Does Concealing Gender Identity Help Women Win the Competition? An Empirical Investigation into Online Video Games

Xinlei (Jack) Chen

Xiaohua Zeng*

Cheng Zhang

Xinlei (Jack) Chen (xlchen@saif.sjtu.edu.cn) is a Professor of Marketing in Shanghai Advanced Institute of Finance, Shanghai Jiao Tong University, China. Xiaohua Zeng (xiaohua.zeng@phbs.pku.edu.cn) is the corresponding author and an Associate Professor at the Peking University HSBC Business School (PHBS), China. Cheng Zhang (zhangche@fudan.edu.cn) is a Professor at the School of Management, Fudan University, China.
Cheng Zhang acknowledges the support by the National Natural Science Foundation of China [Grants 91846302, 71871065 and 71832002].

Does Concealing Gender Identity Help Women Win the Competition? An Empirical Investigation into Online Video Games

Abstract

Signs of the gender gap are ubiquitous in society. Psychological theory suggests that when gender stereotypes are associated with competition, men exert greater effort against women (dominance effect) and women exert less effort against men (submissive effect), which implies that women are at a disadvantage when competing against men. Although multiple factors contribute to the gender gap, attempts to identify these factors have been hampered because gender, as a personal trait, is difficult to manipulate. Herein, the authors investigate submissive and dominance effects in the context of an online video game. They exploit a unique feature of the data: players have two-dimensional gender identities, one birth and one virtual. The results provide support for the dominance but not the submissive effect: when men perceive their opponent as female, they exert increased effort in competition, but women seem unaffected by their opponent's gender, which leads to poorer performance for women when competing against men, unless women conceal their gender. The findings provide important insights for how firms and regulators can help maintain gender equality in online environments. This paper also provides an example of how to assess social disparity with observational data by using a unique feature of the digital world.

Keywords: social disparity, concealed identity, gender, gender stereotypes, competition

1. Introduction

Signs of the gender gap are ubiquitous in the labor market, such as wage differences between men and women and the disproportionately small number of women in executive positions (Fuhrmans 2020). Multiple factors could contribute to this gap (e.g., ability, effort, discrimination), which often create confounding explanations in observational studies. Although experimental approaches have helped researchers draw causal inferences, they are often plagued by the inability to manipulate gender, as a personal trait, experimentally. Given the rapid development of digital technology, social and economic interactions increasingly occur in an online environment (or even the metaverse in the future), in which the parties can be anonymous or use fake signals to indicate their age, gender, and name. If people's decisions are affected by others' virtual identities, this context provides a rare opportunity to assess the gender gap with observational data.

An example of such a scenario is online video games, which had a global market size of US$151 billion in 2019,[1] surpassing the global entertainment (US$101 billion in 2019[2]) and recorded music (US$21.5 billion in 2019[3]) industries combined. Although video games have historically been a male-dominated activity, women have become an increasingly important part of the customer base. According to Statista, women accounted for 41% of computer and video gamers in the United States in 2020.[4] Indeed, in *Honor of Kings*, the most popular mobile game as of April 2020, with more than 100 million daily active users worldwide, women made up 54% of the user base.

[1] See report by Modor Intelligence.
[2] See report by the Motion Picture Association.
[3] See the annual Global Music Report from IFPI.
[4] https://www.statista.com/statistics/232383/gender-split-of-us-computer-and-video-gamers/.

Given the competitive nature of online video games, psychology theory predicts that gender gap could exist in the interpersonal competition. General stereotypes suggest that men are aggressive and women are passive. Moreover, many traits that characterize great competitors are taken as masculine in nature, while those of poor competitors are considered feminine. Men and women might thus perceive a correlation between competitor stereotypes and classic gender stereotypes (Kray et al. 2001), and this mapping process is likely to be implicit (Greenwald and Banaji 1995). When people are made implicitly aware of gender stereotypes, they often behave in conformity with those stereotypes in competition (Kray et al. 2001). For women, the effect may stem from a stereotype threat (Steele 1997): when they perceive a negative stereotype, they may suffer from negative arousal and fear of confirming the stereotype, which may hinder their performance. Conversely, just as negative stereotypes may impair performance, positive stereotypes may enhance performance (Spencer et al. 1999; Steele 1999). When men perceive that stereotypes of their gender could be an advantage in competition, they may perform better than they would have otherwise. In this regard, men may exert greater effort against women (dominance effect), and women may exert less effort against men (submissive effect).

While these theories suggest that women will be at a disadvantage during interpersonal competition, they largely rest on the perception of the opponent's gender. Many people, particularly women, actively conceal their gender identity in online video games. Online video games can be a toxic environment for women. Compared with male players, female players experience more severe harassment, such as stalking, sexual harassment, and sustained bullying, according to a Pew Research Center (2014) survey on online harassment. Many women simply quit the game, citing this reason. For those who stay, some mute their voices to go undetected, while others choose to conceal their identity by playing under ambiguous or male names. A survey by a Norwegian group found that approximately half of all female gamers had actively hidden their gender identity (Bergstrøm 2015). In their survey, Hussain and Griffiths (2008) also find that more women than men had taken on the persona of the opposite gender when playing, to prevent unsolicited male advances.

The phenomenon of concealing gender in online gaming provides a rare opportunity to study the gender gap. For example, if men perform better against women due to the dominance effect, will the advantage diminish if the women signal male virtual gender? Similarly, if women perform worse against men due to the submissive effect, will the disadvantage disappear if the men signal female virtual gender?

In this article, we investigate whether gender stereotypes affect interpersonal competition in an online video game and, if so, whether concealing identity helps eliminate the gender gap. Specifically, we analyze the outcome of player-versus-player (PVP) fights in a popular online video game in China.[5] Fighting is generally perceived as masculine, engendering a correlation between gender stereotypes and competition stereotypes in this context. For example, most people believe that men are more capable in fighting than women, and losing to a woman may even be considered shameful for a man. Such a gender stereotype could arise unconsciously when opposite genders compete against each other. As such, players might react to this gender stereotype by exerting more (for men) or less (for women) effort in a competition.

A common feature in the video game context is that players cannot see one another's faces but can see related images and usernames (at the top of the virtual image). Because the image is less informative of players' gender, players may infer the gender of their opponents through usernames, which we call "virtual gender." Notably, many players choose names that

[5] Given the confidential nature of the data, we do not disclose the name of the game or company.

imply the opposite gender, which creates a two-dimensional gender identity for each player (birth and virtual). We explain our identification strategy and exploit this feature in designing our regression analyses. For example, to test for the dominance effect, we compare a man's performance against virtual men with his performance against virtual women, while holding the opponent's birth gender fixed. Therefore, any observed performance gap would arise only from the male dominance effect (i.e., an increase in effort during competition against a perceived woman). We apply the same logic when designing the analyses that test for the existence of the submissive effect.

Our findings confirm that gender stereotype arises during interpersonal competition and thereby affects the competition outcome. Additional analysis reveals that this effect is due to the dominance, but not the submissive effect. That is, the performance gap in competition is mainly driven by the increased effort that men exert against women, but women seem unaffected by their opponents' gender. In other words, all the effects we observe are driven by men's perception of their opponent's gender; that is, men compete harder against virtual women than virtual men or an ambiguous gender, regardless of the opponents' birth gender. These findings suggest that gender stereotypes indeed affect women unfavorably, though this phenomenon is due to men's reactions to gender stereotypes rather than women's. Thus, concealing gender could help women alleviate such a negative impact and thereby reduce the gender gap in interpersonal competition.

Our paper is broadly related to studies on disparity and discrimination in society. Multiple factors, including generic difference, effort, and discrimination, can contribute to the disparity among individuals and groups; however, they are difficult to manipulate in experiments. Our research provides an example of using unique setting in the digital world to identify the underlying mechanism with observational data. Specifically, we provide direct evidence for how gender stereotype leads to the gender gap in interpersonal competition. By revealing this underlying mechanism, our research demonstrates that sometimes concealing instead of providing information (Busse et al. 2017; Cui et al. 2020) can reduce social gaps. Departing from extant literature in which competition outcome is determined by self-performance (e.g., maze solving; Gneezy et al. 2003), in our case the game outcome is determined by both players, which is a more challenging context to separate various effects.

In terms of managerial relevance, to our knowledge, our study is the first to examine the impact of gender stereotype and concealing gender in the online environment. Harassment is prevalent in the online environment overall, which makes concealing gender an unwilling choice for many people. For example, according to the Women's Media Center Speech Project, chat room participants with female usernames report receiving threatening or sexually explicit private messages 25 times more often than participants with male or ambiguous usernames.[6] Our research provides direct evidence for how concealing gender affects online experiences, which is managerially important to the firms. As we show, good online PVP experience increases players activity level, reduces churn rate, and increases spending in the game. However, when the PVP experience is negatively affected by the dominance effect, the likelihood of players churning and reducing spending increases, though gaming activity seems unaffected. And women are affected more by the dominance effect than men. These findings can help platform managers and regulators improve customer environment and maintain online gender equality.

