$	P-value	Time Window (day)
2	9185.9	< 2.2e-16	0
2	5291.1	< 2.2e-16	7
2	3417.5	< 2.2e-16	14
2	2987.4	< 2.2e-16	21

Under these assumptions, players’ neighbors are divided into two categories: neighbors who cheated without being caught at match time and neighbors who did not cheat. We compare the fractions of players who turned cheaters and have neighbors in one of the categories above: at least one has been cheating or none has been cheating.

**Results:** Figure 2 plots users’ adoption of cheating as a function of number of their observed unpunished cheater neighbors. Three observations can be drawn.

First, overall, compared to no exposure to cheating, players who observed other players’ cheating behavior are more likely to cheat themselves. These results confirm the hypothesis and thus the in-lab experiments in Gino et al.’s study.

Second, when the number of observed uncaught cheaters changes from one to two, there is a dramatic increase in the likelihood of adopting the cheating behavior. For example, the change of cheating after observing two unpunished cheaters is 10 times larger than when observing one unpunished cheater for a time window of 0 (that is, under the assumption that all VAC-bans are delayed the same).

Third, the adoption of cheating decreases when players are exposed to three cheaters compared to when they are exposed to two. Even so, the likelihood to cheat is at least 11.8 times higher than when there is no exposure to unpunished cheating. This can be explained by the over-exposure phenomena in the co-match graph of Figure 1, where repeated exposures have decaying effects on users’ adoption. (We note that this scenario of multiple exposures was not tested in (Gino, Ayal, and Ariely 2009).) A group of Chi-Square tests shown in Table 2 reveal that the number of cheaters significantly differs with the increase of observed unpunished cheater neighbors.

## Factor II: In-group vs. Out-group Influence

**Hypothesis:** People are influenced by members of their groups more than by out-group members.

Gino et al. (Gino, Ayal, and Ariely 2009) discovered the difference between in-group and out-group influence in cheating engagement by doing the same experiment described in Factor I with one difference: the hired actor was making himself an in-group or out-group member by wearing a plain T-shirt, or another university’s T-shirt, respectively. The experimental results support the theory of social norm and social identity (Sherif 1936; Tajfel 2010) according to which people adopt the behavior of other in-group members and reject the same behavior if manifested by out-group members.

**Assumption & Analysis:** In most online multiplayer

games, players are divided into two teams. In this study we use team membership information to distinguish between in-group vs. out-group influence in cheating. Even if the teams are ephemeral, numerous prior studies in social sciences and management have showed that teams affect players’ in-game performance (Spotts and Chelte 2005; Hellerstedt and Aldritch 2008).

We assume that the in-team/out-team influence happens when a player’s teammate/opponent was labeled as a cheater after their match. To exclude mutual influence between in-team and out-team cheaters at the same time, we examined players who have either in-team or out-team influence each time. In the analysis, we compare the fraction of players who were VAC-banned after playing with cheater teammates vs. those playing with cheater opponents.

**Results:** As Figure 3 shows, across all results in four time windows, the level of cheating is dramatically influenced by teammates, and the influence is higher with the number of possible observations of cheating. For example, users’ adoption fraction increases from 0.027 to 0.279 when time window is 0. Our results confirm the above in-lab observations.

Another result (not addressed by Gino et al.) is that an additional in-team or out-team observation after two times does not affect adoption significantly. This tells us that initial exposures might increase contagion probability while further exposures appear to saturate contagion.

## Factor III: Awareness of Repercussions Limits Cheating

**Hypothesis:**The possibility of punishment limits cheating.

The intuition for this hypothesis is in the very existence of the VAC-ban label and the consequences associated to it (i.e., playing restrictions). However, the efficacy of the VAC-ban labels in cheating limitation has not been publicized, if it is known.

**Assumption & Analysis:** To measure the effect of the possibility of punishment on cheating adoption, we consider the visibility of the VAC-ban label to neighbors as such a reminder. This approach makes sense because VAC-bans are publicly visible regardless of the account’s privacy settings, they are permanent, and they are undesired. We divide players into two groups: one group of players played with punished cheater players (and thus, we assume, they observed their VAC-bans), while the others never played with VAC-banned gamers. We measure how much cheating adoption is in each group.

