# Star-Cursed Lovers: Role of Popularity Information in Online Dating

Behnaz Bojd[*]

University of California, Irvine

Hema Yoganarasimhan[*]

University of Washington

March 15, 2021

## Abstract

We examine the effect of user's popularity information on their demand in a mobile dating platform. Knowing that a potential partner is popular can increase their appeal. However, popular people may be less likely to reciprocate. Hence, users may strategically shade down or lower their revealed preferences for popular people to avoid rejection. In our setting, users play a game where they rank-order members of the opposite sex and are then matched based on a Stable Matching Algorithm. Users can message and chat with their match after the game. We quantify the causal effect of a user's popularity (star-rating) on the rankings received during the game and the likelihood of receiving messages after the game. To overcome the endogeneity between a user's star-rating and her unobserved attractiveness, we employ non-linear fixed-effects models. We find that popular users receive worse rankings during the game, but receive more messages after the game. We link the heterogeneity across outcomes to the perceived severity of rejection concerns and provide support for the strategic shading hypothesis. We find that popularity information can lead to strategic behavior even in centralized matching markets if users have post-match rejection concerns.

# 1 Introduction

Throughout human history, people have relied on their extended families, social networks, and religious organizations to help them find romantic partners. However, they are now increasingly turning to online dating for this purpose. The most recent *Singles in America Survey* found that the number one meeting place for singles is now online (Safronova, 2018). According to a study from Pew Research Center, 30% of U.S. adults ($\approx$ 99 million adults) reported that they have used online dating services (Anderson et al., 2020). Indeed, industry revenues for online dating now exceed three billion dollars a year in the United States (IBISWorld, 2019).

Early businesses in this industry were mostly websites that allowed users to create detailed profiles, browse/search other users' profiles, and then establish contact through email exchanges. However, these websites suffered from the problems common to most decentralized two-sided matching markets such as costly search and congestion (Niederle et al., 2008). Not only is browsing and contacting potential partners costly in time and effort, but the efforts are often fruitless due to congestion, i.e., a few attractive people get a ton of messages and most get nothing.

Over the years, mobile dating apps have replaced dating websites as the dominant form of online dating because they address some of the above problems, and offer a much simpler way for users to find matches (Ludden, 2016). First, users are shown a set of potential partners and asked to state their preference for them on some scale (e.g., rank-order them, vote up or down, or swipe right or left) within a fixed period of time. These stated-preferences are then fed into a matching schema/algorithm, which matches users who have expressed some preference for each other. The first step reduces search costs and the second step minimizes rejection concerns. Thus, today's mobile dating apps increasingly resemble centralized matching markets, where a central algorithm allocates matches based on some revealed preferences.

The way information is presented in mobile dating apps has also evolved to reflect the simpler search process. Because users are only given a short (and fixed) amount of time to decide how much they like someone, most dating apps have moved away from showing long detailed profiles. Instead, they show a small set of salient pieces of information that a user can process easily (e.g., photo and age of the potential partner). Many of them also display a summary measure of the popularity of a potential partner (e.g., star-rating, number of likes) next to her/his profile. The benefits of showing users' popularity information are – (a) it is easier to process one cumulative popularity measure instead of parsing through detailed profile data, and (b) popularity measures can provide information on a potential partner's appeal in the dating market, and thereby help users calibrate the likelihood of achieving a match with that person.

However, there is no research that examines the effect of popularity information on users'

demand in a two-sided dating platform. In this paper, we are interested in two related questions. First, we seek to quantify the causal effect of a user's popularity information on her/his demand measures in a centralized dating market. Second, we are interested in identifying the source of these effects (if any), i.e., pin down the mechanism behind them.

In a dating market, popularity information can have both positive and negative impact on demand. On the one hand, revealing that a potential partner is popular can increase her/his appeal, which in turn can increase a user's revealed preference for that potential partner (Hansen, 1977). On the other hand, a very popular potential partner is also more likely to have other options (or interest from other users) and therefore may be less likely to reciprocate any interest. Thus, a user who wants to avoid rejection my reveal lower preference (or *strategically shade* down her/his preference) for a popular user. A priori, it is not clear which of these effects will dominate, and what would be the overall impact of popularity information on demand.

We empirically examine these questions using data from a popular mobile dating app in the United States during the 2014-15 time-frame. Users in the app are matched based on games where they rank members of the opposite sex. Each game consists of four men and four women in a virtual room, where each player has ninety seconds to rank-order members of the opposite sex from one to four, with one indicating the most preferred partner and four the least (see Figure 1). (Throughout the paper, we use the term preference-ranking, which is reverse of ranking, to indicate users' ordered preferences to simplify exposition.[1]) The platform then uses these preference-rankings as inputs into a Stable Match Algorithm and matches each player in the room with a member of the opposite sex. After the game ends, users can initiate contact with their matched players and chat with them (if their matched partner reciprocates).

A key piece of information shown to users during and after the game is a star-rating for each member of the opposite sex (ranging from one to three stars). A user's star-rating is a cumulative measure of all the preference-rankings that s/he received in the past. So users who received higher past preference-rankings are shown with higher stars. Stars are thus a salient and visible indicator of a user's popularity on the platform. At the same time, they do not contain any extra information on the unobserved quality of the user since they are not based on his/her contact/engagement with previous players. Thus, they do not help resolve asymmetric information about the user's quality as a date (unlike star-ratings based on purchase/experience in e-commerce settings).

Our analysis consists of two major components, which mirror our two broad research questions. To answer our first research question, we quantify the causal impact of a user's star-rating on three demand measures: (1) preference-rankings received during a game, (2) likelihood of receiving a first

---

[1]Rank of one denotes a preference-ranking of four, rank of two indicates a preference-ranking of three, and so on.
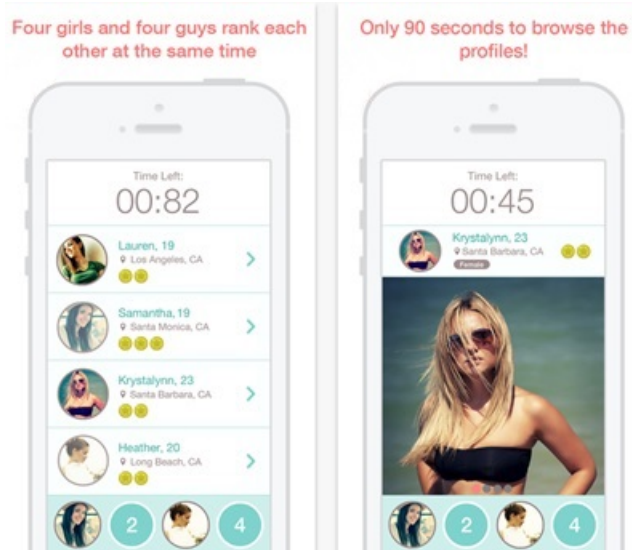
Figure 1: Screen shot of the app during a game (from the perspective of a male user). Players indicate their rank-ordered preference for the players from the opposite sex by dragging their profile pictures into the circles labeled one through four at the bottom of the app. In this example, the focal player has picked his first and third choices, and is yet to decide his second and fourth choices.

message from the matched partner after the game, (3) likelihood of receiving a reply to a message sent after the game. The main challenge here is that users who received high preference-rankings in the past (and hence have higher stars now) are also likely to receive higher preference-rankings now – not necessarily because of their star-rating, but due to their inherent attractiveness, which may be unobservable to the researcher (e.g., great bio descriptions, fun-loving pictures). This can give rise to an upward bias in our estimates of the effect of star-ratings if we use naive estimation strategies. To overcome this challenge, we leverage the fact that a user's star-rating is not static; rather it changes over the course of our observation period as a function of her/his rankings in the previous games. Specifically, we model the first demand measure using a fixed-effects ordered logit model, and the latter two using fixed-effects binary logit models. In all these models, we allow user-specific unobservables (i.e., the fixed-effects) to be arbitrarily correlated with star-ratings.

We find that, everything else being constant, three-star users receive lower preference-rankings compared to two-star users during the game, i.e., popularity has a *negative* effect on preference-rankings. We also find that ignoring endogeneity problems would lead us to draw the exact opposite conclusion. Interestingly, the effect of star-rating is different in after-game outcomes. In particular, three-star users are more likely to receive both first messages and replies after the game. Thus, users respond differently to popularity information at different stages of the matching process.

Next, we focus on our second research question, regarding the source of the popularity effect.

Here, we leverage the differences in the risk of rejection across the observed demand measures and show that the negative effect of star-ratings during the game can be attributed to strategic shading. When a user is ranking a potential partner during the game, she has uncertainty on whether that person is actually interested in a conversation/date. Indeed, even conditional on matching, post-match rejection is very common (i.e., the matched partner does not initiate or respond to messages).[2] In contrast, in the reply message case, the user has already received a message from her/his match and is considering whether to reply or not. Here, rejection is not a concern at all since the other party has already expressed interest. Using the fact that the effect of star-rating in the reply case is strictly positive, we show that the negative effect of star-rating during the game can stem from strategic shading, which can be attributed to post-match rejection concerns. Further, we show that the negative effect of star-ratings on preference-rankings is mainly driven by users who have not had many successful conversations in the past. Since users with a history of being rejected are more likely to have rejection concerns, this finding corroborates our strategic shading hypothesis.

In sum, our paper makes three contributions to the literature. First, we document negative returns to popularity information in two-sided dating markets. Past empirical research has mainly documented positive returns to the revelation of popularity information in e-commerce markets. We show that those results do not always translate to two-sided matching markets where there are rejection concerns (even when the matching is centralized). Second, we are the first to provide empirical evidence for strategic shading in dating markets and directly link it to rejection concerns. While strategic shading has been discussed in the literature, none of the earlier papers have been able to causally identify it. Finally, centralized matching markets have long been proposed as a panacea to the problems that plague decentralized markets. Our findings suggest that centralized matching markets can still lead to strategic behavior if users have post-match rejection concerns. Hence, markets where it is not feasible to enforce binding matches (e.g., dating markets, freelance markets) may suffer from strategic behavior even with centralized matching.

## 2  Related Literature

First, our paper relates to a large stream of literature that has established that popularity information has a positive effect on demand/sales of products and services in a variety of e-commerce settings, such as the music industry (Salganik et al., 2006; Dewan et al., 2017), books (Sorensen, 2007), restaurants (Cai et al., 2009), software downloads (Duan et al., 2009), kidney transplant market (Zhang, 2010), movies (Moretti, 2011), digital cameras on Amazon (Chen et al., 2011), and the

---

[2]Note that even though our setting constitutes a centralized market, there remain significant post-match rejection concerns. In this aspect, our setting is unlike standard centralized matches where matches are binding, e.g., residents and hospitals in NRMP program (Roth and Sotomayor, 1990).

wedding services market (Tucker and Zhang, 2011).[3] These studies have identified three mechanisms for this positive effect: (1) observational learning or quality inference based on others' actions (e.g. purchase statistics), (2) salience effect or awareness of alternative choices, and (3) network effect or increase in value of a product/service as its user base expands. In this paper, we provide the first negative effect of popularity information on demand in an online marketplace, and in a previously unstudied context – a two-sided dating market. We also present evidence for a new mechanism that can moderate the effect of popularity information – strategic shading due to rejection concerns.