The rest of the article proceeds as follows: we first review the relevant literature and then describe the gaming-related data we collected and their characteristics, the variables included in

[6] https://www.womensmediacenter.com/about/wmc-in-the-news/archive/2019/12.

the analysis, and the estimation sample. Next, we explain the empirical strategy and lay out necessary assumptions for identification, followed by three analyses of the data to test for the existence of the dominance and submissive effects. We then present evidence to validate our assumptions and carry out robustness checks to verify our results. We conclude with a discussion of our findings.

2. Literature Review

This research closely relates to the literature on gender gap in society. Although earlier research has attributed the gender gap to discrimination against women (Altonji and Blank 1999; Bertrand and Mullainathan 2004; Goldin and Rouse 2000; Kuhn and Shen 2012; Neumark 1996), more recent studies suggest that gender differences in competition could also be the cause. For example, Gneezy et al. (2003) and Niederle and Vesterlund (2007) find that men are more willing to engage in competition than women, which might be due to factors such as culture (Gneezy et al. 2009), age (Andersen et al. 2013; Sutter and Rützler 2010), nurturing (Booth and Nolen 2012), and task domain (Grosse and Riener 2010). Moreover, research has proposed that some methods mitigate the gender gap in competition, such as gender quotas (Niederle et al. 2013), preferential treatment, repetition of competition (Balafoutas and Sutter 2012), and the influence of female role models (Krishna and Orhun 2021).

Researchers have also examined the role of gender identity in competition. For example, Gupta et al. (2013) find that men are more willing to engage in competition when their opponents are women. Regarding performance, although Price (2008) shows that players perform better when their opponents are women, Antonovics et al. (2009) find such a performance effect only for men, and Gneezy et al. (2003) observe it only for women. Other studies show no effect of gender identity in competition (Dreber et al. 2011; Gneezy and Rustichini 2004; Lavy 2013).

Given the interpersonal nature of competition, gender stereotypes may arise in both mixed- and same-gender conditions; however, extant literature suggests that they are most likely to arise in mixed-gender competition. In addition, research has shown that people are more aware of their own gender when facing the opposite (vs. same) gender (Cota and Dion 1986; McGuire et al. 1979). Furthermore, any advantage or disadvantage implied by the gender stereotype is not perceived in the same gender context because each party is equally advantaged or disadvantaged (Kray et al. 2001).

Finally, our research is also related to the studies of how information provision affects social disparity. For example, auto repair shops are less likely to price discriminate against female customers when they reveal their knowledge about benchmark price (Busse et al. 2017), and acceptance rate of guest reservation requests from African Americans increases to average levels if they have positive review of prior stays with other hosts (Cui et al. 2020). In our case, online players withhold information about their gender identity (the neutral virtual gender case) or even signal the opposite gender identity. We show that these actions can help reduce gender gap in interpersonal competition.

3. Research Context and Data

3.1. Online Video Game

We obtained the data for this study from an action-based massively multiplayer online role-playing game in China. This video game attracts many young and competitive people. According to a survey conducted by the game company in early 2013, among a sample of 19,013 players in its gaming community, 30.2% were between the ages of 18 and 25 years, 47.6% were between the ages of 26 and 33 years, 20.6% were over 33 years, and the remaining 1.6% were under the age of 18 years. Thus, a large proportion of the game community comprises young working people. According to the same survey, 64.2% of the players were full-time employees, 24.8% were self-employed, and only 11.0% were students or unemployed.

In the game, each player can register for up to four characters and can assign a unique name to each character. For each character, the player can choose one of four standard roles: Archer, Cleric, Sorceress, or Warrior.[7] Each role has a distinct function, which is the primary reason for players to choose the role. After choosing a role, players control their characters and fight with the evil elements set by the game master. As with many online games, players' progress is based on their ability levels, and they progress through levels 1 to 40 as they achieve more prespecified goals. After reaching level 40, players can remain in the game, but they cannot achieve further change in their level.

In the game, although players compete primarily against the machine, they also frequently engage in PVP fights to practice their game-playing skills or to compete merely for enjoyment. Players create and join PVP fights through a central system. To initiate a PVP fight, a player simply enters the PVP arena and creates a new PVP fight by specifying the number of people who can enter the PVP room (2–16), the level requirement for participants, the PVP mode (e.g., whether winning is based on points or the number of rounds the player wins),[8] and the conditions for winning (e.g., the minimum number of rounds to win). After players create a PVP fight, it is listed in the system, and other players can search for it and join if they qualify. When a

[7] Our confidentiality agreement prohibits us from providing images of the roles.
[8] The most popular mode is based on the number of rounds, which accounts for 94% of one-versus-one PVP observations. Thus, our analysis focused on this mode.

PVP room has a sufficient number of people, players organize themselves into two teams of equal size (e.g., one vs. one, two vs. two, four vs. four) and begin to fight.

The PVP fights in this game have two notable rules. First, the outcome of a PVP fight does not affect players' level in the game, which is important to our analysis: if the outcomes of PVP fights were linked to players' performance in the overall game, players would strategically choose weaker opponents to avoid losses in PVP fights. Such a selection process would then affect the outcome of a PVP fight, which would be difficult to account for given our current data. Second, only players' skill, not the equipment used, influences the outcome of a PVP fight. This design was intended to encourage players of different levels to play against one another. This feature simplifies our analysis because controlling for the weapon effect in the analysis is difficult given the large number of weapons in the game.

Our PVP data apply to the four-month period from January 1 to April 30, 2011. For each PVP fight, we observed the beginning and ending time, the play mode, each player's character name, the role taken, and the points gained. The majority of PVP fights (73%) were one player versus another player, and 12% were two players versus two players; other types of PVP fights are less common. For ease of analysis, we examined only the one-versus-one PVP fights, which yielded 1,210,828 PVP fights and an average of 10,000 fights per day. These fights included 94,005 characters, or 61,980 unique players (as some players registered more than one character). Most PVP fights are brief, having an average duration of only seven minutes; however, in some cases, a fight can last as long as two hours.

3.2. Key Variables

3.2.1. Birth and virtual gender identities. Players in the video game see only one another's images and names. The images for the four standard roles, which are designed by the game company, have clear genders: two are male (Warrior and Cleric), and two are female (Archer and Sorceress). However, as discussed previously, a player chooses a role mainly for its functionality. For example, according to the company, players most often choose Warrior (59% of the time in the data) over the other three options (Cleric 15%, Archer 15%, Sorceress 11%) because this role is the easiest to control and has a good amount of mobility. Thus, a character's name could be a more salient feature than the character image in identifying a player's gender. We coded players' virtual genders by their character name, which may include numbers, characters (most in Chinese), symbols, or any combination thereof. We asked two Chinese research assistants to independently code each name on a five-point scale (1 = "very male," 2 = "somewhat male," 3 = "difficult to determine," 4 = "somewhat female," and 5 = "very female"). We then coded a name as male (M) if the average score was 2 or less, as female (F) if the average score was 4 or higher, and as neutral (N) otherwise. Table 1 provides examples of the coding.

We also obtained the birth gender of each player in the game. From 2007 onward, the Chinese government has required all online video game companies to collect identity card numbers (ICNs) from their customers.[9] The purpose is to prevent teenagers from becoming addicted to video games. The Chinese ICN contains both age and gender information of the cardholder. When a player is identified as younger than 18 years, he or she is assigned a fatigue index that records the daily cumulative hours of game playing. If the index is exceeded, warnings appear on the screen, and penalties (e.g., losing game points) will apply if the person continues to play.

[9] In 2010, the Ministry of Culture in China released the Interim Measures for the Administration of Online Games, which requires online gaming users to register using their real names with their identity documents.

Players are incentivized to be truthful when providing their ICNs. For example, if they want to recover a password, they must provide a copy of an identification card that matches the ICN they used to register. Occasionally, a player may be unwilling to reveal personal information and enter an invalid ICN. However, the system will detect the invalid ICN and automatically classify the player as a teenager with a fatigue index. Players may also use other people's ICNs obtained by searching online (occasionally, the ICNs of some people are released online for various reasons). In such a case, multiple players may share the same ICN. We asked the game company to check for instances of multiple players using the same ICN. It detected 200 ICNs being shared by multiple players (12% of the players in the initial sample provided), so we omitted these players from the sample.

3.2.2. Performance measure. We observed each player's performance in PVP fights. The play mode we focused on is called "Round," which is similar to a boxing match, in that two players fight for a number of rounds specified by the initiator of the PVP fight. Each round lasts for three minutes. If no character is killed during a round, the round is recorded as a draw, and both players will continue, though any wounds incurred carry over to the next round. However, if one player kills the other, he or she earns one point, while the opponent earns zero points. Afterward, both players are reborn with no wounds, and they begin the next round.