**Results:** Figure 4 shows the fraction of players’ engagement in cheating after seeing the cheating label before matches vs. without any visibility of cheating labels. It turns out that users’ cheating engagement fraction declines up to 83% (declines from 0.006 to 0.001) after exposure to cheating labels, suggesting that being reminded of the possibility of punishment (VAC ban labels) has positive effect in containing cheating. More interestingly, increased pre-exposure times has no prominent effect on adoption, resulting in the same fraction of cheating adoption when exposure frequency increases from one to two.

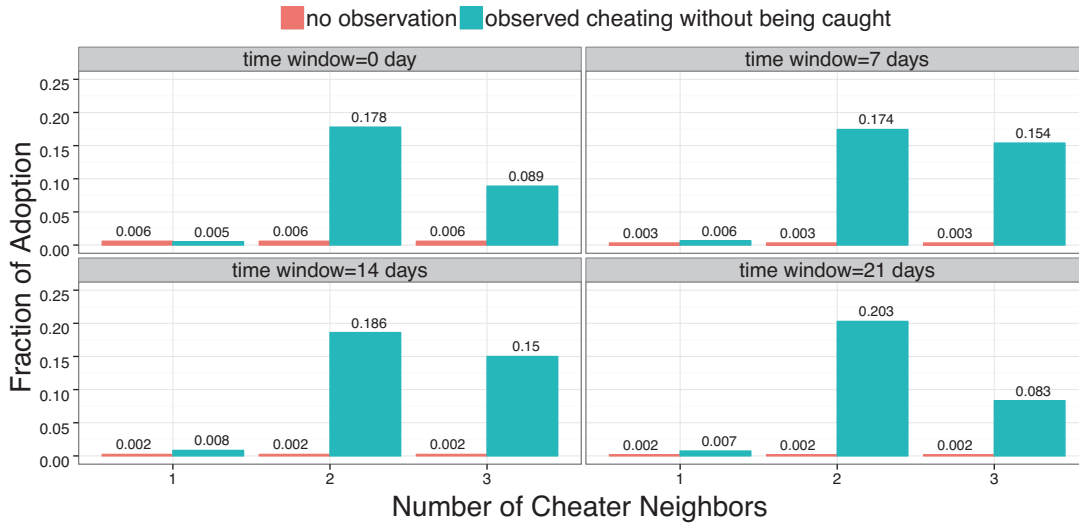


Figure 2: The cheating adoption fraction of players who observed their neighbors’ cheating without being caught in the match vs. those who had not observed any cheating.

#### Factor IV: Social Status Influences Cheating

**Hypothesis:** Social status influences the decision to cheat.

Social status is a way to classify individuals in terms of esteem and prestige acquired through economic success, education and accumulation of wealth (Adler et al. 2000; Oakes and Rossi 2003). The role of social class has been of long-standing interest to social scientists to understand how it can influence people’s thought and behavior. One group of theories sustain that upper-class individuals are more self-ish (Piff et al. 2010), donate smaller proportion of their incomes to charity (Hodgkinson and Weitzman 1990), and are more likely to engage in unethical behavior than lower-class individuals (Piff et al. 2012).

However, most of these studies did not consider the effect of penalties for unethical behavior that exist in real life and some online communities. We verify Piff’s theory of unethically inclined higher class individuals in an environment with punishment.

**Assumption:** We consider the Steam level as an indicator of a player’s social status in the gaming world. The Steam level is a number that summarizes a player’s gaming skills and obtained badges, shows off trading card collection and participation in Steam events, etc. Overall, it is a way to know how much time and effort someone has invested in their Steam account. Players can increase their levels by purchasing games, earning badges, etc. Higher levels come with benefits such as higher limit on the number of friends (general players are limited to 250 friends).