Second, our paper relates to the literature on the empirical measurement of mate preferences in marriage and dating markets. Early work in this stream mostly used data on observed marriages to estimate population-level mate preferences under the assumption of no search frictions (Wong, 2003; Choo and Siow, 2006). More recently, researchers have been able to access data from speed-dating and online dating platforms. In these settings, search frictions are minimal and researchers have direct visibility into the search process employed by users and their preferences. This has led to a stream of literature that attempts to directly estimate users' preferences for mates along a variety of dimensions, e.g., age, income, race, physical attractiveness (Kurzban and Weeden, 2005; Fisman et al., 2006, 2008; Eastwick and Finkel, 2008; Hitsch et al., 2010a,b; Shaw Taylor et al., 2011; Bapna et al., 2016; Lee, 2016a; Jung et al., 2019).

An important concern when measuring user preferences is the possibility of strategic behavior – users may shade down their revealed preference for appealing users (physically attractive, popular, etc.) to avoid the psychological cost of being rejected (Cameron et al., 2013). If users shade their revealed preferences, and we do not explicitly account for this in the estimation, then our estimates of user preferences will be biased. The effect of users' perceived probability of being rejected on their revealed preference has been examined by a few papers. In an early paper, Hitsch et al. (2010b) employ empirical tests and show that strategic shading is not a concern in their setting. However, their results may not hold if we have variables that directly affect the perceived risk of rejection (e.g., popularity information). We use the difference in the perceived risk of rejection across outcomes (within-game ranking behavior and post-game reply behavior), and show that users strategically shade their ranking for popular users because of rejection concerns. Fong (2020) shows that an increase in market size increases selectivity while an increase in competition decreases selectivity. However, this is conceptually different from the strategic shading that we document, where users

---

[3]A related stream of work examines the effect of WOM or online reputation on demand outcomes (Chevalier and Mayzlin, 2006; Sun, 2012; Yoganarasimhan, 2012, 2013). However, in these papers, ratings are given after the interaction between the buyer and seller, and therefore help resolve asymmetric information on the quality of the product/seller. In contrast, in our case ratings are purely measures of popularity and do not convey any information on the unobserved quality of the user.

strategically avoid popular and desirable mates because of rejection concerns, which in turn leads to negative returns to popularity signals in dating markets.

Finally, our work relates to the literature on two-sided matching markets. There are two types of two-sided matching markets: centralized and decentralized. In decentralized markets, there is no central match-maker for the matching process. Instead, each agent engages in her/his own search process, and makes/accepts offers over a period of time. It has been shown that these markets are prone to market failures that can lead to inefficient matching because of search costs, unraveling, and/or congestion (Roth, 2008; Niederle and Yariv, 2009). In particular, congestion can cause users to strategically avoid making offers to their top preferences because of rejection concerns (Roth and Xing, 1997; Che and Koh, 2016; Arnosti et al., 2019).

Centralized markets have long been proposed as the panacea to the problems plaguing decentralized matching (Roth and Sotomayor, 1990). In their seminal work, Gale and Shapley (1962) proposed an algorithm that requires users to submit rank-ordered lists of their preferences for the opposite sex, and allocates stable matches for all users. Versions of this algorithm are used today in centralized markets such as National Residency Matching Program (to match residents and hospitals), and for matching students with public schools in New York City and Boston (Roth, 2008; Abdulkadiroğlu and Sönmez, 2013). Our work contributes to the matching literature by showing that centralized markets can still lead to preference-shading and strategic behavior if agents matches are non-binding and there are non-negligible costs of being rejected.

## 3 Setting

### 3.1 Mobile Dating App

Our data come from a popular online dating iOS mobile application in the United States. The app (or platform) is targeted at a younger demographic, and those using it are often looking for a fun chat rather than long-term dating/marriage partners. To join and use the app, users need a Facebook ID. When the user first logs in to the app (using his/her Facebook ID), the user's name, gender, age, education and employment information, and Facebook profile picture are automatically imported from his/her Facebook account into the user's dating profile in the app. Users cannot change this information in their dating profile directly.[4] However, they can upload up to five more pictures, and add a short bio to their profile. Further, the app has access to a user's real-time geographic location (based on the GPS in the mobile device) when the user is actively using the app.

The app requires users to participate in a structured matching game, which is described in detail below. Users cannot directly access or browse other users' profiles through the app; the only way to

---

[4]The app did not update this information (from Facebook) during our observation period.

6

use the app is to play the ranking game described in §3.2.

## 3.2 Description of the Game

### 3.2.1 Game Assignment

Initiation and completion of a game requires the live participation of four men and four women. When a user logs in to the app and decides to play a game, s/he is assigned to a game-room by the platform. Among the available players, only two criteria are used by the platform to assign players to games – proximity in geographic location and age. The exact algorithm is as follows: the geographic location of the first player assigned to a game-room is set as the initial center point of that game; the next player is then assigned to that game if he/she is within 500 miles of this center point. The center point is then updated as the average location of the first two players. The third player assigned to the game has to be within 500 miles of the new center point and after s/he is assigned to the game, the geographic center is again updated. This continues until four men and four women have been added to the game. Similarly, the platform ensures that the age gap between any two members in a game is no more than six years (older or younger). In the data, we find that this constraint is trivially satisfied because a vast majority of players belong to a small age bandwidth. Therefore, conditional on geography and age, the assignment of users to games is random.

### 3.2.2 Game Activity

When a game starts, participants can see a list of four short profiles of the members of the opposite sex. As shown in the left panel of Figure 1, these short profiles display a thumbnail version of users' profile picture, name, age, location and their star-rating. Tapping on the thumbnail leads to the full profile of the user (right panel of Figure 1), which contain a larger version of the profile picture (and possibly additional photos) and other information, such as bio, education, and employment. Each user then indicates his/her rank-ordered preference for the four members of the opposite sex. All users have exactly 90 seconds from the start of the game to finalize their rank-orderings.[5]

Two points are worth noting. First, players do not know the identities and attributes of the other members of their own sex in the game, i.e., men (women) do not know which other men (women) are in the same game. Second, players' actions are simultaneous and private, i.e., each user only has visibility into his/her own actions and at no point is the rank-ordering of the other players revealed to them (though they may be able to make some inferences after the game based on their match assignments). Hence, while choosing their rank-orderings, they cannot use information on other players' preferences to make their own choices.

---

[5]If one or more users leave the game or do not complete their rank-ordering, the game is deemed incomplete and no matches are assigned. In our data, we see a very high rate (over 97%) of game completions.

Figure 2: Screen shots of the application before and after a game.

### 3.2.3 Match Allocation

The platform uses the rank-ordered preferences of all players in a game to derive a set of "stable matches", where the concept of stability is based on the canonical Stable Marriage Problem (SMP): "Given $n$ men and $n$ women, where each person has ranked all members of the opposite sex in order of preference, match the men and women such that there are no two people of opposite sex who would *both* prefer each other over their current partners," (Gale and Shapley, 1962).

There are a few noteworthy points about the SMP. First, for any combination of preferences, there always exists at least one solution/stable match to a SMP. Second, the SMP can have more than one solution even for a relatively small number of players and the optimality of these solutions can depend on the algorithm used. For instance, Gale and Shapley (1962) show that a "Men-proposing Gale-Shapley Deferred Acceptance algorithm" is men-optimal, i.e., none of the men can do better under a different algorithm.[6] In our case, the platform first calculates all possible solutions for a game by considering all combinations of matches and checking for stability. If a game has a unique solution, then the platform allocates matches based on this solution. If there are two or more solutions, the solution that offers the highest average match is chosen. Thus, the platform does not optimize for either men or women, but instead tries to pick the best globally optimal solution.[7]

---

[6]Similarly, a women-proposing Gale-Shapley Deferred Acceptance algorithm is women-optimal, i.e., none of the women can do better using a different algorithm.

[7]The average match of a solution is calculated as follows: take the ranking that each player gave the person s/he is paired with in a stable match and sum this number over all players. In case there are multiple solutions with the same average match, ties are broken randomly.

The entire matching process takes less than a second and users can see the match assigned to them as well as all the other matches allocated in the room (see the right panel of Figure 2).

### 3.2.4 Post-Game Actions

After they have been assigned a match, users have the option to send a message to their match; see Figure 2 on the right panel. Users can also play another game, go to the home page or close the app. However, if they choose any of the latter actions without first sending a message to their matched partner, they lose the option to communicate with them in the future (unless the matched person sends them a message, in which case they can respond to it and continue the conversation). Once users initiate or receive a message, the message stays in their Inbox, and they can continue to communicate with that person in the future, if they choose to. Finally, note that users cannot start or receive any communication from other players in the game with whom they have not been matched.

## 4  Data

Our data comprises of 94,386 games played by 24,653 unique users during the ten month period from September 15th 2014 to July 15th 2015. The data can be categorized into three groups: (1) User-level data, (2) User-User level data, and (3) User-Game level data. We now describe the variables in each of these categories and present some summary statistics on them.

### 4.1  User-level Data

We first describe the time-invariant attributes associated with each user $i$.

- $gender_i$: A dummy variable indicating user $i$'s gender; is 1 for men and 0 for women.
- $age_i$: User $i$'s age.[8]
- $bio_i$: The length of user $i$'s bio in his/her profile (i.e., number of words).
- $education_i$: Categorical variable that denotes the user $i$'s highest education level (either earned or working towards), where $1 =$ High-school, $2 =$ College, and $3 =$ Graduate school.
- $employment_i$: Number of positions/companies mentioned in user $i$'s profile.
- $initial\_game_i$: Total number of games played by user $i$ before the data collection period.
- $total\_game_i$: Total number of games played by user $i$ during the data collection period.
- $num\_pic_i$: Number of uploaded pictures in the dating profile.

In addition, we also have access to the profile picture of user $i$. To obtain a measure of the physical attractiveness of a user's profile picture, we conducted a survey. We asked 384 heterosexual subjects in a research lab to rate the profile pictures of the opposite sex (men rated women and vice-versa),

---

[8]Age (calculated based on the user's Facebook birthday) changes for 26.87% users (6,378 users) during our sample period. However, this is a deterministic change i.e., age can increase only by one in the 10-month window.

| Variables | Mean | Std. Dev | 25th | 50th | 75th | (Min, Max) | Size |
|---|---|---|---|---|---|---|---|
| $age_i$ | 21.53 | 5.41 | 19 | 21 | 22 | (13, 109) | 22024 |
| $bio_i$ | 67.04 | 275.58 | 0 | 0 | 63 | (0, 29519) | 22948 |
| $employment_i$ | 1.29 | 1.60 | 0 | 1 | 2 | (0, 68) | 24653 |
| $initial\_game_i$ | 59.50 | 64.32 | 0 | 48 | 90 | (0, 2146) | 24653 |
| $total\_game_i$ | 31.27 | 37.90 | 6 | 18 | 45 | (1,1069) | 24653 |
| $num\_pic_i$ | 4.26 | 1.01 | 4 | 4 | 4 | (0, 6) | 22669 |
| $pic\_score_i$ | 0.00 | 0.68 | -0.52 | -0.09 | 0.43 | (-2.88, 3.29) | 17739 |
| $gender_i$ | (0) female: 42.45% | | (1) male: 57.55% | | | | 24653 |
| $education_i$ | (1) high-school: 19.24% | | (2) college: 78.12% | | (3) graduate: 2.64% | | 21604 |

Table 1: Summary statistics of user-level data.

on a scale of 1 to 7, with 1 being "not at all attractive" and 7 being "very attractive". The subjects were undergraduate students at a large state university in the west coast, with an equal fraction of male and female, and their ages ranged between 18-25 (with a median age of 21). This demographic distribution closely mimics the age and gender distribution of the app users.