Because PVP fights differ in the number of rounds, players in fights with more rounds may obtain more points than those in fights with fewer rounds. As such, comparing performance across PVP fights simply by considering the points earned would be inaccurate. We addressed this concern by employing a relative measure of performance. Specifically, in a PVP fight, if Player A earned S_A points and Player B earned S_B points, we measured Player A's performance as $\left(\frac{S_A}{S_A+S_B}\right)$ and Player B's performance as $\left(\frac{S_B}{S_A+S_B}\right)$. Here, (S_A+S_B) refers to the total number of times a player killed the opponent, and $\left(\frac{S_A}{S_A + S_B}\right)$ refers to Player A's likelihood of killing Player B. If the players reached draws in all rounds (i.e., $S_A = 0,\ S_B = 0$), we set the performance of both players to .5 (i.e., in the limited time, we consider both players equally capable of killing the other).

3.2.3. Player's skill. Because the outcome of a PVP fight is highly dependent on each player's skill, using a proper measure of skill was critical. Fortunately, we had data on each player's ability level in the game. In contrast with the relative performance outcome in a PVP fight, which depends on the opponent, the system calculated and updated players' ability level over time depending on the direct outcomes of each player competing against the machine; thus, this measure was objective and comparable across players. For each PVP fight, we used the most recent level achieved by each player.

In addition to using the system's recorded ability level, we included two other measures to control for a player's skill: (1) the number of days since a player had registered at the time of each PVP fight and (2) the number of PVP fights a player had participated during the data period. Because a PVP fight includes two players, we used relative measures in the analysis. For example, for the first measure, we used the number of days since player *i* registered divided by the number of days since player *j* registered.

3.3. Calibration Sample

Taking all the aforementioned information into account, we then selected a valid sample for our analysis. First, as noted, we focused on one-versus-one PVP fights for easier identification. Second, we needed PVP observations with information on both players' birth gender and virtual name. In our data, information on players' birth gender was not available for 46% of the users, either because they registered with the game company before 2007 (when they were not required to provide personal identity information) or because they submitted invalid information. Therefore, we omitted these players and their PVP observations from our data. Third, we excluded players at the highest level (i.e., level 40), who accounted for 28% of the players. As we noted previously, after level 40, players cannot further increase their level even when they become more skillful by continuing to engage in PVP fights; thus, these players' level may not reflect their true skill. In addition, they are likely to have different objectives from lower-level players because they have finished all the prespecified tasks and may be staying in the game for purposes other than winning (e.g., to make friends). Our final sample contains 79,957 PVP fights among 23,851 characters (20,144 unique players; birth gender 85% male, 15% female).

We report descriptive statistics in Panels A–C of Table 2. Panel A shows that the majority (85%) of the characters are registered by men, which reflects that the customer base of this online video game is primarily male. Both men and women prefer male-oriented roles: men chose male characters in two-thirds of the cases, and women chose them more than half the time. In addition, players commonly conceal their birth gender by signaling the opposite gender through their virtual name: not only do the majority (79%) of the characters registered by women have male-oriented or neutral names, but nearly one-third of the characters registered by men also have female-oriented or neutral names. In more than half the cases (54%), women identified their characters using male-oriented names, and men occasionally gave female-oriented names to their characters (12% of the cases). Panel B shows the gender composition for the PVP fights. Regarding players' birth gender, most PVP fights were between two men (77%), approximately 20% were between a man and a woman, and only 3% were between two women. Note that players were aware only of their opponents' virtual gender, not their birth gender. When we analyzed the data solely on the basis of players' virtual gender, the distribution became more even, with 43% of PVP fights between two virtual males, 16% between a virtual male and a virtual female, 2% between two virtual females, and the remainder between neutral genders. Panel C shows the descriptive statistics for PVP activities. On average, PVP fights involve six rounds and last seven minutes. In addition, the players on the two sides are, on average, within four ability levels of each other. Men and women are similar in their skill levels; however, men appear to play more and win slightly more rounds in PVP fights. Most of the statistics are consistent between the original and calibration samples, except for the level of the players (because the calibration sample omits players at level 40). Furthermore, players with missing information on their personal identity are most likely long-time customers who began playing before 2007 and thus have a higher average level.

4. Empirical Tests

The goal of our empirical tests was to examine how gender identity affects competition as a result of the submissive or dominance effect. We first explain the empirical strategy and lay out necessary assumptions to identify the effect of gender on the outcome of mixed-gender competition. Because the validation of some assumptions coincides with the tests for gender effects, we discuss the validation part in the next section. Next, we conduct three analyses to identify the gender stereotype in the competition; for Analyses 1 and 2, we first explain the test design and hypothesis and then conduct regression analysis followed by result interpretation. In Analysis 1, we demonstrate the gender gap in mixed-gender competition. In Analysis 2, we test the submissive and dominance effects. In Analysis 3, we demonstrate that neutral virtual gender does not activate gender stereotypes.

4.1 Empirical Strategy

As discussed previously, multiple factors could contribute to the gender gap. If women perform worse against men, it could be due to poorer skill or less effort owing to stereotype conformity. Without perfect control of player's skill, it is impossible to identify the effect of gender stereotype on the competition outcome with observational data. In addition, if skill and gender are highly correlated, an experimental approach cannot deliver a satisfactory answer either, because the gender or skill could not be randomly assigned among players.

Fortunately, in our context (as well as in many other digital contexts) a person has a two-dimensional gender identity: the birth gender and the virtual gender. Most people in the virtual world do not know each other in person, so they must rely on virtual identity to infer others' birth identity. Thus, if some people conceal their gender identities, a separation can occur between genetic factors (e.g., skill, ability) determined by birth gender and gender identity perceived by others. As long as the potential competition outcome is independent from the manipulation of the virtual gender conditional on the player's attributes (the "unconfoundedness" assumption), we are able to attribute the observed gender gap in the competition outcome to specific gender effects. For example, if we fix the birth genders of a pair of PVP players but alternate the virtual gender of the opponents, the gap in the competition outcome for the focal player, if any, can be attributed to the focal player's reaction to the virtual gender of the opponents due to stereotype conformality.

In general, we need to make four assumptions for our context:

A1: Players use their opponent's virtual gender to infer birth gender.

A2: The potential outcome of competition is independent from the selection of virtual gender conditional on the player's attributes.

A3: Selection of the opponent in PVP is not strategic.

A4: A player's behavior is determined by his or her birth gender.

Here, A1 is the cornerstone of our identification strategy, which could be violated if players communicate and actually know each other's birth gender. A2 and A3 refer to two selection mechanisms, which, if they exist, could confound our theory and bias the estimates. Finally, A4, if violated, suggests an alternative mechanism of gender stereotype.

For each assumption, we do some of the following: (1) provide theoretical support from literature, (2) demonstrate evidence from the data pattern, (3) conduct robustness checks to assess the extent of bias, or (4) rule out the alternative explanations. Some of our analyses use the same design as our main empirical tests in this section; therefore, we present the discussion of the assumptions in a separate section.

4.2 Analysis 1: Gender Gap in Competition

In Analysis 1, we examine the gender gap in competition—that is, whether men have an advantage over women in mixed-gender competition. Both dominance and submissive effects will lead to the gender gap. Our data show that the average performance of men playing against women is .55, which is significantly higher than .5 ($t = 17.5$, $p < .001$; $N = 17{,}217$). This statistic indicates that men perform better than women in mixed-gender competition before we control for other factors, including skill.

4.2.1. Method and hypothesis. We denoted each player by using a combination of two letters, the first indicating the player's birth gender and the second indicating the player's virtual gender. For example, MF indicates that a player is male with a female virtual identity. For this analysis, we chose players whose virtual gender was consistent with their birth gender (i.e., MM and FF). We examined players' performance variation in the following scenarios:

Scenario 1: MM vs. MM/FF

Scenario 2: FF vs. MM/FF

For Scenario 1, to examine an MM player's performance variation between competing against another MM player and competing against an FF player, we performed the following regression:

$$Y_{ij} = \alpha_i + \beta F_j + \gamma Z_{ij} + \varepsilon_{ij}, \qquad (1)$$

where i is the MM player, j is his opponent (i.e., MM or FF), and Y_{ij} is the logit-transformed relative performance score of player i when playing against j (i.e., $Y_{ij} = \ln[S_{ij}/(1 - S_{ij})]$). Because Y_{ij} is undefined when S_{ij} is 0 or 1, we use a modified form of the transformation by adding $1/(2n)$ to an $S_{ij} = 0$ observation and by subtracting $1/(2n)$ from an $S_{ij} = 1$ observation, where n is the number of rounds of the PVP fight. Next, α_i is the individual fixed effect for player i. The dummy variable F_j indicates whether j is a female player (FF in this case), which is the focal variable in our analysis. Note that we are unable to control for player j's fixed effect because it will be correlated with F_j. Z_{ij} includes both i and j's ability levels, the number of days since each player's registration at the time of each PVP fight, and each player's total number of PVP fights during the data period. Z_{ij} also includes role interaction fixed effects. Because each of the four roles has special skills and thus may interact with other roles in different ways, we obtained $4 \times 4 = 16$ role interaction fixed effects. However, because we also have individual fixed effects at the character level and because one character is associated with only one role, we included only 12 role interaction fixed effects in the model to avoid linearity. In addition, we clustered the standard errors at the individual level to control for the potential interdependency of individual decisions across competitions. We then performed a similar regression for Scenario 2.