Next, we describe how to divide Steam players into ten groups. We built a web crawler to collect users’ Steam levels from their profiles. However, because of the privacy settings of user profiles, Steam levels are available only for users whose profiles are public. Table 3 gives the statistics of players’ levels in our dataset. We group players into dif-

ferent classes according to their levels. Specifically, we sort all users’ Steam levels from high to low, then partition them into ten groups called top  $x\%$  level (the level values in top 0–10% are larger than the values in top 10–20%), and each group have no overlapped individuals. Naturally, players in the top 0–10% group can be seen as upper-status individuals, and players in the bottom group (top 90–100%) can be seen as lower-status individuals. The first two columns of Table 4 give players’ distribution in each group. A Chi-Square test reveals that the value of levels significantly differs in each group ( $\chi^2(8)=5,865,100, p<2.2e-16$ ).

**Results:** Figure 5 depicts the relationship between players’ Steam levels and their likelihood of cheating in each group. We make two key observations from these results. First, globally, players with high levels demonstrate higher fraction of cheaters, which is consistent with the above theory. However, players with lower social status in our dataset presents similar cheating trends as the upper ones, having a high fraction of cheaters, which is inconsistent with the controlled experimental results in (Piff et al. 2010). We believe this difference is mainly caused by the existence of the Steam VAC-ban punishment. A VAC-banned account cannot play on any VAC-secured servers, cannot trade or sell items, and loses game items. Because of this penalty, players who invested a large amount of time and money in their accounts usually are careful not to lose them. On the contrary, individuals with lower status are at the initial phase to maintain their accounts, thus less money and effort have been put into the accounts (e.g., less weapons/games were bought, less badges and achievements were gained). Naturally, a VAC-ban is less costly for players who have not much to lose in terms of reputation or games – they can even abandon the banned account and create a new one. Even if upper-status players’ inventory is several thousands dollars

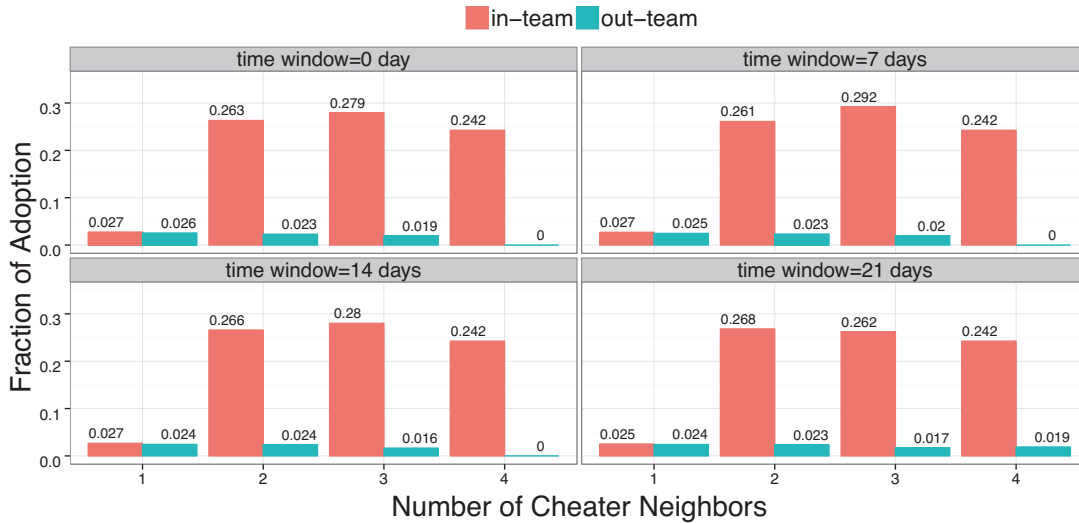


Figure 3: The cheating adoption fraction of players who were exposed to in-team members vs. those who were exposed to out-team members in scenarios of four different time windows.

and more than hundreds of hours spent on gaming, according to our results, upper-status individuals cheat more than other groups even comparable to low-status players.

Second, as the value of levels decrease, the fraction of cheaters declines and reaches the lowest point at the fifth group, where players’ levels rank in top 41 – 50%. After that, cheating has an upturn, and this tendency continues as level values further reduce.