During the lab study, each subject rated 100 randomly pictures in approximately 20 minutes. On average, each profile picture was rated by five subjects to ensure that the ratings captured average appeal rather than idiosyncratic preferences of a specific subject. It is possible that some subjects give consistently higher or lower ratings than other subjects. We therefore standardized each rating by subtracting the mean rating given by the subject and dividing by the standard deviation of the subject's ratings, as advocated by Biddle and Hamermesh (1998). We then take the average of all the standardized ratings that user $i$'s picture received in our study and denote it as:

- $pic\_score_i$: The average physical attractiveness score of user $i$'s profile picture.

Finally, because of constraints in subject-pool time, we could only obtain the picture-scores for a random sub-sample of users instead of the full pool of users; thus we have picture-score information for 17,753 of the 24,653 unique users.

The summary statistics of all the user-level variables are shown in Table 1. Of the 24,653 users, 14,189 (57.55%) are male and 10,464 (42.45%) are female. The median user is 21 years old, has no bio written on her/his profile, has/is working towards a college degree, and one employment-related information is listed on her/his profile. S/he has played 48 games before the data collection period and plays 18 games during it. However, there is quite a bit of variation across users in the extent of activity, with some users playing over 1000 games during our observation period.

Finally, note that the above user-specific variables are treated as time invariant because users lack the ability to change most of their profile information after it is first imported from their Facebook profile (name, gender, age, education and employment information, and profile picture). The two pieces of information that users can change in the app are – (1) the five extra pictures that they are

allowed to upload (in addition to the profile picture), and (2) their short bio. However, we do not believe that this was a frequent occurrence for the reasons discussed in §6.3.2.

## 4.2   User-User level data

Each game consists of eight unique users – four men and four women. For each man-woman pair in a game, we have data on the preference-ranking that they gave each other, their match outcome, and their post-match interactions. We describe these variables in detail below.

- $pref_{ijt}$: The preference-ranking that user $i$ *receives* from user $j$ in game $t$; it can take values from one to four, with four indicating the highest preference and one the lowest.

  Users rank members of the opposite sex from one through four (as shown in Figure 1), with a rank of one indicating their highest preference and four indicating the lowest. We convert these rank orderings to preference-rankings, such that rank of one denotes a preference-ranking of four, rank of two indicates a preference-ranking of three, and so on. The transformed variable $pref$ is easier to interpret because higher values of this variable correspond to more preference.

- $match_{ijt}$: A dummy variable indicating whether user $i$ is matched with player $j$ in game $t$. In each game, all players are uniquely matched with one other player from the opposite sex. So for woman (man) $i$ in a game, this variable is equal to one for only one man (woman).

- $first_{ijt}$: A dummy variable indicating whether user $i$ receives the first message from the matched partner (denoted by $j$) after game $t$. Since users cannot communicate with players that they have not been matched with, by default, this variable is zero if $match_{ijt} = 0$.

- $reply_{ijt}$: A dummy variable indicating whether user $i$ receives a reply message from the matched partner $j$ after game $t$, conditional on user $i$ initiating the first message. By default, this variable is zero if $first_{jit} = 0$.

The summary statistics of these variables are shown in Table 2. The sample sizes of $pref$ and $match$ reflects the fact that there are 32 observations per game.[9] The distributions of $pref$ and $match$ are determined by the game structure, and their summary statistics are as expected. The sample size of $first_{ijt}$ reflects the fact that there are eight users matched with each other, and each of them can potentially initiate the first message. It is worth noting that the mean of $first_{ijt}$ is around 0.05 (of the 713,014 matches, only 39,377 messages were initiated). The observed number of first messages (39,377) defines the sample size of $reply_{ijt}$. The mean of $reply_{ijt}$ is around 0.08 (among 39,377 first messages only 3380 receive a reply). Interestingly, 76% of the conversations are

---

[9]Eight users participate in each game and each user receives four preference-rankings from players of the opposite sex. So we have a total of $8 \times 4 = 32$ preference-rankings per game. Also, since each user can be matched with only one of the four potential mates, $match_{ijt}$ becomes one once and zero thrice. Thus, for each game we have $8 \times 1 + 8 \times 3 = 32$ data points for $match_{ijt}$. Therefore, the size of $pref_{ijt}$ and $match_{ijt}$ should be the number of games (94,386) $\times$ 32 = 3,020,652. However, there were some discrepancies in the data for 42 users, so we exclude them from our analysis.

| Variables | Mean | Std. Dev | 25th | 50th | 75th | (Min, Max) | Size |
|---|---|---|---|---|---|---|---|
| $pref_{ijt}$ | 2.5 | 1.12 | 2 | 3 | 4 | (1, 4) | 3008560 |
| $match_{ijt}$ | 0.25 | 0.43 | 0 | 0 | 0.5 | (0, 1) | 3008560 |
| $first_{ijt}$ | 0.05 | 0.23 | 0 | 0 | 0 | (0, 1) | 713014 |
| $reply_{ijt}$ | 0.08 | 0.28 | 0 | 0 | 0 | (0, 1) | 39377 |

Table 2: Summary statistics of user-user level data.

initiated by men, which indicates that women are less likely to approach men after being matched. Further, men receive a reply to their messages 5% of the times, and women receive a reply 20% of the times. These statistics are consistent with previous research on online dating, which find that men are more likely to initiate contact and respond to emails/messages, compared to women (Kurzban and Weeden, 2005; Fisman et al., 2006; Hitsch et al., 2010b).

### 4.3 User-Game level data

We now describe user-game level variables, i.e., user-specific data that varies with each game.

- $match\_level_{it}$: An integer variable that denotes how much user $i$ prefers his match in game $t$.

$$match\_level_{it} = pref_{jit} \quad \text{where} \quad match_{ijt} = 1 \tag{1}$$

- $total\_game_{it}$: Total number of games that user $i$ has played before game $t$. This is updated by one after each game played by user $i$.

- $star_{it}$: Indicates the user's star-rating in game $t$; see Figure 1 for an example. User's star-rating is updated in real time after each game and is calculated as follows:

$$star_{it} = \begin{cases} 1, & \text{if } 1 \leq popularity_{it} < 2 \\ 2, & \text{if } 2 \leq popularity_{it} < 3 \\ 3, & \text{if } 3 \leq popularity_{it} \leq 4, \end{cases} \tag{2}$$

where popularity is defined as the average of the preference-rankings that user $i$ has received *before* the $t^{th}$ game, such that $popularity_{it} = \frac{\sum_{q=1}^{total\_game_{it}} \sum_{j=1}^{4} pref_{ijq}}{4 \times total\_game_{it}}$. While users know their own star-rating before each game, and members of the opposite sex in the game room can observe a user's star-rating, the platform does not reveal a user's popularity scores to her/him or to anyone else in the platform.

Figure 3 illustrates the relationship defined in Equation (2). Intuitively, an individual's star-rating captures how popular or sought after s/he was in her/his past games. Three-star users, on average, are those who were among the top two choices of other players. Two-star players are those who, on average, were the second or third choice of players in the past . Finally, one-star
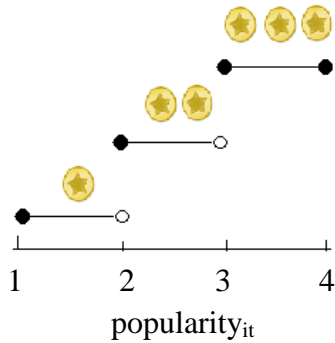
12

Figure 3: Pictorial representation of the star-rating rule.

| Variables | Mean | Std. Dev | 25th | 50th | 75th | (Min, Max) | Size |
|---|---|---|---|---|---|---|---|
| $match\_level_{it}$ | 3.19 | 0.95 | 3 | 3 | 4 | (1, 4) | 752140 |
| $total\_game_{it}$ | 74.75 | 74.25 | 29 | 59 | 97 | (0, 2194) | 752140 |
| $star_{it}$ | 2.00 | 0.10 | 2 | 2 | 2 | (1,3) | 745037 |

Table 3: Summary statistics of user-game level variables.

players, on average, are those who were the third or fourth choice of others in the past. Thus, there is a clear monotonic relationship between past popularity and current star-rating.

The summary statistics of all the user-game level variables are shown in Table 3. There are a few interesting points of note. First, the average $match\_level$ is 3.19, which implies most users get matched with their first or second top choices, on average.[10] We also find that the median of $total\_game_{it}$ is 59, i.e., most users have played a good number of games before a median game in the observation period. Moreover, we see that users are shown with a two-star rating on average.

Finally, we examine the within-user variation in star-ratings. Of the 24,653 users in our data, 85.83% (21,159 users) are shown with two stars in all their games, i.e., they never experience a star change. However, 3,494 users experience a star change. Of these, 36.83% (1,287 users) were shown with a minimum of one star and a maximum of two stars, and 62.54% (2,185 users) were shown with a minimum of two and maximum of three stars. Very few users (22) experienced a minimum of one star and a maximum of three stars. In sum, while a majority of users never experience a star change, there is a sufficiently large portion that goes through at least one star change.

---

[10]If user $i$ is matched with her most preferred player in game $t$, then in that game, $match\_level_{it} = 4$, and if she is matched with her least preferred player, $match\_level_{it} = 1$. If preferences were purely vertical, i.e., if all the men in a game had the same rank-ordering for women (and vice-versa), and users report their preferences without strategic shading, then the mean $match\_level$ would be 2.5. Instead, if preferences were purely horizontal, then the mean $match\_level$ would be 4. The fact that the average of $match\_level$ is 3.19, suggests that users' preference-rankings are a combination of vertical attribute, horizontal attributes, and other factors such as strategic shading.

# 5   Descriptive Analysis

We now examine the relationship between a user's star-rating and three measures of her/his demand – preference-rankings received during the game, and whether s/he receives a first message, or reply message after the game – using simple model-free analyses. In this section, we focus on users who experienced at least one change in their star-rating during our observation period.