We illustrate our hypothesis in Panel A of Table W1 in the Online Appendix. Each cell (a, b) indicates the adjustment in effort for each player when he or she played against an opponent. Here, “a” is the adjustment for player *i*, “b” is the adjustment for player *j*, “+” indicates an increase in effort, “–” indicates a decrease in effort, and “0” indicates no adjustment in effort.

In Scenario 1, when an MM player competes against an FF player, the MM player will increase his effort because of the dominance effect, whereas the FF player will decease her effort because of the submissive effect. Thus, the effort adjustment for two players is represented as (+, –) in the table. However, such an adjustment in effort will not occur when an MM player competes against another MM player, which is represented as (0, 0). Therefore, we expect that an MM player will perform better when competing against an FF player than when competing against another MM player. Similarly, in Scenario 2, when an FF player competes against an MM player, the FF player will decrease her effort because of the submissive effect, while the MM player will increase his effort because of the dominance effect. Thus, the effort adjustment for two players is represented as (–, +). However, such an adjustment in effort will not occur when an FF player competes against another FF player, which is represented as (0, 0). Therefore, we expect that an FF player will perform better when competing against another FF player than when competing against an MM player. We essentially determined the estimate of the coefficient for F_j by comparing the performance of player *i* when he or she competed against an MM player with the performance when he or she competed against an FF player. In both Scenarios 1 and 2, we expect the coefficient of F_j to be positive.

4.2.2. Results. We report the results in Table 3. Consistent with our hypothesis, the coefficients for F_j are positive and significant in both regressions, indicating that men (women) perform better (worse) in mixed-gender competition than in single-gender competition. These findings support the existence of a gender gap in competition, which may be due to both submissive and dominance effects.

4.3. Analysis 2: Dominance Versus Submissive (Own Reaction)

We demonstrate the gender gap in the competition in Analysis 1. In Analysis 2, we aimed to identify which effect, dominance or submissive, leads to the gender gap. We carefully designed the test to identify the dominance and submissive effects through a player's performance gap induced by his or her own adjustment in effort.

4.3.1. Method and hypothesis. The design of the test is straightforward. Our hypothesis predicts that men will increase their effort when competing against women because of the dominance effect while women will decrease their effort when competing against men because of the submissive effect. Therefore, we examine a simple setting in the PVP fight. Suppose that a player encounters two opponents in two PVP fights. If the two opponents' virtual names suggest they are of different genders from each other but they actually share the same birth gender, will the player's performance vary in the two PVP fights?

To test this idea, we examined the following scenarios in the PVP fights to identify the submissive and dominance effects:

Dominance effect:

- Scenario 1: MM vs. MM/MF
- Scenario 2: MM vs. FM/FF

Submissive effect:

- Scenario 3: FF vs. MM/MF
- Scenario 4: FF vs. FM/FF

In each scenario, we selected observations from PVP fights involving one type of player competing against two types of opponents. While the opponents are actually of the same birth gender, their names suggest they have different genders. We examined the performance of an MM player to identify the dominance effect and the performance of an FF player to identify the submissive effect. Similar to Analysis 1, we performed a regression analysis for each scenario as in Equation 1 to examine whether the player's performance changed when he or she faced different types of opponents; the only difference was that now F_j indicates whether j's virtual gender is female.

We illustrate our hypothesis regarding the players' adjustment in effort in Panel B of Table W1 in the Online Appendix. In each scenario, given that the birth genders of both i and j are fixed, i's performance varies only in his or her reaction to the opponent's virtual gender. In Scenarios 1 and 2, the performance variance is due to the dominance effect, while in Scenarios 3 and 4, it is due to the submissive effect. Take scenario 2 (MM vs FM/FF) as an example. When MM competes with FM, MM perceives FM as a male and therefore does not adjust his effort, but FM will decrease her effort due to submissive effect. Therefore, the prediction for the two players' effort adjustment is (0, –). When MM competes with FF, MM will increase his effort due to dominance effect and FF will decrease her effort due to submissive effect. Therefore, the prediction for the two players' effort adjustment is (+, –). Comparing these two cases, the focal player MM will perform better when the opponent is FF, purely due to the dominance effect from focal player MM, while his opponent's submissive effect in two cases will be canceled out in the comparison.

4.3.2. Results. We report the regression results in Table 4. In terms of the relative performances, the results show that MM players perform better when they compete against opponents with female virtual names. In Scenario 1, the estimated coefficient for F_j is .136, indicating that the MM player's odds of winning when he competes against MF players are higher than when he competes against MM players. In Scenario 2, the coefficient for F_j is .177, indicating that for an MM player, the odds of winning when he competes against FF players are higher than when he competes against FM players. These results provide support for the dominance effect; that is, men increase their effort when they believe they are competing against women, even if the opponents are actually men. By contrast, women seem unaffected by their opponent's gender. In both Scenarios 3 and 4, the coefficients for F_j are nonsignificant, suggesting that FF players' relative performance does not depend on the virtual gender of their opponents. Thus, we find no evidence of the submissive effect in two regressions.

In summary, we find that men become more aggressive when competing against women, whereas women are insensitive to gender stereotypes when competing against men. Therefore, we find support for the dominance effect but not the submissive effect. In Part D of the Online Appendix, we offer another test of the dominance and submissive effects using opponents' reactions to the focal player's virtual gender and find similar results.

4.4. Analysis 3: Neutral Gender

In the sample, a significant proportion of players used gender-neutral virtual names. As noted previously, a gender-neutral name can be a symbol, a number, a word, or a phrase, thus providing no information about the player's gender. Conceptually, a neutral gender will not trigger gender stereotypes; therefore, we expect men to react to the neutral virtual gender similarly to the male virtual gender and women not to react to any gender signal, as in the previous analysis.

To test this assumption, we used the opponents whose virtual gender was neutral as the control, and we included two dummies in the model: M_j, which indicates whether j's virtual gender is male, and F_j, which, as before, indicates whether j's virtual gender is female. We report the results in Table 5. We find that women typically do not react to the gender signal from their opponents, with only one regression yielding a significant F_j dummy estimate. In addition, men's behavior remains the same whether they are playing against virtual men or virtual neutrals, though they show significantly increased effort when playing against virtual women than when playing against virtual neutrals.

5. Assumption Validation

As stated previously, our empirical strategy rests on several key assumptions. A1 is the cornerstone of the analysis, which cannot be violated. A2 and A3, if violated, will bias the estimation but could be fixed with selected methods and better controls. A4 offers an alternative explanation that can be tested using our data.

In this section, we explain theoretic reasoning for A1 and A4, provide data evidence for A1 and A3, conduct a robustness check for A2, and rule out the alternative explanation for A4. Given space limitations, we report the full results in the Online Appendix but provide important results herein. We also discuss the potential bias if some of the assumptions are violated in the general discussion.

5.1. A1: Players Use Opponent's Virtual Gender to Infer Birth Gender

Certainly, players in the game could be aware that their opponents' virtual names may not be consistent with birth gender. If so, why do men exhibit discriminatory behavior toward virtual gender despite potentially knowing that most players are men? One explanation is conversational norms (Grice 1975; Sperber and Wilson 1986); that is, players are likely take the virtual gender as norm and favor the acceptance of it as the birth gender unless cues for suspicion are perceived. A second explanation is confirmatory bias (Hoch and Ha 1986; Hoch and Deighton 1989); that is, players could take the virtual name as a hypothesis and test it in subsequent interactions (e.g., chatting with and observing their opponent's behavior). If the information from these interactions is ambiguous, confirmatory bias in gender recognition may arise.

As a third explanation, players could be suspicious of the credibility of a virtual name all the time. According to rational expectation, their belief will be consistent with the aggregate distribution in the population. Table 2 shows that when the virtual name signals a woman, the probability that the player is female is nearly twice that when the virtual name signals a man; when the virtual name signals a man, the probability that the player is male is 1.3 times that when the virtual name signals a woman. Therefore, if players adjust their effort according to the statistical belief about the opponent's birth gender conditional on virtual gender, a male player will exert more effort when facing a virtual woman than a virtual man. Conversely, a female player will exert less effort when facing a virtual man than a virtual woman.