These two observations provide strong evidence that in the gaming world with penalty systems, both high and low status are more likely to cheat, and the most ethical people appear in the middle. This might be the unique feature of the Steam Community caused by the mechanism of VAC-ban system, which could be extended to the real-world cheating behavior when cheating prevention/penalty systems apply.

**Excluded Factors:** To exclude some factors that might influence the above analysis other than social status itself, we try to answer two questions that closely relate to the two variables (fraction of cheaters and values of level) in Figure 5. First, are the value of levels (social classes) mainly influenced by players’ gaming time? Second, do players have more opportunities to cheat (accordingly, the fraction of cheaters is higher), if they spend longer time in the gaming community? To answer these two questions, we take a closer look at the correlation between time spent in gaming, level values and fraction of cheaters in each group. Because there is no information available on how much time a player spent playing games, we use each gamer’s account age to estimate his gaming hours. The account age is defined as the time period between the creation of the Steam account (marked on Steam Community profiles) and our crawling time, July 15th 2015. Table 4 gives two sets of Pearson correlations for each group, i.e., the value of levels vs. account age and fraction of cheaters vs. account age. The low correlation scores show almost no correlation between ac-

Table 3: Summary of the dataset of Steam levels.

Gamers	# of Cheaters	Range	Mean
903,585	58,291 (6.5%)	[0, 717]	10.8

count age, level, and fraction of cheaters. The only exception ( $cor.=0.2741$ ) happens with the 10th group (top 90 – 100%) where players have the lowest levels. But this correlation is still weak. It is reasonable that gaming time has higher influence on levels when low-level players have no other resources like advanced weapons or plenty of virtual coins. Upgrading to higher levels also means players are equipped with more gaming resources that could dilute the influence of gaming time. Overall, these results further indicate that players’ gaming time has no strong influence on the value of levels and fraction of cheaters in each group.

## Related Work

Social influence and homophily are two factors that always being discussed in social contagion processes. Do people befriend others who are similar to them, or do they become more similar to their friends over time via contagion? Distinguishing them requires longitudinal data on social relationships, individual attributes and complex methodologies (Lewis, Gonzalez, and Kaufman 2012; Shalizi and Thomas 2011). Many prior studies have attempted to distinguish these two factors. For example, Davin et al. (Davin, Gupta, and Piskorski 2014) studied user behavior on adoption of mobile applications and introduced latent space to control for homophily factors through simulations to separate the effects of homophily from peer influence. Aral et al. (Aral, Muchnik, and Sundararajan 2009) developed a dynamic matched sample estimation framework to distinguish



Figure 4: The cheating adoption fraction of players who were exposed to cheating labels before matches vs. those who were not exposed in scenarios of six different time windows.

Table 4: The Pearson correlation in top  $x\%$  levels. acc: account

Top % Level	# of players	cor. (level vs. acc. age)	p-value	cor. (frac. of cheaters vs. acc. age)	p-value
0-10%	33	-0.0250	0.8919	0.0037	0.9841
10-20%	44	0.0661	0.6738	-0.0592	0.7060
20-30%	58	0.1515	0.2652	-0.2036	0.1323
30-40%	104	-0.0931	0.3543	-0.0898	0.3721
40-50%	203	0.0089	0.9016	-0.0719	0.3153
50-60%	431	-0.0206	0.6723	0.0162	0.7395
60-70%	1,619	-0.0550	0.0287	0.0049	0.8445
70-80%	6,448	0.0046	0.7153	-0.0088	0.4833
80-90%	73,631	0.0044	0.2365	-0.0309	< 2.2e-16
90-100%	821,014	0.2741	< 2.2e-16	-0.0665	< 2.2e-16

influence and homophily effects in dynamic networks.

Unethical behavior is defined as a harmful, illegal or morally unacceptable behavior toward others in the larger community (Brass, Butterfield, and Skaggs 1998), which has been studied in psychology and sociology. It covers a variety of negative behavior, such as toxic behavior (Kwak, Blackburn, and Han 2015), cheating, trolls (Hardaker 2010), grieving (Foo and Koivisto 2004), etc.