The relationship between a user's star-rating in a given game and the average preference-ranking that s/he receives in that game is illustrated in Figure 4. The solid increasing line shows the relationship between the average preference-rankings received for all user-game observations calculated for each star-rating.[11] We see that in observations where users have higher star-ratings, they also receive higher preference-rankings. However, there is an obvious issue of correlated unobservables here, i.e., users with higher star-ratings are likely to be more attractive on other unobserved dimensions (e.g., physical attractiveness) as well. To examine if this conjecture is true, we plot the average of users' $pic\_score$ for each star-rating. As shown in Figure 5, users with higher star-ratings also have higher physical attractiveness score, on average. Thus, the effects shown by the solid line in Figure 4 cannot be interpreted as causal.
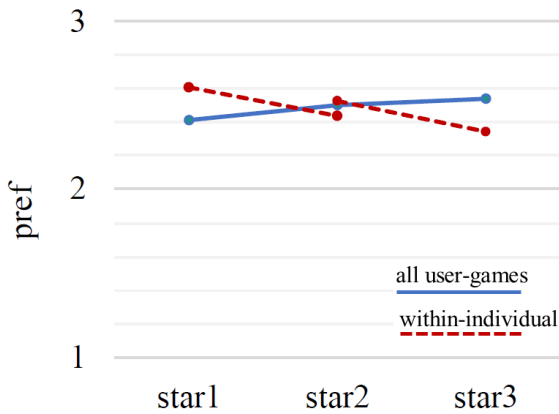
Figure 4:  Relationship between star-ratings and average preference-rankings received.
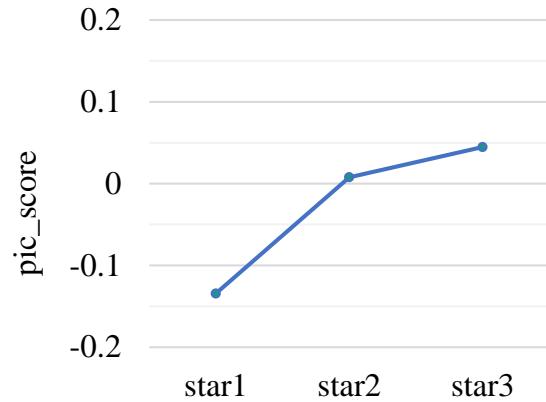
Figure 5:  Relationship between star-ratings and average physical attractiveness score.

One way to cleanly capture the effect of star-ratings is to look at the effect of star-ratings *within* an individual, i.e., if we compare preference-rankings received by the same individual when s/he is shown with different star-ratings, then our comparisons are less likely to be subject to endogeneity concerns. Consider an individual who was shown with a minimum of one star and a maximum of two stars and calculate two averages: (1) the average of preference-rankings received in games where s/he is shown with one star, and (2) the average of preference-rankings received in games

---

[11]For example, the average preference-ranking for the data point at $star1$ on the solid line is $\frac{\sum_i \sum_t \sum_j (pref_{ijt} \mid star_{it}=1)}{4 \times \sum_i \sum_t I(star_{it}=1)}$.

where s/he is shown with two stars. We then perform an analogous exercise for users who were shown with a minimum of two stars and a maximum of three stars. These two comparisons are shown using dashed lines in Figure 4. As we can see, on average, the same set of users receive higher preference-rankings when they are shown with one star compared to two stars. Moreover, on average, the same set of users receive higher preference-rankings when they are shown with two stars compared to three stars. This implies that higher star-ratings leads to lower preference-rankings, i.e., users avoid those with higher stars! Note that the direction of the effect of star-rating on preference-rankings in solid line and dashed lines in Figure 4 are exactly opposite. This discrepancy implies that controlling for the endogeneity between star-ratings and unobserved factors that affect user attractiveness is essential to deriving the causal impact of star-ratings in our setting.

Next, we conduct an analogous analysis on the relationship between a user's star-rating and the likelihood of receiving the first message and receiving a reply if s/he initiates a message, and present the results in Figures 6 and 7. First, we see that observations where users have higher star-ratings are more likely to receive both first messages and replies (solid lines in the figures). Second, for the within-user analysis, we see that, on average, the same set of users are more likely to receive first messages when they are shown with one star compared to two stars (though this is not the case when we compare two and three stars). In the case of *reply*, the same set of users are more likely to get a reply when shown with higher star-ratings.
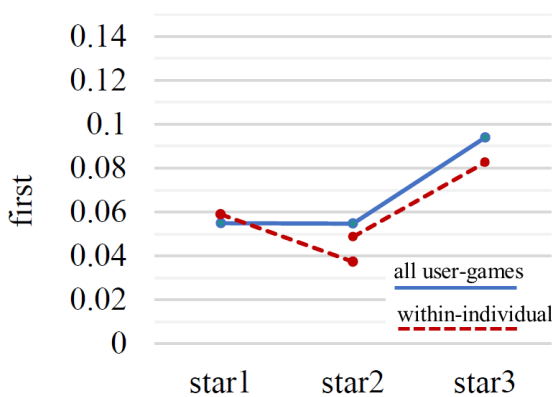


Figure 6: Relationship between star-ratings and the average likelihood of receiving the first message.
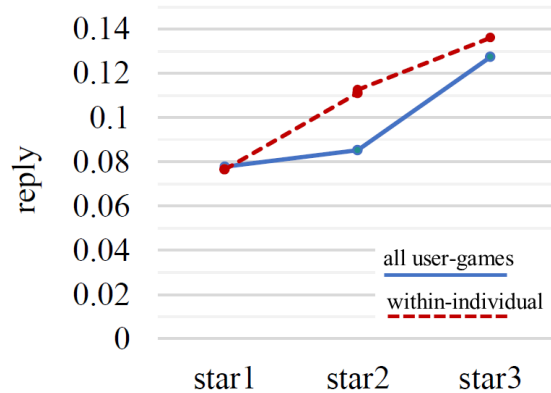


Figure 7: Relationship between star-ratings and the average likelihood of receiving a reply message.

In sum, when we look at the simple correlation between star-ratings and revealed preferences, we always see a positive effect. However, when we look at within-individual comparisons, the findings are quite different. Interestingly, the effect of higher star-ratings seems to be negative for preference-rankings during the game, partially-negative for initiating communication after the game (first message), and positive when it comes to replying to messages after the game. In the rest of

the paper, we focus on deriving the unbiased causal effect of star-ratings on these three revealed preference measures, and exploring the mechanisms driving these effects.

# 6 Effect of Star-ratings on Preference-Rankings

In this section, we formalize the causal impact of a user's star-rating on the preference-rankings that s/he receives during the game. In §6.1 and §6.2, we present the model specification and estimation. In §6.3, we discuss the identification and we discuss our findings in §6.4.

## 6.1 Model Specification

The outcome variable of interest here is $pref_{ijt}$, which denotes the preference-ranking that user $i$ *receives* from $j$ during game $t$. Note that $pref$ is an ordinal integer value going from 1 to 4, with one representing the lowest preference-ranking and four indicating the highest preference-ranking. Therefore, we use an ordered logit model[12] that relates the observed outcome variable $pref_{ijt}$ to a *latent* variable $pref_{ijt}^*$ where:

$$pref_{ijt}^* = \beta_1 star1_{it} + \beta_2 star3_{it} + \gamma z_i + \eta_i + \epsilon_{ijt}, \tag{3}$$

The latent variable $pref_{ijt}^*$ is modeled as a linear function of:
- $star1_{it}$, $star3_{it}$ – indicator variables for the star-rating of user $i$ in game $t$, where $star2_{it}$ is considered as the base.
- $z_i$ – set of user-specific observables that can affect $j$'s ranking of $i$, e.g., gender of $i$.
- $\eta_i$ – set of unobservable (to the researcher) characteristics of user $i$ that is visible to $j$ and affects $j$'s ranking of $i$. These could include the aspects of user $i$'s physical attractiveness not captured in our lab study (e.g., other photos of the user), details in her/his bio description, employment details, her geographic location, etc.
- $\epsilon_{ijt}$ – factors uncorrelated to the star-rating of user $i$ that can affect the preference-ranking s/he receives from $j$ in game $t$. We assume that $\epsilon_{ijt}$s have a logistic cumulative distribution. Three key sets of variables are subsumed in $\epsilon_{ijt}$. First, it includes $j$'s attributes (both observable $z_j$ and unobservable $\eta_j$) since there is no correlation between $j$ and $i$'s attributes. Second, it includes all the attributes of the other three players of $i$'s gender who $i$ is being compared with, in game $t$.[13] The reason neither of the above two sets of variables affect our inference on star-ratings is because the app adds users into a game randomly. Thus, there is no correlation between the

---

[12] We discuss other modeling frameworks such as rank-ordered logit or Regression Discontinuity Design (RDD) in Appendix §A and explain why they are not appropriate for our setting.

[13] Including the observations of all the competitors in a game can create within-game correlation in our analysis. We address this issue in §8.3.

attributes of users within a game.[14] Third, it can capture idiosyncratic factors that affect $j$'s ranking of $i$ within the game, e.g., $j$'s mood for going on a date with someone of $i$'s type etc. We then model the relationship between $pref_{ijt}$ and $pref_{ijt}^*$ as follows:

$$pref_{ijt} = k \quad \text{if } \mu_k < pref_{ijt}^* \leqslant \mu_{k+1} \quad \forall \quad k = 1, 2, 3, 4, \tag{4}$$

where the thresholds $\mu_k$ are strictly increasing. Further, we assume that $\mu_1 = -\infty$ and $\mu_5 = \infty$. This specification is the ordinal choice analog of a binary logit model. Thus, $pref_{ijt}$ can take four possible values, denoted by $k$. Because the error terms are drawn from a logistic distribution, we can write the cumulative probability function of $\epsilon_{ijt}$ as

$$F\left(\epsilon_{ijt} \mid X_{it}, \beta_1, \beta_2, \gamma, \eta_i, \mu_k, \mu_{k+1}\right) = \frac{1}{1 + exp(-\epsilon_{ijt})} \equiv \Lambda(\epsilon_{ijt}), \tag{5}$$

where $X_{it} = \{star1_{it}, star3_{it}, z_i\}$. Therefore, the probability of observing outcome $k$ in game $t$ for a pair of users (where user $i$ receives a rank $k$ from user $j$) can be written as:

$$
\begin{aligned}
Pr\left(pref_{ijt} = k \mid X_{it}, \beta_1, \beta_2, \gamma, \eta_i, \mu_k, \mu_{k+1}\right) &= \Lambda\left(\mu_{k+1} - \beta_1 star1_{it} - \beta_2 star3_{it} - \gamma z_i - \eta_i\right) \\
&- \Lambda\left(\mu_k - \beta_1 star1_{it} - \beta_2 star3_{it} - \gamma z_i - \eta_i\right) \quad (6)
\end{aligned}
$$

Using this model formulation, we can then write the log-likelihood of the preference-rankings observed in the data as:

$$LL(\beta_1, \beta_2, \gamma, \eta_i, \mu_k, \mu_{k+1}) = \sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{j=1}^{4} \sum_{k=1}^{4} \ln\left[Pr\left(pref_{ijt} = k \mid X_{it}, \beta_1, \beta_2, \gamma, \eta_i, \mu_k, \mu_{k+1}\right)^{I(pref_{ijt}=k)}\right],$$

where $N$ is the total number of users observed and $T_i$ is the total number of games played by user $i$. The unknown parameters in Equation (7) are $\beta_1, \beta_2, \gamma, \eta_i s, \mu_2, \mu_3, \mu_4$.