Naturally, if players know each other's birth gender, we should expect that the virtual gender poses little effect on the competition outcome. Although we have no information about players' social connection in the real world, we can use the frequency of encounters as a proxy of familiarity among players, reasoning that as the familiarity increases, players may communicate and get to know each other better. Specifically, we measure the familiarity between two players in a PVP by the number of times they have played PVP against each other before. Our calibration sample has 43,825 pairs in total. Among them, 68% of pairs played only once, 17% played twice, 7% played three times, 3% played four times, and 5% played five or more times.

We replicate Analysis 2 using the following equation:

$$Y_{ij} = \alpha_i + \beta D_{ij} F_j + \gamma Z_{ij} + \varepsilon_{ij}, \tag{2}$$

where D_{ij} is a vector of dummy variables indicating the number of times i and j had played against each other before this PVP fight. The other variables are the same as in Equation 1. We report the estimation results in Table W2 in the Online Appendix. It shows that the gender effect decreases with the increasing number of encounters before a PVP, which confirms our hypothesis.

5.2. A2: The Potential Outcome of Competition Is Independent of the Selection of Virtual Gender Conditional on the Player's Attributes

An alternative explanation is that players choose virtual names according to factors correlated with the competition outcome, such as ability. For example, players choosing male virtual identities possess stronger ability than those choosing female virtual identities. This may explain why players perform better when competing against virtual women than virtual men.

While we could not completely rule out such a possibility, we used propensity score matching to control for the selection of treatment (in this case, the selection of virtual gender). Specifically, for each player i, we used propensity score matching to construct a sample of opponents who are comparable in terms of the likelihood of choosing female virtual gender. We executed our matching procedure in two steps.

First, we estimated a player's propensity to choose virtual female gender as a function of observed variables, including birth gender, age, the virtual role of the character, and proxies of ability. We obtained the age information from the players' ICNs. For each player, we construct two variables to measure his or her ability. One is the average level ("Avg. level") across this player's PVP fights during the data period. The other one measures the speed of progressing. We use the maximum level achieved by the player during data period divided by the number of days from the character registration to the end of data period. Although we have controlled for levels in regressions, including level information in the propensity score helps to strengthen our results.

Using the full calibration sample, we ran a binary logistic regression to compute the propensity score for 18,645 characters who chose either male or female virtual gender. We report the estimation result in Table W3 of the Online Appendix. Players who are female, are younger, and adopt the female role image are more likely to choose a female virtual gender. In addition, players with higher speed of progressing are less likely to choose female virtual gender. However, the propensity to choose a female virtual gender is positively correlated with the average level, which reflects the statistics that in our sample the average level of virtual women is higher than that of virtual men, as shown in the next section.

Second, we used the one-nearest-neighbor algorithm to match players that are closest together. We conducted the matching analysis for Analysis 2, in which the selection problem confounds with the gender stereotype theory. Take the analysis of MM vs. MM/MF as an example. For each PVP between player *i* and an opponent MM, we matched it with another PVP between player *i* and an opponent MF, such that this MF player had the closest propensity score to that of the MM player. If the MF plays multiple PVPs with player *i*, we randomly selected a PVP. Following Gordon et al (2019), we matched on the log-odds ratio to linearize the value on the unit interval to improve estimation (Rubin 2001). Unfortunately, this method resulted in a poor match. To improve the matching, we imposed a caliper that is 20% of the standard deviation of the log-odds ratio of the propensity scores across players in the calibration sample (Austin 2011). We performed the matching with replacement because it reduces bias and does not depend on the sort order of the data (Gordon et al. 2019). We report the absolute standardized differences of covariate means before and after the matching in Table W4, and we plot the density of the propensity scores before and after matching in Figure W1, both in the Online Appendix. Our findings show that matching with caliper delivers satisfactory matching samples for most analyses, except for FF versus FM/FF, for which the sample size is too small to deliver a satisfactory matching sample.

With the matched sample, we reran Analysis 2, and Table 6 presents the estimation results. All the results are consistent with our original findings, though the coefficient of F_j is only marginally significant for analysis of MM versus FM/FF (p = .079), potentially due to the significant drop in sample size. We also note that the results for analysis of FF versus FM/FF are not reliable due to the poor matching in the sample.

5.3. A3: Selection of the Opponent in PVP is Not Strategic

Players could be strategic when choosing their opponents in PVP. For example, competitive men might be more likely to choose virtual women as opponents, and competitive women might be more prone to choosing virtual men as opponents. If virtual men are on average more competitive than virtual women, these selection processes could inaccurately indicate that the dominance effect is in effect but not the submissive effect.

We took several steps to address these concerns. First, we needed to control for players' ability in the analysis. Ideally, we would use the fixed effect to control for opponents' skills, but as mentioned previously, doing so was impossible because the opponent's fixed effect would be correlated with the dummy variable of the opponent's virtual gender in most of our analyses. Therefore, the best we could do is to control for opponent's level in the analysis. Second, we checked whether it is the case that virtual men are on average more competitive than virtual women. The calibration sample contains 18,414 characters of either virtual male or female gender and with level information. A character may appear multiple times in the PVP observations, such that the calibration sample has 159,914 character appearances. In Table W5 in the Online Appendix, we report the distribution of virtual gender conditional on birth gender, the average level for each virtual gender, and the sample t-test for the level difference between virtual genders. At the character appearance level, in both birth male and birth female groups, the average level of virtual women is higher than that of virtual men, and the sample t-test also confirms this. At the character level, each character may appear multiple times in the data, and with a different level for each appearance; therefore, we calculate the average level for each character in the sample period. The distribution is quite similar to that of character appearance. Similarly, the average level of virtual woman is higher than that of virtual man in each birth gender group, again confirmed by sample t-tests. Thus, the data show that virtual women are more competitive than virtual men.

Does a player's competitiveness affect his or her preference of opponents' virtual gender? To answer this question, we took all players' PVP observations as the sample and ran a logistic regression with the opponent's virtual gender as the dependent variable. For the independent variables, we included the focal player's birth gender, own level, and the interaction between these two variables. The other independent variables are the same as in Equation 1 except that we do not have focal player's fixed effect because that is correlated with focal player's birth gender. The estimation results (reported in Table W6 in the Online Appendix) show that a player's level does not significantly affect his or her preference regarding the opponent's virtual gender, and there is no significant difference between birth male and birth female players in terms of this preference.

Therefore, our analysis shows that (1) competitive players do not appear more likely to choose virtual women as opponents and (2) virtual female players are more competitive than virtual male ones. These data patterns imply that the performance gap between a focal player and a virtual woman will shrink, owing to the high skill of virtual women. If the opponent's ability is perfectly controlled, no estimation bias would result. Otherwise, it makes our tests for dominance and submissive effect both conservative. We further discuss the impact of this possibility in the general discussion.

5.4. A4: A Player's Behavior Is Determined by Birth Gender

Each player in the video game has two gender identities, one real and one virtual. Although we assume that a player's behavior is determined by his or her birth gender, the identity priming literature (for a review, see Bargh 2006) suggests that the process of choosing a virtual name might constitute a priming process, such that when a man chooses a female virtual name, he may prime a female identity and consequently behave like a woman, and a similar priming process may occur for a woman who chooses a male name. Identity priming is thus an alternative explanation for the effect of gender identity on people's performance in competition. In our case, this priming effect explains some of our findings. For example, when an MM player competed against MM/MF players, we observed that the MM player performed better when competing against an MF player than when competing against an MM player. Whereas our theory attributes the gender gap solely to the dominance effect (i.e., the MM player increases his effort during competition against an MF player), the identity priming theory suggests that the MF player may behave as he perceives that a woman would behave and therefore decreases his effort when playing against an MM player. Thus, according to identity priming theory, the gender gap is caused by a combination of both dominance and submissive effects.

However, we surmise that the identity priming theory will not dominate in our context for several reasons. First, the priming from true gender is a self-stereotype, while the priming from the opposite gender is an other-stereotype. Moreover, the behavioral effects of other-stereotypes may be dominated by self-stereotypes (Wheeler and Petty 2001). Therefore, this type of priming process will not be strong enough to affect the behavior effect of the process we propose herein. Second, our empirical tests provide evidence to rule out this alternative explanation. We first present in Table 7 the hypothesis of the adjustment in effort for Analysis 2 according to both our explanation and the identity priming theory. In our tests for the dominance effect (MM vs. MM/MF, MM vs. FM/FF), we found that both explanations predict that an MM player will perform better when competing against opponents with a female virtual gender. Our explanation suggests that this result is solely due to the dominance effect, whereas the identity priming theory suggests that the outcome is due to the combination of both the dominance and submissive effects. In this test, we could not distinguish between these two explanations.