Cheating as a specific unethical behavior has been a topic of interest from sociology and psychology to computer science. McCabe et al. (McCabe, Trevino, and Butterfield 2001) studied cheating in academic institutions and found that even if both individual and contextual factors have influence on cheating, contextual factors like peers' behavior are the most powerful influence. Gino et al. (Gino, Ayal, and Ariely 2009) conducted a group of controlled experiments to test the effects of different factors in cheating, which inspired

our work. Piff et al. (Piff et al. 2012) revealed that people in higher social classes are more likely to cheat than people in lower classes. Kang et al. (Kang et al. 2013) proposed a framework for bot detection in online games through user behavior analysis. Blackburn et al. (Blackburn et al. 2012; 2014) discovered that cheaters in online gaming worlds are well embedded in the social networks, and a fair player is more likely to become a cheater himself if he has more cheater friends. Others studied cheaters who gathered and exchanged virtual goods in online games for real money in black market (Keegan et al. 2010; Woo et al. 2011). In contrast to previous studies, by using a large dataset, this paper confirms theories about cheating that have remained untested outside of controlled laboratory experiments or only in small, survey-based studies.



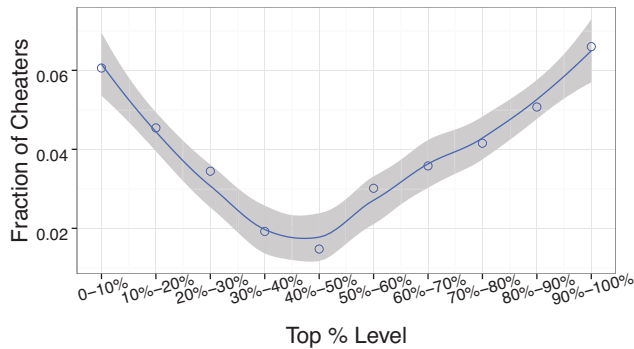


Figure 5: Fraction of cheaters in different groups where players are divided into 10 social status based on level. The shaded area represents the 95% confidence interval.

## Discussion and Conclusions

This paper presents a data-driven study of the factors that contribute to an online game player to adopt the cheating behavior of a neighbor in the social network. Our datasets include most of popular games from different companies, which represent behavior in general for gaming. Although the dataset gives us a unique opportunity to study influential factors behind unethical behavior, it has several limitations. First, because of the delay between the cheating time and the application of the VAC ban, we do not have the exact time of when a player cheats. But by using different time windows, covering all possible delay slots from days to weeks, we can approximate the evolution of cheating behavior in the network. Second, it is difficult to trace in which match a cheater was caught, and we do not know how many times a player cheated before being marked as a cheater on his profile. Thus, we estimate the number of players who witnessed cheating by counting all users who played with the cheater before his cheating label was shown. Third, a user may vaguely be aware of the cheating labels on his friends' profiles. It is possible that some players focused only on gaming seldom check their partners' or opponents' profiles, or they check these profiles after the match.

A typical concern in this type of data-driven analysis is distinguishing between homophily and influence (Lewis, Gonzalez, and Kaufman 2012; Aral, Muchnik, and Sundararajan 2009). In this case, however, we note that the *influence* of cheating has been confirmed in environments where users cannot express homophilious preferences, such as attendance of particular universities or a military service academy in the US. Thus, we are inclined to believe that the same phenomenon is at play in online gaming, and thus the observations we made are not the result of homophily but rather of influence.

Even if our datasets have limitations, our study represents an important step towards evaluating hypotheses of adoption of unethical behavior on network data of this scale, and provides a better understanding of cheating behavior and its influential factors.

## Acknowledgments

This research was supported by the U.S. National Science Foundation under Grants No. CNS 0952420 and IIS 1546453.