## 6.2 Estimation

We are mainly interested in estimating the effect of star-ratings ($\beta_1$ and $\beta_2$). The challenge comes from the potential correlation between $\eta_i$ and $star_{it}$, i.e., we expect that $E[star_{it} \cdot \eta_i] \neq 0$. We now discuss three estimation strategies that address this problem in varying degrees.

---

[14]In principle, because the app only adds new users within a 500 mile radius of users already in the game, the geographic locations of users in a game are correlated. However, conditional on being in the same room, there is no correlation between the location of two users, and the distance between the users is random. In other words, if we denote the geographic location of users by $g$, then we can write the location of $j$ as: $g_j$, where $g_j = g_i + \delta$, where $g_i, g_j, \delta$ are two dimensional vectors (latitude, longitude) such that $||g_j - g_i|| \leq 500$. Since we already control for user $i$'s location ($g_i$) through $\eta_i$, the remaining $\delta$ is random noise.

First, a pooled ordered logit model, which ignores the user-specific unobservables $\eta_i$. It simply involves pooling all the user-game data, ignoring the user-specific variables ($z_i$, $\eta_i$), and then maximizing the log-likelihood in Equation (7). However, it is important to recognize that the estimates from this approach will be biased in the presence of correlated unobservables.

Second, a pooled ordered logit model with control variables, that includes user-specific variables ($z_i$) to control for the correlation between $star_{it}$s and $\eta_i$. For example, controlling for users' physical attractiveness ($pic\_score_i$) may reduce the bias in estimates of $\beta_1$ and $\beta_2$. However, this method is unable to control for the correlation between $star_{it}$s and $\eta_i$.

Third, a fixed-effects ordered logit model, where we allow the user-specific unobservables $\eta_i$ to be arbitrarily correlated with the star-ratings. A naive approach to estimation with fixed-effects is to treat the $\eta_i$'s as parameters and maximize the log-likelihood in Equation (7) directly. However, such a Maximum Likelihood Estimator (MLE) is inconsistent with large $N$ and finite $T$ due to the well-known incidental parameters problem (Neyman and Scott, 1948). As a result, the estimates of $\beta_1$ and $\beta_2$ from this approach will be inconsistent too. Chamberlain (1980) provides an elegant solution to the incidental parameters problem by dichotomizing the ordered outcome variable. In §6.2.1, we describe how to apply the Chamberlain estimator to our setting, in §6.2.2 we describe how the Chamberlain estimators can be combined to form an efficient Minimum Distance estimator.

### 6.2.1 Chamberlain's Conditional Maximum Likelihood Estimator

The ordered outcome variable $pref_{ijt}$ can take $K = 4$ possible integer values, $\{1,2,3,4\}$. Therefore, we can transform the random variable $pref_{ijt}$ into $K - 1 = 3$ possible binary variables $pref_{ijt}^k$ where:

$$pref_{ijt}^k = I(pref_{ijt} \geq k), \quad \text{where} \quad k = 2, 3, 4. \tag{7}$$

For example, the binary variable $pref_{ijt}^4$ indicates whether user $i$ received a preference-ranking of 4 from user $j$ in game $t$, or not. Similarly, the binary variable $pref_{ijt}^3$ indicates whether user $i$ receives a preference-ranking of 3 or higher (i.e., 3 or 4) from user $j$ in game $t$, or not. We can specify Chamberlain's Conditional Maximum Likelihood (CML) estimator on each of these transformed binary variables. For each $k$, $pref_{ijt}^k$ is a binary logit variable such that:

$$Pr(pref_{ijt}^k = 1 \mid X_{it}, \beta_1, \beta_2, \gamma, \eta_i, \mu_k) = 1 - \Lambda(\mu_k - \beta_1 star1_{it} - \beta_2 star3_{it} - \gamma z_i - \eta_i) \tag{8}$$

We denote $pref_i^k$ as the entire history of preference-rankings at level $k$ received by user $i$ over time, i.e. $pref_i^k = \{pref_{i11}^k, pref_{i21}^k, pref_{i31}^k, pref_{i41}^k, ..., pref_{i1T_i}^k, pref_{i2T_i}^k, pref_{i3T_i}^k, pref_{i4T_i}^k\}$. Further, let $s_i^k$ be the sum of all the binary transformed preference-rankings at level $k$ received by user $i$ such that $s_i^k = \sum_{t=1}^{T_i} \sum_{j=1}^{4} pref_{ijt}^k$. In other words, $s_i^k$ shows the count of ones in the set of $pref_i^k$. Let

$B_i^k$ be the set of all possible vectors of length $4 \times T_i$ with $s_i^k$ elements equal to 1, and $4 \times T_i - s_i^k$ elements equal to 0.[15] That is:

$$B_i^k = \{d \in \{0,1\}^{4 \times T_i} \mid \sum_{t=1}^{T_i} \sum_{j=1}^{4} d_{jt} = s_i^k\} \tag{9}$$

Now, we can write the conditional probability of $pref_i^k$ given $s_i^k$ as:

$$Pr\left(pref_i^k \mid star1_{it}, star3_{it}, s_i^k, \beta_1, \beta_2\right) = \frac{exp\left(pref_i^k \cdot (\beta_1 star1_{it} + \beta_2 star3_{it})\right)}{\sum_{d \in B_i^k} exp\left(d \cdot (\beta_1 star1_{it} + \beta_2 star3_{it})\right)} \tag{10}$$

A key observation is that this conditional probability does not depend on $\eta_i$'s (or the thresholds $\mu_k$'s or $z_i$'s), i.e., $s_i^k$ is a sufficient statistic for $\eta_i$. Thus, we can now specify a Conditional Log-Likelihood that is independent of $\eta_i$s and $\mu_k$s as shown below:

$$CLL(\beta_1^k, \beta_2^k) = \sum_{i=1}^{N} \sum_{t=1}^{T_i} \ln\left[Pr(pref_i^k \mid star1_{it}, star3_{it}, s_i^k, \beta_1^k, \beta_2^k)\right] \tag{11}$$

Since we can dichotomize $pref_{ijt}$ into three binary variables at each of the three cutoffs ($pref_{ijt}^4$, $pref_{ijt}^3$, and $pref_{ijt}^2$), the above CLL can be specified for each $pref_{ijt}^k$, where $k \in \{2, 3, 4\}$. Maximizing each of these CLLs gives us three separate but consistent estimates of $\beta_1, \beta_2$, which we denote as $\{\beta_1^k, \beta_2^k\}$, where $k \in \{2, 3, 4\}$. These are referred to as Chamberlain CML estimators.

However, these three estimates are inefficient because each of them only uses part of the variation in the data for identification. Intuitively, at any cut-off $k$, only the variation around $k$ is used for identification because of dichotomization; for example, the CLL for $k = 4$ only considers whether $pref_{ijt}$ is greater than or equal to 4 and ignores the variation in $pref_{ijt}$ when it is less than 4. Thus, while Chamberlain's CML estimator at each $k$ is consistent, it is not efficient because it does not exploit all the variation in data.[16]

---

[15]Note that the size of $B_i^k = \binom{4 \times T_i}{s_i^k}$. Consider user $i$ who plays two games ($T_i = 2$). For $k = 4$, we have $pref_{ijt}^4 \in \{0, 1\}$ that denotes whether user $i$ has received a preference-ranking of 4 from user $j$ or not. Now, let's consider a scenario where user $i$ receives a preference-ranking of four only in her first game and from $j_1$, i.e. , $pref_i^4 = \{1, 0, 0, 0, 0, 0, 0, 0\}$. Thus, $s_i^4 = 1$. Next, we can write $B_i^4$ or the set of all possible ways that user $i$ can get only one preference-ranking of 4 in her games by $B_i^4 = \{(1, 0, 0, 0, 0, 0, 0, 0), (0, 1, 0, 0, 0, 0, 0, 0), ..., (0, 0, 0, 0, 0, 0, 1, 0), (0, 0, 0, 0, 0, 0, 0, 1)\}$. Note that each element of $B_i^4$ is itself a vector with eight elements, because user $i$ has played two games and in each game s/he receives four preference-rankings ($4 \times 2 = 8$). We denote each element of set $B_i^4$ with vector $d$. Also, notice that the size of $B_i^4$ is eight, because $\binom{4 \times 2}{1} = 8$.

[16]For individuals who have played a large number of games (large $T_i$) and have a large number of positive values of $pref_{ijt}^k$ (large $s_i^k$), calculating all combinations of outcomes can lead to numerical overflow and computational issues.

### 6.2.2 Minimum Distance Estimator

To address the efficiency issue in Chamberlain's CML, Das and Van Soest (1999) proposed a Minimum Distance (MD) estimator that combines all the Chamberlain estimates. We now describe the application of their method to our context. Recall that we have $K - 1 = 3$ estimates for each of $\{\beta_1, \beta_2\}$: $\{\beta_1^1, \beta_2^1\}, \{\beta_1^2, \beta_2^2\}, \{\beta_1^3, \beta_2^3\}$. Since each of these three estimates are consistent, any weighted average of these estimates will be consistent too. The main idea in Das and Van Soest (1999) is to use the variance and co-variances of $K - 1$ estimators as weights and generate one efficient estimate. It thus involves solving the minimization problem:

$$\hat{\beta}^{MD} = \underset{b}{\mathrm{argmin}}(\tilde{\beta} - Mb)'var(\tilde{\beta})^{-1}(\tilde{\beta} - Mb), \tag{12}$$

where $\tilde{\beta}$ is the $6 \times 1$ matrix of Chamberlain estimators, $M$ is the matrix of 3 stacked 2-dimensional identity matrices, and $var(\tilde{\beta})$ is the variance-covariance matrix of the stacked Chamberlain estimates. In other words, we need to find $b_1$ and $b_2$ such that:

$$\underset{b_1, b_2}{\mathrm{argmin}}
\begin{pmatrix} \tilde{\beta}_1^2 - b_1 \\ \tilde{\beta}_2^2 - b_2 \\ \tilde{\beta}_1^3 - b_1 \\ \tilde{\beta}_2^3 - b_2 \\ \tilde{\beta}_1^4 - b_1 \\ \tilde{\beta}_2^4 - b_2 \end{pmatrix}'
\begin{pmatrix} var(\tilde{\beta}_1^2) & & & & & \\ cov(\tilde{\beta}_2^2, \tilde{\beta}_1^2) & var(\tilde{\beta}_2^2) & & & & \\ . & . & var(\tilde{\beta}_1^3) & & & \\ . & . & . & var(\tilde{\beta}_2^3) & & \\ . & . & . & . & var(\tilde{\beta}_1^4) & \\ cov(\tilde{\beta}_2^4, \tilde{\beta}_1^2) & . & . & . & cov(\tilde{\beta}_2^4, \tilde{\beta}_1^4) & var(\tilde{\beta}_2^4) \end{pmatrix}^{-1}
\begin{pmatrix} \tilde{\beta}_1^2 - b_1 \\ \tilde{\beta}_2^2 - b_2 \\ \tilde{\beta}_1^3 - b_1 \\ \tilde{\beta}_2^3 - b_2 \\ \tilde{\beta}_1^4 - b_1 \\ \tilde{\beta}_2^4 - b_2 \end{pmatrix}$$

The solution to the above minimization problem ($b_1$ and $b_2$) is a weighted average of the Chamberlain estimators and is equal to:

$$\hat{\beta}^{MD} = \{M'var(\tilde{\beta})^{-1}M\}^{-1}M'var(\tilde{\beta})^{-1}\tilde{\beta} \tag{13}$$

and its variance is given by $var(\hat{\beta}^{MD}) = \{M'var(\tilde{\beta})^{-1}M\}^{-1}$. We implement this MD estimator using the Stata code developed by Hole et al. (2011). For more details about this method and a comparison with other methods, see Baetschmann et al. (2015).