However, in the tests for the submissive effect (FF vs. MM/MF, FF vs. FM/FF), though both explanations predict that an FF player will perform better when competing against opponents with a female virtual gender, the reasons underlying these predictions differ: our explanation suggests that this outcome is purely due to the submissive effect, whereas the identity priming theory suggests that the outcome results from both the dominance and submissive effects. Because we found no significant difference in performance, our explanation suggests that there is no evidence to support the submissive effect. However, the alternative view cannot offer an explanation because in the test for the dominance effect, the combination of both the dominance and submissive effects is significant.

6. Robustness Checks

We checked the robustness of our findings using other designs and methods, such as different gender combinations and pooled regression. We report most estimation results in the Online Appendix except those from pooling regression. In Part A in the Online Appendix, we also include an analysis to examine the effect of role gender—that is, whether role gender also triggers gender stereotypes in the competition.

6.1. Test with Other Gender Combinations

In the previous tests for the dominance effect, we examined an MM player's reactions, and in the tests for the submissive effect, we examined an FF player's reactions. In other words, we focused on the reactions of the players whose virtual gender was consistent with their birth gender. Alternatively, we created a similar design by examining the reactions of MF and FM players to determine whether these players react to gender stereotypes in the same way. To examine an MF player's reactions, we can compare, for example, MF versus FF players and MF versus FM players, with the MF versus FM player scenario as the control condition (as both players perceive the opponent as having the same gender, no stereotype is activated). Similarly, for an FM player, we can compare FM versus MF players and FM versus MM players, with the FM versus MF player scenario as the control condition.

We illustrate our hypothesis in Table W8a and report the findings in Table W8b in the Online Appendix. Most of the findings are consistent with our previous findings. For example, we confirm that women (i.e., FM players) are insensitive to the virtual gender of their opponent. Regarding the dominance effect (i.e., MF players), we replicated our findings in one analysis (MF players vs. MM/MF players), though the effect is nonsignificant in another analysis (MF player vs. FM/FF players).

6.2. Pooled Regression

In our current analysis, for a given player *i*, we fixed the birth gender of *i*'s opponents to test the effect of virtual gender on competition outcome. An alternative method is to allow a common baseline against which an alternative combination of birth and virtual genders could be evaluated—the so-called pooled regression. By incorporating multiple gender and virtual gender combinations into one analysis, pooled regression not only allows us to test both dominance and submissive effects simultaneously, but also enables us to test the dominance and submissive effects in different ways.

For Analysis 2, we first ran a pooled regression for dominance effect by combining the MM versus MM/MF and MM versus FM/FF analyses using following equation:

$$Y_{ij} = \alpha_i + \beta_1 MF_j + \beta_2 FM_j + \beta_3 FF_j + \gamma Z_{ij} + \varepsilon_{ij}, \tag{3}$$

where MF_j, FM_j, and FF_j are dummies indicating the birth and virtual gender combination of opponent *j*. We take the opponent to be MM as the baseline. The other variables are the same as in Equation 1. Table 8 displays the estimation results, which indicate that, compared with the baseline (MM as opponent), MM performs significantly better when facing MF (β_1 = .142), indicating a dominance effect. By contrast, MM performs the same when facing FM compared with the baseline, indicating no submissive effect. Finally, MM performs better when facing FF compared with the baseline (β_3 = .137), indicating that at least one effect (dominance or submissive) exists. The estimate of MF indicates that it is the dominance effect.

Similarly, we ran another pooled regression for submissive effect by combining the FF versus MM/MF and FF versus FM/FF analyses using following equation:

$$Y_{ij} = \alpha_i + \beta_1 FM_j + \beta_2 MF_j + \beta_3 MM_j + \gamma Z_{ij} + \varepsilon_{ij}. \tag{4}$$

Here, we took the opponent to be FF as the baseline and report the estimation results in Table 9. First, FF performed significantly worse when facing MM compared with the baseline ($\beta_3 = -.373$), indicating that at least one effect (dominance or submissive) exists. However, FF performed the same when facing FM compared with the baseline, indicating no submissive effect. Second, FF performed only marginally significantly worse when facing MF compared with the baseline ($\beta_1 = -.324$, $p = 0.08$), indicating a marginal dominance effect at best, potentially due to the small sample.

Note that in both pooled regressions, we were able to test the dominance and submissive effects at the same time. In addition, the method to test these effects could be different from our previous tests in Analysis 2. For example, in the first pooled regression (reported in Table 8), we tested submissive effect by comparing MM's performance competing against FM and competing against the baseline (MM), whereas in the second pooled regression (reported in Table 9), we tested dominance effect by comparing FF's performance competing against MF and competing against the baseline (FF).

7. General Discussion and Conclusion

Our research presents an example of how to assess social disparity with observational data using the increasing popular digital context. Our focus herein is the gender gap in competition. Research shows that women often conceal their gender identity in online video games to escape harassment. Our study reveals another benefit of concealing identity: it also helps women better compete in these games. Using players' relative performances in PVP fights in an online video game in China, we found support for the dominance effect but not for the submissive effect. Specifically, men tended to exert more effort and, thus, perform better when they perceived their opponents as female, even if these opponents were actually male. In other words, men adjusted their effort level depending on the perceived gender of the opponent they faced in competition. When players were perceived as female, their performance was thus weaker, potentially hurting their gaming experience. However, concealing identity could help women alleviate such an effect.

Why, then, do women not care about stereotypes? One explanation could be that women who play online video games are a specific type who are comfortable with their own gender and therefore are less concerned with fulfilling gender stereotypes. Although we have no information to verify this speculation, we know that such self-selection is common in many competitive environments, in which players choose competitive situations rather than being randomly assigned. Women who are more competitive and thus self-select into a competition could differ to a greater extent from the rest of the population. Moreover, literature suggests that stereotypes arise only when the test performance is diagnostic of one's stereotype-relevant ability (Steele and Aronson 1995) or when the domain is important (e.g., for defining him or herself) (Wheeler and Petty 2001). Because the video game context is considered more of a male domain, women may be less likely to be judged in such situations. Thus, they do not face a stereotype threat (or do not feel the pressure to perform well).

In addition to this study's academic contribution, our findings have significant marketing implications. Naturally, interpersonal competition is an important part of online video game experiences. While players enjoy socializing and exploring in the virtual world, winning against other players is unarguably an important part of the consumption experience as well. As we show in Parts B and C in the Online Appendix, the outcome of PVP competition is positively correlated with gaming activity level, retention rate, and spending in the game. However, the existence of gender stereotype affects the competition outcome and, therefore, the online experience. We found that for players affected by the dominance effect, the extent of the dominance effect's influence increases the churn rate and reduces the frequency of purchasing game coins, though their gaming activity remains unaffected. Notably, in terms of churn rate, women are more affected by the dominance effect than men.

Our findings have important policy implications as well. They suggest that in the current, often toxic online environment, allowing women to conceal their identity could be an effective way to not only protect them from online harassment but also enhance their gaming experience. However, the question arises: Is it fair that women must do this? Policy makers should consider whether there are other ways to alleviate such a negative effect from a gender stereotype. Perhaps the most important issue is to educate consumers about the existence of such a stereotype. Kray et al. (2001) show that when the stereotype is explicitly activated, people exhibit stereotype reactance and behave inconsistently with the stereotype. In addition, although gender identity and related stereotypes make people perceive the opposite gender as a member of a distinct group, Kray et al. show that a commonly shared, positive identity helps transcend the between-groups distinction. As a result, it encourages cooperative behavior and a greater focus on mutual beneficial agreement.

This study has three limitations. First, although we have discussed the assumption that players do not choose the opponents strategically (Assumption 3), the pairing of players in a PVP fight might still be endogenous, and competitors may be friends in real life and coordinate their play on the platform. Although we did not have real-life friendship data for our sample, we obtained a small survey of 2,000 PVP fights from the game company. We found that only 13%

of the players were online friends in the game. Given that offline friends could also be online friends, it seems likely that only a small portion of the PVP fights would occur between friends.

Second, in terms of the submissive effect, though nonsignificant, the sign of all the estimates do point in the direction of submissive effect. One plausible explanation is the small sample size in many of the tests for submissive effects (as shown in Table 4). However, this explanation does not hold in every case. For example, in the analysis with other gender combinations (Table W8b in the Online Appendix), one regression for the submissive effect (FM vs. MM/MF) has a sufficient number of observations (5,751 observations) but still revealed no effect. Furthermore, in the pooled regression MM versus MM/MF/FM/FF (Table 8), we were able to test the submissive effect by comparing the estimates of FM with the baseline (MM), and we observed an insignificant difference, which indicates no submissive effect. Note that in this regression, the observations for MM and FM are 48,678 and 4,346, which should be sufficient.

Another explanation is that men and women may exhibit different performance sensitivity: given that women have on average lower skill than men in this game, they may have lower sensitivity to the performance variation of opponents, which would explain smaller performance gap for women when competing against various opponents. Indeed, in Analysis 2 (Table 4), the performance gap for women is smaller than that for men, judging by the magnitude of the estimates. However, in Analysis 1 (Table 3), we found just the opposite: when facing both MM and FF, women's performance gap (.411) is much larger than that of men (.112), implying a greater performance sensitivity for women.