## References

- Aboukhadijeh, F. 2011. Cheating in video games.
- Adler, N. E.; Epel, E. S.; Castellazzo, G.; and Ickovics, J. R. 2000. Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy, white women. *Health Psychology* 19(6):586–592.
- Aral, S.; Muchnik, L.; and Sundararajan, A. 2009. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences* 106(51):21544–21549.
- Ariely, D., and Jones, S. 2012. *The (honest) truth about dishonesty*. Harper Audio.
- Blackburn, J., and Iamnitchi, A. 2013. Relationships under the microscope with interaction-backed social networks. In *1st International Conference on Internet Science*, 199.
- Blackburn, J.; Simha, R.; Kourtellis, N.; Zuo, X.; Ripeanu, M.; Skvoretz, J.; and Iamnitchi, A. 2012. Branded with a scarlet "c": cheaters in a gaming social network. In *Proceedings of the 21st International Conference on World Wide Web*, 81–90. New York, NY, USA: ACM.
- Blackburn, J.; Kourtellis, N.; Skvoretz, J.; Ripeanu, M.; and Iamnitchi, A. 2014. Cheating in online games: A social network perspective. *ACM Transactions on Internet Technology (TOIT)* 13(3):9.
- Blackburn, J. 2014. *An Analysis of (Bad) Behavior in Online Video Games, Jeremy Blackburn*. Ph.D. Dissertation, University of South Florida.
- Brass, D. J.; Butterfield, K. D.; and Skaggs, B. C. 1998. Relationships and unethical behavior: A social network perspective. *Academy of Management Review* 23(1):14–31.
- Carrel, S. E.; Malmstrom, F. V.; and West, J. E. 2008. Peer effects in academic cheating. *Journal of human resources* 43(1):173–207.
- Centola, D., and Macy, M. 2007. Complex contagions and the weakness of long ties<sup>1</sup>. *American Journal of Sociology* 113(3):702–734.
- Cha, M.; Mislove, A.; and Gummadi, K. P. 2009. A measurement-driven analysis of information propagation in the flickr social network. In *Proceedings of the 18th international conference on World wide web*, 721–730. ACM.
- Cosley, D.; Huttenlocher, D. P.; Kleinberg, J.; Lan, X.; and Suri, S. 2010. Sequential influence models in social networks. *ICWSM* 10:26.
- Crookes, D. 2014. Time for fair play: Online cheating by gamers is increasing-and it's seriously spoiling the fun. *The Independent*.
- Davin, J. P.; Gupta, S.; and Piskorski, M. J. 2014. Separating homophily and peer influence with latent space. *Available at SSRN*.