---

For example, if user $i$ plays 100 games ($T_i = 100$) and receives one preference-ranking of four in each game, then $s_i^4 = 100$ and $\binom{4 \times 100}{100} = 2.24e + 96$. Therefore, we limit our empirical analysis to users' first 100 games. Of the 3,494 users who experience a star change, only 352 (10%) users play more than 100 games. The consistency of the estimates is not affected if we choose a subset of games for players who have played a large number of games.

### 6.3 Identification

We start with a description of the types of variation that we need to see in the data for identification, and then explain why they can be treated as plausibly exogenous in our setting.

#### 6.3.1 Variation in the Data

We need two types of variation in the data for the identification of the $\{\beta_1^k, \beta_2^k\}$s in the CLL at each $k$ (as described in §6.2.1).

First, we need within-user variation in $star1_{it}$ and $star3_{it}$. Intuitively, this estimator takes advantage of the variation in star-ratings "within" a user for identifying the effect of star-ratings. This allows us to circumvent the problem of user-specific correlated unobservables since they remain constant for the user across time. If the same user $i$ receives lower preference-rankings when s/he is shown with three stars as opposed to two stars, that difference can be directly attributed to the change in star-rating since it is the only variable that has changed across time (assuming that the inherent attractiveness of the user remains constant over the duration of observation).

Second, we need within user variation in the outcome variable $pref_{ijt}^k$ because users with constant $pref_{ijt}^k$ do not contribute to the CLL for cut-off $k$.[17] We now illustrate this condition using an example. For $k = 4$, consider a user $i$ who has either received a preference-ranking of 4 in all her games, or never ever received a preference-ranking of 4 in any of her games. This user does not contribute to the CLL because her outcome ($pref_{ijt}^4$) is constant over time even if her/his star-rating varies over time. Thus, the only users who contribute to the identification of $\{\beta_1^k, \beta_2^k\}$ are those for whom we have across-time variation in both the outcome variable ($pref_{ijt}^k$) and the independent variables ($star1_{it}$, $star3_{it}$) at a given $k$. In the MD estimator, we combine the estimates across all $k$s. Therefore, all users who saw any variation in their outcomes ($pref_{ijt}$) and star-ratings will contribute to identification of $\{\beta_1, \beta_2\}$.

#### 6.3.2 Exogeneity of Variation in Star-ratings

While within-user variation in star-ratings and outcomes (preference-rankings) is necessary for identification, it is not sufficient. This brings us to the second condition necessary for valid inference: the within-user variation in star-ratings needs to be plausibly exogenous. We now provide arguments for why this is a reasonable assumption in our setting.

In order to be able to manipulate their star-rating in any period $t$, users need to be aware of and be able to meaningfully change their popularity score ($popularity_{it}$) by manipulating their profile information. This is not feasible for a few reasons. First, as discussed in §4.1, users lack the ability to change many aspects of their profile in response to their star-ratings. While they can add few

---

[17]Constant $pref_{ijt}^k$ means that all elements of $B_i^k$ are either zero or one.

additional pictures and/or modify their bio, both of these are not very critical since they are not shown in the main screen of a game (see Figure 2). Therefore, while a user can change these in response to their star-ratings, we do not believe that this was a frequent occurrence. Second, while users are aware of their star-rating at any given point in time ($star_{it}$), they do not observe any of the ranks that they received in the past games or their popularity score ($popularity_{it}$) at any point in time. (They are simply shown the person they are matched with after each game; the rankings that they received from other players are never revealed to them.) Moreover, users were never informed of the threshold rule used by the platform to assign the star-ratings. While users may have correctly inferred that their star-ratings are correlated with their prior rankings, they are unlikely to have inferred the exact rule. Finally, the marginal effect of the rankings received in a new game on the popularity score is vanishingly small as the number of games played increases (see Appendix B.2.2 for details). Thus, as user's gain experience, it is increasingly hard for them to move the needle on their popularity score (and their star-rating).

In sum, users lack the ability to modify the key aspects of their profile information, are unaware of the exact rule used to calculate their popularity scores and star-ratings, and have little ability to move the needle on their popularity scores in most cases. Therefore, we believe that it is reasonable to assume that the changes in a user's star-ratings are plausibly exogenous. That said, we cannot prove that users did not change their bio and/or additional picture in response to ratings, i.e., we cannot completely rule out potential confounds. This is a limitation of our observational setting.

This brings us to the question of: "where does the variation in star-ratings (or popularity scores) of a user come from?" It comes from two main sources. First, there is significant heterogeneity in players' taste for people of the opposite sex, i.e., rank-givers' preferences for people of the opposite sex is not purely vertical. So the same user often gets different preference-rankings from different users. Indeed, the average match-level in the data is 3.19, which suggests that, on average, users are matched with their first or second choices (see Table 3 and the discussion in Footnote 10 for more details). Second, the ranks that a user receives in game $t$ are in comparison to her/his competitors in that game. However, users have no control over whom they compete with in a given game and there is considerable randomness in the set of participants in a game (see details in §3.2.1). Both these factor induce variation in the preference-rankings (and star-ratings) of a user over time. Importantly, they are exogenous because a user has no control over the preferences of the opposite-sex players who are ranking her/him or the attributes of her/his competitors in a game (as described in §3.2.1).

## 6.4 Results

The results from the estimation exercise are presented in Table 4. As discussed in §6.2, we estimate three ordered logit models: (1) Model M1 – a simple specification that only includes star-ratings as

|  | (M1) | (M2) | (M3) |
|---|---|---|---|
|  | (Ordered Logit) | (Ordered Logit) | (FE Ordered Logit) |
| $star1_{it}$ | -0.14452*** | -0.12991*** | 0.02852 |
|  | (0.02315) | (0.02312) | (0.01804) |
| $star3_{it}$ | 0.06063*** | 0.06863*** | -0.05101*** |
|  | (0.01560) | (0.01893) | (0.01464) |
| Controls ($z_i$) |  | ✓ |  |
| Fixed Effects ($\eta_i$) |  |  | ✓ |
| $\mu_2$ | -1.09924*** | -1.11212*** |  |
|  | (0.00203) | (0.01197) |  |
| $\mu_3$ | -0.00053 | -0.00893 |  |
|  | (0.00188) | (0.01192) |  |
| $\mu_4$ | 1.09828*** | 1.09239*** |  |
|  | (0.00205) | (0.01192) |  |
| Individuals | 24393 | 16461 | 3494 |
| Observations | 2980148 | 2339168 | 630160 |

Standard errors (in parentheses) are clustered at the user level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Controls ($z_i$) in Model M2 include: $age_i$, $college_i$, $graduate_i$, $pic\_score_i$, $num\_pic_i$, $employment_i$, and $bio_i$.

Table 4: Ordered logit estimates of the effect of star-rating on preference-rankings received.

the independent variable, (2) Model M2 – a more elaborate model that also includes user-specific observables ($z_i$), and (3) Model M3 – a fixed-effects model using MD estimator that controls for $\eta_i$s.

In the basic ordered logit model (model M1), we see a positive and significant effect for higher star-ratings. That is, one-star users receive lower preference-rankings compared to two-star users, and two-star users receive lower preference-rankings compared to three-star users. This result is consistent with Figure 4 (solid line). Next, we estimate model M2, which controls for all the user-specific observables because a user's current star-rating is likely to be positively correlated to user-specific observables such as physical attractiveness, age, education, etc.[18] However, the direction of the results remain unchanged. Nevertheless, without explicitly controlling for the endogeneity concerns discussed earlier ($E[star_{it}.\eta_i] \neq 0$), our estimates are likely to be biased. Therefore, we now focus on the results from the fixed-effects MD estimator (model M3). Interestingly, here we find that the effect of star-rating is *negative* – a user gets worse preference-ranking when s/he is

---

[18]Note that the numbers of individuals are different in models M1 and M2. This is due to the fact that model M1 does not include any controls, whereas model M2 includes user-specific observables as controls. As summarized in Table 1, some of these control variables are missing for some users in the data. Since model M2 includes all the control variables, it only consists of observations where all the control variables are non-missing.

shown with three stars as opposed to two stars. We do not find any significant effect of one star compared to two stars. In §8, we present a battery of robustness checks to confirm the validity of these empirical findings.

The main takeaway from our findings is that popularity information has a negative effect on popular users' demand during the game. As discussed in §2, past empirical research has mainly documented positive gains to popularity information or herding effects. In our setting, there could be multiple reasons for the deviation from the standard positive results. It could be because users dislike the popular users. Or, they may like popular users but avoid them due to rejection concerns: rank-givers may be concerned that popular users are hard to get (at both the match and post-match conversation stage), and therefore shade their preferences for popular users to avoid rejection. In §9, we formalize the discussion of the mechanism behind the negative effect of popularity information, and rule out alternative mechanisms.

In sum, our findings suggest that researchers and managers need to understand the behavioral underpinnings of the mechanism through which popularity information operates in a given market instead of assuming positive effects based on prior work.

## 7 Effect of Star-ratings on Messaging Behavior

In this section, we examine the causal impact of a user's star-rating on her likelihood of receiving messages. We focus on two variables: (1) $first_{ijt}$: a dummy variable indicating whether user $i$ receives a first message from her match $j$ after game $t$, and (2) $reply_{ijt}$: a dummy variable indicating whether user $i$ receives a reply message from player $j$ after game $t$, conditional on user $i$ initiating the first message. We present the model and estimation in §7.1 and discuss the results in §7.2.