Thus, the final and most likely explanation is the one we discussed for Assumption 3: the virtual woman possesses greater skill than the virtual man, which reduces the performance gap in all cases. If players' ability is not perfectly accounted for, considering that the sample for submissive tests was smaller than that for dominance tests, these combined factors may have caused the insignificance of the submissive effect. Future research could confirm this supposition with a larger test sample and a better control for players' ability.

Third, although we have shown men compete more aggressively against women than against men, there is no direct evidence that it is solely due to dominance effect. There could be reasons other than gender stereotype which prompt men exert more effort playing against female opponents. For example, some men maybe more interested and focused to play against opponents with opposite genders, which is not unusual in a social platform like video game. While these factors do not suggest harassment, they lead to similar disadvantage to female players in the competition. While the answer to this question is inconclusive in this paper, future research with primary information regarding players' thoughts will provide direct link between the gender effect and the underline process.

References

Altonji JG, Blank RM (1999) Race and gender in the labor market. Ashenfelter O, Card D, eds. *Handbook of Labor Economics*, Volume 3C (Amsterdam: North-Holland):3143–3259.

Andersen S, Ertac S, Gneezy U, List JA, Maximiano S (2013) Gender, competitiveness , socialization at a young age: Evidence from a matrilineal and a patriarchal society. *Review of Economics and Statistics* 95(4):1438-1443.

Antonovics K, Arcidiacono P, Walsh R (2009) The effects of gender interactions in the lab and in the field. *Review of Economics and Statistics* 91(1):152-162.

Austin PC (2011) Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. *Pharmaceutical Statistics* 10(2):150-161.

Balafoutas L, Sutter M (2012) Affirmative action policies promote women and do not harm efficiency in the laboratory. *Science* 335(6068):579-582.

Bargh JA (2006) What have we been priming all these years? On the development, mechanisms , ecology of nonconscious social behavior. *European Journal of Social Psychology* 36(2):147-168.

Bergstrøm II (2015) Online computer games force women into the closet. Kilden genderresearch.no (February 17). https://partner.sciencenorway.no/computer-games-forskningno-harassment/online-computer-games-force-women-into-the-closet/1414246.

Bertrand M, Mullainathan S (2004) Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination. *American Economic Review* 94(4):991-1013.

Booth A, Nolen P (2012) Choosing to compete: How different are girls and boys? *Journal of Economic Behavior & Organization* 81(2):542-55.

Busse MR, Israeli A, Zettelmeyer F (2017) Repairing the damage: The effect of price knowledge and gender on auto repair price quotes. *Journal of Marketing Research* 54(1):75-95.

Cota A, Dion KL (1986) Salience of gender and sex composition of ad hoc groups: An experimental test of distinctiveness theory. *Journal of Personality and Social Psychology* 50:770–776.

Cui RM, Li J, Zhang D (2020) Reducing discrimination with reviews in the sharing economy: Evidence from field experiments on Airbnb. *Management Science* 66(3):1071-1094.

Dreber A, von Essen E, Ranehill E (2011) Outrunning the gender gap—Boys and girls compete equally. *Experimental Economics* 14(4):567-82.

Fuhrmans V (2020) Where are all the women CEOs? *The Wall Street Journal* (February 6). https://www.wsj.com/articles/why-so-few-ceos-are-women-you-can-have-a-seat-at-the-table-and-not-be-a-player-11581003276.

Gneezy U, Leonard, KL List JA (2009) Gender differences in competition: Evidence from a matrilineal and a patriarchal society. *Econometrica* 77(5):1637-1664.

Gneezy U, Niederle M, Rustichini A (2003) Performance in competitive environments: Gender differences. *Quarterly Journal of Economics* 118(3):1049-1074.

Gneezy U, Rustichini A (2004) Gender and competition at a young age. *American Economic Review* 94(2):377-381.

Goldin C, Rouse C (2000) Orchestrating impartiality: The impact of blind auditions on female musicians. *American Economic Review* 90(4):715-741.

Gordon BR, Zettelmeyer F, Bhargava N, Chapsky D (2019) A comparison of approaches to advertising measurement: Evidence from big field experiments at Facebook. *Marketing Science* 38(2):193-225.

Greenwald AG, Banaji MR (1995) Implicit social cognition: Attitudes, self-esteem, stereotypes. *Psychological Review* 102:4-27.

Grice HP (1975) Logic and conversation. Cole P, Morgan JL, eds. *Syntax and Semantics, Vol. 3: Speech Acts* (New York: Academic Press):41-58.

Grosse ND, Riener G (2010) Explaining gender differences in competitiveness: Gender-task stereotypes. Jena Economic Research Papers No. 2010,017.

Gupta DN, Poulsen A, Villeval MC (2013) Gender matching and competitiveness: Experimental evidence. *Economic Inquiry* 51(1):816-835.

Hoch SJ, Deighton J (1989) Managing what consumers learn from experience. *Journal of Marketing* 53(2):1-20.

Hoch SJ, Ha Y-W (1986) Consumer learning: Advertising and the ambiguity of product experience. *Journal of Consumer Research* 13(2):221–233.

Hussain Z, Griffiths MD (2008) Gender swapping and socializing in cyberspace: An exploratory study. *CyberPsychology & Behavior* 11(1):47-53.

Kray LJ, Thompson L, Galinsky A (2001) Battle of the sexes: Gender stereotype confirmation and reactance in negotiations. *Journal of Personality and Social Psychology* 80(6):942-958.

Krishna A, Orhun AY (2021) Gender (still) matters in business school. *Journal of Marketing Research* doi: 10.1177/0022243720972368.

Kuhn P, Shen K (2013) Gender discrimination in job ads: Evidence from China. *Quarterly Journal of Economics* 128(1):287-336.

Lavy V (2013) Gender differences in market competitiveness in a real workplace: Evidence from performance-based pay tournaments among teachers. *Economic Journal* 123(569):540-73.

McGuire WJ, McGuire CV, Winton W (1979) Effects of household sex composition on the salience of one's gender in the spontaneous self-concept. *Journal of Experimental Social Psychology* 15:77–90.

Neumark D (1996) Sex discrimination in restaurant hiring: An audit study. *Quarterly Journal of Economics* 111(3):915-941.

Niederle M, Segal C, Vesterlund L (2013) How costly is diversity? Affirmative action in light of gender differences in competitiveness. *Management Science* 59(1):1-16.

Niederle M, Vesterlund L (2007) Do women shy away from competition? Do men compete too much? *Quarterly Journal of Economics* 122(3):1067-1101.

Pew Research Center (2014) Online harassment. http://www.pewinternet.org/2014/10/22/online-harassment/.

Price J (2008) Gender differences in the response to competition. *Industrial and Labor Relations Review* 61(3):320-333.

Rubin DB (2001) Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services Outcomes Res. Methodology* 2(3/4):169–188.

Spencer SJ, Steele CM, Quinn DM (1999) Stereotype threat and women's match performance. *Journal of Experimental Social Psychology* 35:4-28.

Sperber D, Wilson D (1986) *Relevance: Communication and Cognition* (Cambridge, MA: Harvard University Press).

Steele CM (1997) A threat in the air: How stereotypes shape intellectual identity and performance. *American Psychologist* 52:613-629.

Steele CM (1999) Stereotype threat and performance. Paper presented at the 11th Annual Conference of the American Psychological Society, Denver.

Steele CM, Aronson J (1995) Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology* 69(5):797-811.

Sutter M, Rützler D (2010) Gender differences in competition emerge early in life. IZA Discussion Paper No. 5015.

Wheeler SC, Petty RE (2001) The effects of stereotype activation on behavior: A review of possible mechanisms. *Psychological Bulletin* 127:797-826.