- Dransfield, I. 2014. Cs:go cheaters banned in record numbers.
- Economist, T. 2011. All the world's a game.
- Foo, C. Y., and Koivisto, E. M. 2004. Defining grief play in mmorpgs: player and developer perceptions. In *Proceedings of the 2004 ACM SIGCHI International Conference on Advances in computer entertainment technology*, 245–250. New York, NY, USA: ACM.
- Fortress2, T. 2013. Rampant cheating on valve tf2 servers.
- Gino, F.; Ayal, S.; and Ariely, D. 2009. Contagion and differentiation in unethical behavior the effect of one bad apple on the barrel. *Psychological science* 20(3):393–398.
- Hardaker, C. 2010. Trolling in asynchronous computer-mediated communication: From user discussions to academic definitions.
- Hellerstedt, K., and Aldritch, H. E. 2008. The impact of initial team composition and performance on team dynamics and survival. In *Academy of Management Proceedings*, number 1, 1–6. Briarcliff Manor, NY, USA: Academy of Management.
- Hodas, N. O., and Lerman, K. 2014. The simple rules of social contagion. *Scientific reports* 4.
- Hodgkinson, V. A., and Weitzman, M. S. 1990. *Giving and volunteering in the United States: findings from a national survey*. Washington, D.C.: Independent Sector, 1990 edition.
- Huizinga, J. 1950. *Home Ludens: A Study of the Play-Element in Culture*. Boston, Massachusetts: Ed. Beacon Press.
- Kang, A. R.; Woo, J.; Park, J.; and Kim, H. K. 2013. Online game bot detection based on party-play log analysis. *Computers & Mathematics with Applications* 65(9):1384–1395.
- Keegan, B.; Ahmed, M. A.; Williams, D.; Srivastava, J.; and Contractor, N. 2010. Dark gold: Statistical properties of clandestine networks in massively multiplayer online games. In *Social Computing (SocialCom), 2010 IEEE Second International Conference on*, 201–208. IEEE.
- Kulik, B. W.; O'Fallon, M. J.; and Salimath, M. S. 2008. Do competitive environments lead to the rise and spread of unethical behavior? parallels from enron. *Journal of Business Ethics* 83(4):703–723.
- Kwak, H.; Blackburn, J.; and Han, S. 2015. Exploring cyberbullying and other toxic behavior in team competition online games. In *The SIGCHI Conference on Human Factors in Computing Systems*, 3739–3748. New York, NY, USA: ACM.
- Lewis, K.; Gonzalez, M.; and Kaufman, J. 2012. Social selection and peer influence in an online social network. *Proceedings of the National Academy of Sciences* 109(1):68–72.
- Maiberg, E. 2014. Hacks! an investigation into the million-dollar business of video game cheating.
- McCabe, D. L.; Trevino, L. K.; and Butterfield, K. D. 2001. Cheating in academic institutions: A decade of research. *Ethics & Behavior* 11(3):219–232.
- Oakes, J. M., and Rossi, P. H. 2003. The measurement of ses in health research: current practice and steps toward a new approach. *Social Science and Medicine* 56(4):769–784.
- Piff, P. K.; Kraus, M. W.; Côté, S.; Cheng, B. H.; and Keltner, D. 2010. Having less, giving more: the influence of social class on prosocial behavior. *Journal of personality and social psychology* 99(5):77.
- Piff, P. K.; Stancato, D. M.; Côté, S.; Mendoza-Denton, R.; and Keltner, D. 2012. Higher social class predicts increased unethical behavior. *Proceedings of the National Academy of Sciences* 109(11):4086–4091.
- Pritchard, M. 2000. How to hurt the hackers: The scoop on internet cheating and how you can combat it. [http://www.gamasutra.com/view/feature/131557/how\\_to\\_hurt\\_the\\_hackers\\_the\\_scoop\\_.php](http://www.gamasutra.com/view/feature/131557/how_to_hurt_the_hackers_the_scoop_.php).
- Rettinger, D. A., and Kramer, Y. 2009. Situational and personal causes of student cheating. *Research in Higher Education* 50(3).
- Romero, D. M.; Meeder, B.; and Kleinberg, J. 2011. Differences in the mechanics of information diffusion across topics: idioms, political hashtags, and complex contagion on twitter. In *Proceedings of the 20th international conference on World wide web*, 695–704. New York, NY, USA: ACM.
- Shalizi, C. R., and Thomas, A. C. 2011. Homophily and contagion are generically confounded in observational social network studies. *Sociological methods & research* 40(2):211–239.
- Sherif, M. 1936. The psychology of social norms.
- Sinclair, B. 2015. Gaming will hit billions this year - newzoo.
- Spotts, H. E., and Chelte, A. F. 2005. Evaluating the effects of team composition and performance environment on team performance. *Journal of Behavioral and Applied Management* 6(2):127.
- Szell, M., and Thurner, S. 2010. Measuring social dynamics in a massive multiplayer online game. *Social networks* 32(4):313–329.
- Tajfel, H. 2010. *Social identity and intergroup relations*. Cambridge University Press.
- Valve. 2015. Steam support: I've been banned. <https://www.valvesoftware.com>.
- Wilson, C.; Boe, B.; Sala, A.; Puttaswamy, K. P. N.; and Zhao, B. Y. 2009. User interactions in social networks and their implications. In *Proceedings of the 4th ACM European conference on Computer systems*, 205–218.
- Woo, K.; Kwon, H.; chul Kim, H.; kwon Kim, C.; and Kim, H. K. 2011. What can free money tell us on the virtual black market? *ACM SIGCOMM Computer Communication Review* 41(4):392–393.