### 7.1 Model and Estimation

The outcome variables $first$ and $reply$ are binary. Hence, we consider logit formulations that relate them to latent variables $first_{ijt}^*$ and $reply_{ijt}^*$ as follows:

$$first_{ijt} = \begin{cases} 1, & \text{if } first_{ijt}^* > 0 \\ 0, & \text{else} \end{cases} \qquad reply_{ijt} = \begin{cases} 1, & \text{if } reply_{ijt}^* > 0 \\ 0, & \text{else} \end{cases} \tag{14}$$

These latent variables are defined as:

$$first_{ijt}^* = \beta_1^f star1_{it} + \beta_2^f star3_{it} + \gamma^f z_i + \eta_i^f + \epsilon_{ijt}^f, \tag{15}$$

$$reply_{ijt}^* = \beta_1^r star1_{it} + \beta_2^r star3_{it} + \gamma^r z_i + \eta_i^r + \epsilon_{ijt}^r, \tag{16}$$

24

where the interpretations of $\{\beta_1^f, \beta_2^f, \gamma^f, \eta_i^f, \epsilon_{ijt}^f\}$ and $\{\beta_1^r, \beta_2^r, \gamma^r, \eta_i^r, \epsilon_{ijt}^r\}$ are similar to that in §6.1. Further, following the same arguments, we allow for $\eta_i^f$ and $\eta_i^r$ to be arbitarily correlated to $star1_{it}$ and $star3_{it}$. Assuming that $\epsilon_{ijt}$s are IID and drawn from a logistic distribution, the probability that user $i$ receives a first message from user $j$ (conditional on $i$ and $j$ being matched in game $t$) is:

$$Pr(first_{ijt} = 1 \mid match_{ijt} = 1, X_{it}, \eta_i^f) = \frac{exp(\beta_1^f star1_{it} + \beta_2^f star3_{it} + \gamma^f z_i + \eta_i^f)}{1 + exp(\beta_1^f star1_{it} + \beta_2^f star3_{it} + \gamma^f z_i + \eta_i^f)}$$

Similarly, the probability that user $i$ receives a reply from user $j$ (conditional on them being matched in game $t$ *and* user $i$ having initiated the first message) can be written as:

$$Pr(reply_{ijt} = 1 | match_{ijt} = 1, first_{jit} = 1, X_{it}, \eta_i^r) = \frac{exp(\beta_1^r star1_{it} + \beta_2^r star3_{it} + \gamma^r z_i + \eta_i^r)}{1 + exp(\beta_1^r star1_{it} + \beta_2^r star3_{it} + \gamma^r z_i + \eta_i^r)}$$

As in the case of the ordered logit model, we can use these probabilities to specify two Conditional Log-Likelihoods (CLLs) that are independent of $\eta_i$s and then maximize the two CLLs to derive consistent estimates of $\{\beta_1^f, \beta_2^f\}$ and $\{\beta_1^r, \beta_2^r\}$. Since these steps are very similar to that described in §6.2, we relegate the details to Appendix §C.

## 7.2   Results

The results for both message outcomes are shown in Table 5. We start with a discussion of first messages (shown in models M4 and M5). Model M4 is a pooled logit model that only controls for the observable attributes of the (potential) receiver, but ignores the unobservables. Model M5 is a fixed-effects logit model that accounts for the endogeneity between star-ratings and user-specific unobservables. Both models control for for the time-invariant attributes of the sender $j$, i.e., $z_j$, and $j$'s star-rating to avoid selection problems.

In model M4, we find that three-star users are more likely to receive first messages compared to two-star users. We do not find any significant effect of one star compared to two stars. However, after controlling for the endogeneity issues in model M5, we find that a user is more likely to receive first messages when s/he is shown with one or three stars as opposed to two stars. This is consistent with dashed-lines in Figure 6. These results are somewhat different from those in model M3 (that characterizes the effect of star-ratings on preference-rankings). On the one hand, the positive effect for one star suggests that rejection concerns may be at play since players may expect one-star users to be more responsive to their message. On the other hand, the positive effect of three stars suggests that players may value higher-star users more. These results can be explained by a combination of both higher utility for higher star users as well as lower rejection concerns. Next, we discuss the results from the analysis of the reply messages, which helps us tease out the mechanism better.

25

|  | First Message | | Reply Message | |
|---|---|---|---|---|
|  | (M4) | (M5) | (M6) | (M7) |
|  | (Logit) | (Logit FE) | (Logit) | (Logit FE) |
| $star1_{it}$ | 0.14989 | 0.51448*** | -0.05065 | -0.09053 |
|  | (0.13078) | (0.12034) | (0.22566) | (0.31932) |
| $star3_{it}$ | 0.63824*** | 0.73056*** | 0.46113*** | 0.40377** |
|  | (0.09112) | (0.07482) | (0.15023) | (0.18095) |
| Controls ($z_i$) | ✓ |  | ✓ |  |
| Controls ($z_j$) | ✓ | ✓ | ✓ | ✓ |
| Fixed Effects ($\eta_i$) |  | ✓ |  | ✓ |
| Constant | -4.17057*** |  | -2.04840*** |  |
|  | (0.08135) |  | (0.30083) |  |
| Individuals | 16364 | 1797 | 3446 | 385 |
| Observations | 436652 | 83693 | 25062 | 6566 |

Standard errors (in parentheses) are clustered at the user level.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Controls ($z_i$) include: $gender_i$, $age_i$, $college_i$, $graduate_i$, $pic\_score_i$, $num\_pic_i$, $employment_i$, and $bio_i$.

Controls ($z_j$) include: $age_j$, $college_j$, $graduate_j$, $pic\_score_j$, $num\_pic_j$, $employment_j$, $bio_j$, $star1_{jt}$, and $star3_{jt}$.

Table 5: Effect of star-rating on messages received.

We present the results for *reply* behavior in models M6 and M7, which are analogous to M4 and M5. Note that both models control for sender $j$'s attributes because the outcome variable (receiving a reply or not) is conditioned on user $i$ sending a first message to user $j$ in the first place; and $i$'s decision to send a first message can be function of $j$'s characteristics.

Interestingly, we find that conditional on initiating a message, a user is more likely to receive a reply message when s/he is shown with three stars as opposed to two stars. That is, the effect of star-ratings on preference-ranking and replies are quite different (compare models M3 and M7). The main takeaway here is that, in the case of replies, the effects are consistent with the earlier literature that documents positive returns to popularity on demand. Intuitively, when sending a reply message, users are unlikely to be concerned about rejection and therefore rejection concerns may not play any role in their reply behavior. In §9, we formalize and discuss the mechanism that can explain the difference in the effect of star-ratings on preference-ranking and reply behavior in greater detail.

# 8   Robustness Checks

We now present a set of analyses to establish the robustness of the results presented in 6.4 and 7.2.

## 8.1 Effect of Stars on Preference-Rankings - Linear Model

First, we examine whether the substantive results from the non-linear models in §6.4 hold if we directly model the outcome as a linear function of star-ratings and other relevant variables. We therefore consider three linear specifications – (1) a simple model that only includes star-rating variables as the independent variable, (2) a slightly more elaborate model that includes all the user-specific observables ($z_i$), and (3) a linear fixed-effects model. These are the linear analogs of models M1, M2, and M3 in Table 4. The estimates from these models are substantively similar to those from the ordered logit models. Please see Appendix §B.1 for details of the model and results.

## 8.2 Estimation Sample

Next, we examine if our results are driven by the estimation sample used. The MD estimator for the fixed-effects ordered logit model utilizes only a subset of the data for inference – data on users who experienced at least one star change during the observation period. In principle, this sub-population can be different from the full population, and the fixed-effects estimates could simply reflect that difference. We therefore perform a few validation checks. First, we re-estimate model M1 with the sample used in model M3. We see that the results are similar to those obtained from the full sample. Second, we show that the variation in the number of star-changes a user experiences in the observation period is mainly a function of whether the user is new to the app or not. Third, we find no systematic differences between users who go through at least one star change compared to those who do not go through any star change. Please see Appendix §B.2 for details.

Finally, recall that the effect of star-ratings on preference-ranking and replies were quite different. Our explanation of this difference was based on the difference in the perceived probabilities of being rejected. However, this might be due to the differences in the estimation samples used in models M3 and M7. Recall that model M3 includes all users who experienced a star-change while model M7 includes users who experienced a star-change and also initiated a message with their match. As a robustness check, we therefore re-estimate model M3 with the sample used in model M7. We find that the results from this exercise are similar to as those presented in M3 (see Appendix B.2.4).

## 8.3 Within Game Correlation

Recall that $\epsilon_{ijt}$s can include the attributes of the other three players of $i$'s gender who $i$ is being compared with in game $t$. Technically, this can create a correlation between the error $\epsilon_{ijt}$s in one game, if we include the observation of all competitors in one game in our analysis. As discussed in §6.1, this correlation does not affect the consistency of our results, i.e., the estimates are unbiased. However, it can affect the efficiency of our results. To examine if this is an issue, we conduct another robustness check.

Note that a majority of users in our sample never experienced a star change, and recall that the observations of those competitors who never experienced a star change are dropped from our analysis. Therefore, to confirm that our results are not affected by the within game correlation between the errors, we re-estimate the fixed-effects ordered logit model M3 with the games in which only one of the four competitors experienced a star change in the observation period. We find that the results remain similar to those presented in model M3. (See Appendix §B.3 for details.)

## 8.4 Star Configuration in a Game

Users may self-select their entry time when they expect certain types of competitors and this may affect the star configuration of the games. So for the set of users in the estimation sample, we calculate the probability of being in a game with a specific configuration of competitors and present these probabilities in Table A6 in Appendix B.4. We find that the star configuration of the competing players faced by a focal user $i$ is not really function of $i$'s own star-rating. We find that $i$ is competing with three other two-star users in over 94% of the cases. Therefore, regardless of when a three-star or one-star user decides to play a game, they are almost always being compared to other two star users. This ensures that the effect of star-ratings is not driven by users' self-selection into games.

# 9 Discussion of Mechanism

We now examine the mechanism behind the effects established in §6 and §7.

## 9.1 Players' Ranking and Messaging Strategy

We start by formally defining players' ranking strategy during the game, and messaging decisions after the game (with their match).

### 9.1.1 Ranking Strategy During the Game

Let $EU_{ijt}$ denote the expected utility that user $j$ gets conditional on being matched with $i$, such that:

$$EU_{ijt}(star_{it} \mid match_{ijt} = 1) = U(star_{it}) \cdot \mathcal{P} - C \cdot (1 - \mathcal{P}) \tag{17}$$

Here, $U(star_{it})$ denotes the utility that user $j$ expects to receive if she successfully converses with $i$ upon matching. $\mathcal{P}$ denotes $j$'s perceived probability of having a successful conversation with $i$, either by receiving a first message from $i$, or by receiving a reply from $i$ (in response to $j$'s first message). Finally, if $i$ does not respond to $j$ after the match, user $j$ may incur a rejection cost of $C$. $C$ can be interpreted as the psychological cost of rejection because $j$ can infer that $i$ is not interested in pursuing a conversation/date with him/her.[19] Together, $\mathcal{P}$ and $C$ capture $j$'s post-match rejection

---

[19]Users may also get some dis-utility from remaining single and having no one to converse with. Without loss of generality, we normalize this dis-utility to zero.

concerns when s/he is ranking $i$.