Table 1. Examples of Virtual Names

Examples of Virtual Name	Virtual Gender as Coded
四爷	Male
冷漠丨 John 丶	Male
獨愛大刀	Male
George	Male
丶草莓味◎	Female
风铃、晓	Female
卡芙莲	Female
ImUrQueen	Female
天涯星空	Neutral
离凡	Neutral
snow	Neutral
369258147	Neutral

Table 2. Gender Composition and Activities

A: Gender Distribution: Birth, Role, and Virtual

Birth Gender	Role Gender	Virtual Gender	# of Characters	Percentage
M	M	M	9,779	41.0%
		F	981	4.1%
		N	2,491	10.4%
	F	M	3,655	15.3%
		F	1,485	6.2%
		N	1,766	7.4%
F	M	M	1,282	5.4%
		F	211	0.9%
		N	437	1.8%
	F	M	712	3.0%
		F	542	2.3%
		N	510	2.1%

B: Gender Composition Among PVPs

Player 1	Player 2	# of PVPs	Percentage
Birth Gender			
M	M	61,250	77%
M	F	16,496	20%
F	F	2,211	3%
Virtual Gender			
M	M	34,748	43%
M	F	13,027	16%
F	F	1,425	2%
N	M	22,035	28%

N	F	3,973	5%
N	N	4,749	6%

C: PVP Activities

Variable	N	Mean	SD	Min	Max
Calibration Sample					
In General					
# of rounds of each PVP	79,957	5.91	3.45	0.00	42.00
Play time of the PVP (min.)	79,957	7.15	5.05	0.02	85.95
Diff. in levels of the PVP pair	76,081	3.60	3.99	0.00	29.00
Birth Gender = M					
Player's level	19,920	23.66	8.90	10.00	39.00
PVP performance	138,996	0.50	0.38	0.00	1.00
# of PVPs played per month	20,157	2.82	7.84	0.25	204.00
Birth Gender = F					
Player's level	3,647	23.06	8.68	10.00	39.00
PVP performance	20,918	0.48	0.38	0.00	1.00
# of PVPs played per month	3,694	2.21	5.67	0.25	126.00
Original Data (including players with missing IDs and those at level 40)					
# of rounds of each PVP	1,210,828	5.69	3.29	0.00	53.00
Play time of the PVP (min)	1,210,828	6.08	4.13	0.01	149.10
Diff. in levels of the PVP pair	1,168,812	1.91	4.21	0.00	30.00
Player's level	2,376,381	36.17	7.21	10.00	40.00

Table 3. Analysis 1: Gender Gap

	MM vs. MM/FF	**FF vs. MM/FF**
	DV: MM's Performance	DV: FF's Performance
F_j (opponent's gender is female)	.112* (.058)	.411** (.151)
Opponent's level	−.067** (.003)	−.064** (.010)
Own level	.064** (.003)	.054** (.013)
Days after registration (i divided by j)	.007** (.002)	−.003 (.006)
Number of PVPs played (i divided by j)	.009** (.001)	.016** (.004)
Role interaction fixed effects	Yes	Yes
Individual fixed effects	Yes	Yes
Number of observations	50,230	2,001
Adj. R^2	.357	.403

Notes: Standard errors in parentheses are clustered at the individual level. DV: dependent variable.

**$p < .01$, *$p < .05$.

Table 4. Analysis 2: Dominance/Submissive Effects (Own Reaction)

	Dominance Effect		Submissive Effect	
	MM vs. MM/MF	**MM vs. FM/FF**	**FF vs. MM/MF**	**FF vs. FM/FF**
	DV: MM's Performance	DV: MM's Performance	DV: FF's Performance	DV: FF's Performance
F_j (opponent's virtual gender is female)	.136** (.031)	.177* (.076)	.075 (.125)	.099 (.273)
Opponent's level	−.067** (.002)	−.073** (.007)	−.066** (.009)	−.038 (.021)
Own level	.063** (.003)	.066** (.009)	.051** (.013)	.038 (.025)
Days after registration (*i* divided by *j*)	.005** (.002)	.000 (.004)	−.004 (.006)	−.009 (.006)
Number of PVPs played (*i* divided by *j*)	.009** (.001)	.006** (.001)	.014** (.004)	.031* (.012)
Role interaction fixed effects	Yes	Yes	Yes	Yes
Individual fixed effects	Yes	Yes	Yes	Yes
Number of observations	56,978	5,962	2,148	520
Adj. R^2	.352	.469	.353	.581

Notes: Standard errors in parentheses are clustered at the individual level. DV = dependent variable.
**$p < .01$, *$p < .05$.

Table 5. Neutral Gender

	MM vs. MM/MF/M0 (M0 as base)	**MM vs. FM/FF/F0 (F0 as base)**	**FF vs. MM/MF/M0 (M0 as base)**	**FF vs. FM/FF/F0 (F0 as base)**
	DV: MM's Performance	DV: MM's Performance	DV: FF's Performance	DV: FF's Performance
M_j	.037 (.023)	−.000 (.062)	.012 (.108)	.235 (.314)
F_j	.179** (.035)	.158* (.083)	.130 (.138)	.548* (.275)
Opponent's level	−.065** (.002)	−.076** (.006)	−.061** (.009)	−.058** (.019)
Own level	.060** (.003)	.063** (.008)	.056** (.014)	.044* (.022)

Days after registration (i divided by j)	.003* (.001)	.001 (.003)	−.007 (.005)	−.009 (.005)
Number of PVPs played (i divided by j)	.008** (.001)	.006** (.001)	.016** (.004)	.008* (.004)
Role fixed effects	Yes	Yes	Yes	Yes
Individual fixed effects	Yes	Yes	Yes	Yes
Number of observations	75,985	9,997	2,974	730
Adj. R^2	.340	.455	.344	.560

M0: Male with neutral virtual name.

F0: Female with neutral virtual name.

Notes: Standard errors in parentheses are clustered at the individual level. DV = dependent variable.

$^{**}p < .01$, $^{*}p < .05$.

Table 6. Analysis 2 with Propensity Score Matching

	Dominance Effect		Submissive Effect	
	MM vs. MM/MF	**MM vs. FM/FF**	**FF vs. MM/MF**	**FF vs. FM/FF**
	DV: MM's Performance	DV: MM's Performance	DV: FF's Performance	DV: FF's Performance
F_j (opponent's virtual gender is female)	.119* (.049)	.195 (.110)	.005 (.188)	.153 (.619)
Opponent's level	−.048** (.006)	−.047* (.020)	−.035 (.023)	.039 (.033)
Own level	.053** (.008)	.060* (.028)	.021 (.032)	.033 (.046)
Days after registration (*i* divided by *j*)	.009* (.004)	−.007 (.013)	.027** (.010)	.049 (.028)
Number of PVPs played (*i* divided by *j*)	.007** (.001)	.002 (.002)	.015 (.008)	.118* (.055)
Role interaction fixed effects	Yes	Yes	Yes	Yes
Individual fixed effects	Yes	Yes	Yes	Yes
Number of observations	35,610	1,090	1,002	66
Adj. R^2	.328	.352	.316	.812

Notes: Standard errors in parentheses are clustered at the individual level. DV = dependent variable.

$^{**}p < .01$, $^{*}p < .05$.

Table 7: Hypotheses/Expectations from Gender Stereotype and Identity Priming Theory

	Gender Stereotype			Identity Priming		
Player i	Player j (Opponent)		Expected Coefficient of F_j	Player j (Opponent)		Expected Coefficient of F_j
Scenario 1						
	MM	MF		MM	MF	
MM	(0, 0)	(+, 0)	+	(0, 0)	(+, −)	+
Scenario 2						
	FM	FF		FM	FF	
MM	(0, −)	(+,−)	+	(0, 0)	(+, −)	+
Scenario 3						
	MM	MF		MM	MF	
FF	(−, +)	(0, +)	+	(−, +)	(0, 0)	+
Scenario 4						
	FM	FF		FM	FF	
FF	(−, 0)	(0, 0)	+	(−, +)	(0, 0)	+

Notes: Each cell (a, b) indicates the adjustment in effort for each player when he or she played against an opponent. Here, "a" is the adjustment for player *i*, "b" is the adjustment for player *j*, "+" indicates an increase in effort, "−" indicates a decrease in effort, and "0" indicates no adjustment in effort.

Table 8: Pooling Regression for Analysis 2 (MM vs MF/FM/FF/MM)

	Estimate	**Std. Error**
F_j (opponent is MF)	.142**	.030
F_j (opponent is FM)	.049	.034
F_j (opponent is FF)	.137*	.062
F_j (opponent is MM) as baseline	—	—
Opponent's level	−.067**	.002
Own level	.062**	.003
Days after registration (*i* divided by *j*)	.005**	.001
Number of PVPs played (*i* divided by *j*)	.009**	.001
Role interaction fixed effects	Yes	
Individual fixed effects	Yes	
Number of observations	62,940	
Adj. R^2	.350	

Notes: Standard errors are clustered at the individual level.

**$p < .01$, *$p < .05$.

Table 9: Pooling Regression for Analysis 2 (FF vs FM/MF/MM/FF)

	Estimate	**Std. Error**
F_j (opponent is FM)	−.309	.195
F_j (opponent is MF)	−.324	.187
F_j (opponent is MM)	−.373*	.148
F_j (opponent is FF) as baseline	—	—
Opponent's level	−.059**	.008
Own level	.043**	.013
Days after registration (*i* divided by *j*)	−.003	.006
Number of PVPs played (*i* divided by *j*)	.015**	.004
Role interaction fixed effects	Yes	
Individual fixed effects	Yes	
Number of observations	2,668	
Adj. R^2	.381	

Notes: Standard errors are clustered at the individual level.

**$p < .01$, *$p < .05$.