Note that $U(star_{it})$ can also be a function of other observed $i$ and $j$ specific variables. Similarly, $\mathcal{P}$ and $C$ can also be a function of $i$'s and $j$'s attributes; for instance, $j$ may suffer higher rejection costs if $i$ is popular (three-star) or attractive. However, these dependencies do not affect any of the arguments used to demonstrate strategic shading in §9.2.1 and therefore we simply denote them as $U(star_{it})$, $\mathcal{P}$, and $C$ to keep the notation simple.

Next, we state a key assumption on users' behavior during the game.

**Assumption 1. Truthfulness:** We assume that the preference-ranking that user $j$ gives to user $i$ is higher than that she gives to $i'$ during game $t$, i.e., $pref_{ijt} > pref_{i'jt}$, if and only if $EU_{ijt} > EU_{i'jt}$.

Assumption 1 states that users are truth-telling, i.e., the relationship between users' latent expected utilities for any pair of potential partners is consistent with their stated preference-rankings. If user $j$'s preferences for four potential partners $1, 2, 3$, and $4$ satisfy the following relationship: $EU_{1jt} > EU_{2jt} > EU_{3jt} > EU_{4jt}$, then the user's revealed preference-rankings is truthful such that: $pref_{1jt} > pref_{2jt} > pref_{3jt} > pref_{4jt}$.

This assumption essentially implies that the ranking game does not induce strategic motivations to deviate from truthfulness. In Appendix §D, we discuss the background for this assumption in detail and empirically validate it.

Finally, it is important to recognize that truth-telling in this context refers to truthfully ranking based on the expected utility from the match (i.e., $EU_{ijt}$s), and not $U(star_{it})$. This is an important distinction that plays a key role in §9.2, when we formally discuss strategic shading.

### 9.1.2 Messaging Strategy after the Game

After the game, each user makes a decision on whether to initiate a message with her/his match and whether to reply to a message (if s/he receives one from her match). The decision to send a first message is not central to our discussion, so we do not define it in the text. However, the decision to reply to a received (first) message is important. So we now formally define it.

We assume that user $j$ replies to the message sent by user $i$ based on her underlying expected utility. Let $EU_{ijt}^{reply}$ denote the expected utility that user $j$ gets from replying to $i$ conditional on receiving the first message from $i$. Since $i$ initiated the first message, $j$ is unlikely to have any rejection concerns when replying to $i$. Thus, unlike Equation (17), there is no rejection probability or cost in the expected utility that user $j$ gets from replying to $i$. Thus, we can write:

$$EU_{ijt}^{reply}(star_{it} \,|\, first_{jit} = 1) = U(star_{it}). \tag{18}$$

We assume that user $j$ replies to $i$, if and only if $EU_{ijt}^{reply} > 0$.

### 9.2 Strategic Shading

We now formally define *strategic shading*.

**Definition 1. Strategic shading:** User $j$'s revealed preference for a potential partner $i$ is *not* just based on the expected utility from a successful conversation/date with $i$ (i.e., $U(\cdot)$). Instead, user $j$'s revealed preference also takes into account the perceived probability of being rejected and rejection costs. This distortion of revealed preference away from $U(\cdot)$ is referred to as *strategic shading*.

Strategic shading can be easily understood in our setting as follows: suppose that users value more popular users, i.e., expect higher utility ($U(\cdot)$) from dating a popular partner. However, if there is a non-zero probability of being rejected (i.e., $\mathcal{P} < 1$), they may reveal lower preferences for popular users. That is, users would strategically shade down their preferences for popular users in order to avoid being rejected in the post-match conversations.

#### 9.2.1 Evidence for Strategic Shading

We can identify the presence of strategic shading in our setting based on the differences in the effect of popularity information (star-ratings) on two revealed preference measures that vary only in the severity of rejection concerns: preference-rankings during the game and reply choice after the game.

We start by invoking the empirical findings on the $reply$ message from §7, which suggests that user $j$ is more likely to send a reply message to a three-star match (who has initiated a first message) compared to two-star match. This implies that:

$$EU_{ijt}^{reply}(star_{it} = 3 \,|\, first_{jit} = 1) > EU_{ijt}^{reply}(star_{it} = 2 \,|\, first_{jit} = 1). \tag{19}$$

Then, based on Inequality (19) and Equation (18), we can infer that:

$$U(star_{it} = 3) > U(star_{it} = 2). \tag{20}$$

This implies that users receive higher utility from a conversation/date with a three-star partner compared to a two-star partner. Next, we characterize the empirical findings from §6 (on $pref$), which suggests that user $j$ is more likely to give a lower preference-ranking to $i$, when $i$ is presented with three stars compared to two stars. This implies that:

$$EU_{ijt}(star_{it} = 3) < EU_{ijt}(star_{it} = 2). \tag{21}$$

The above inequality is based on Assumption 1, which asserts that users' ranking behavior during the game reflects their true preferences, i.e., preference-rankings reflect users' underlying expected

utilities. Since we know from Inequality (20) that $U(star_{it} = 3) > U(star_{it} = 2)$, Inequality (21) can only be explained by rejection concerns, i.e., due to perceived positive probability of rejection $\mathcal{P} < 1$ and non-zero cost of being rejected ($C > 0$). Thus, the negative effect of star-ratings during the game can therefore be directly attributed to rejection concerns.

### 9.2.2 Discussion: Sources of Strategic Shading

We now discuss the sources of strategic shading in our setting in greater detail. First, we start with a brief discussion of standard centralized matching markets, e.g., in the medical labor market (NRMP). In these markets, the underlying assumption is that matches are binding ($\mathcal{P} = 1$), i.e., both hospitals and residents cannot renege on the matches. In such cases, it has been empirically shown that agents have no strong incentives to deviate from ranking potential partners based on their post-match utility, i.e., $U(\cdot)$.[20] That is, when $\mathcal{P} = 1$, users' revealed preferences over potential partners align with their true post-match utilities from those partners. In these cases, even as users recognize that the probability of match with popular partners is low, they continue to give higher preference-ranking to popular agents because if they fail to match with their top choice, they will be automatically considered for their second-best choice, and so on.

Our setting is different from standard centralized matching markets because matches are not binding in our case; there is a high probability of post-match rejection (most matches do not lead to successful conversations, i.e., $\mathcal{P} < 1$). If users expect popular users to be less responsive post-match, then they will shade away from popular users at the ranking stage. Indeed, users may believe that three-star users are less likely to be responsive post-match based on their prior dating experiences or pop culture media. Interestingly, in our data, we found no evidence to suggest that three-star users are less responsive than two-star users after the match.

However, we do find that users' prior success in post-match conversation shapes their ranking strategy. We stratified rank-givers into two groups based on their prior conversation history as *successful* and *unsuccessful*. Successful rank-givers are defined as those who have had more successful conversations with their past matches compared to the median user. Here a successful conversation from a user's perspective is defined as one where s/he either received a first message from the matched partner, or received a reply to a message that s/he initiated. We find that the negative effect of popularity (or three-star rating) comes mainly from the unsuccessful rank-givers (see Appendix E for the details of the model and the table of results). Indeed, we see that users who have been successful in engaging in post-match conversations actually give higher preference

---

[20]Roth (1982) formally shows that there is no mechanism for the stable marriage problem in which truth-telling is the dominant strategy for both men and women. However, a large stream of empirical papers have shown that in most real markets, there is little incentive to distort rankings away from true preferences, $U(\cdot)$ (Roth and Peranson, 1999; Lee, 2016a). We refer you to Appendix D for a more detailed discussion of truth-telling in our setting.

rankings to three-star users. This suggests that the strategic shading mainly stems from users who have not had much success in post-match conversations in the past, and therefore avoid popular users. Moreover, the personal nature of dating can give rise to significant psychological costs of rejection ($C > 0$). If users suffer from being rejected, then they will shade away from popular people whom they perceive as less likely to reciprocate in the post-match conversation stage.[21]

### 9.3 Alternative Mechanisms

We now consider and rule out a few other alternative explanations for the results in §6 and §7.

First, the negative effect of three-stars during the game could be due to the salience effect. Since most users are shown with two-stars (see Table A6 in Appendix B.4 for the distribution of stars in a game), three-star users may be more salient and people may pay more attention to them. However, salience cannot explain the negative effect of popularity because of two reasons. First, salience effect should also come into play for one-star users, but we see no significant effect for one-star users during the game. Second, usually demand increases when we increase the salience of a positive attribute; however we see a negative effect for three-star users.

A second alternative explanation for the negative effect of higher stars during the game could be that users dislike popular users. However, our results show that three-star users are more likely to receive a reply to their first messages after the game. This implies that users receive higher utility from a conversation with a three-star partner (i.e., Inequality (20)). Thus, we can rule out the explanation this explanation. Finally, a third possible reason for the negative effect of higher star-ratings during the game could be the reference-point effect: when a user (rank-giver) sees a potential partner with a higher star-rating, s/he may set a higher reference-point for the rank-receiver. As such, that person is held to a higher standard (for attractiveness/appeal) and if they do not match up to that reference point, a loss component may be added to them. We can rule out this explanation using the same argument as the one used above, i.e., such behavioral biases are not supported by the fact that three-star users receive more replies after the game.

## 10  Conclusion

In this paper, we examine effect of a user's popularity on her/his demand in a mobile dating app at different stages of the matching process, and the drivers of these effects. Specifically, we document the causal impact of a user's star-rating on the preference-rankings that s/he receives during a game and her likelihood of receiving messages after a game. We show that, everything else being constant,

---

[21]In Appendix F, we provide additional evidence in support of strategic shading due to rejection concerns – the negative effect of popularity is mainly driven by rank-givers who are less-attractive than average, when they are considering attractive potential partners.

compared to two-star users three-star users receive lower preference-rankings during the game but receive more reply messages after the game. We then link the heterogeneity across outcomes to the perceived severity of post-match rejection concerns, and establish strategic shading of as the underlying mechanism for the negative effect of popularity during the game.

Our results suggest that managers of online dating markets (and other two-sided matching markets) should take the dampening effect of popularity information into account when designing their user-interface. On the one hand, displaying popularity information can simplify users' search process and help them quickly evaluate potential partners. On the other hand, doing so can have unintended consequences on the demand for popular users. Whether the decrease in search costs offsets the strategic incentives, and how these factors jointly affect the platform's overall health is an empirical question, and worthy of future research.

Our findings have important implications for the design and implementation of centralized matching markets. Centralized matching has been long proposed as a solution to efficiently match agents and avoid the common problems associated with decentralized settings such as costly search and congestion (Roth, 2008). However, our findings suggest that centralized matching markets are also prone to strategic behavior and shading if users have post-match rejection concerns. It is not feasible to enforce binding matches or ignore psychological costs of rejection in markets with inter-personal interactions (e.g., dating markets, freelance markets). Market designers should therefore take these factors into account when designing matching mechanisms for these cases. For instance, even the celebrated success stories of centralized matching, such as the NYC public high school admissions process (Abdulkadiroğlu et al., 2005; Toch and Aldeman, 2009) are likely subject to strategic incentives. While the matches are binding from the school's perspective, they are not so for students. That is, the best students who apply to public high schools may still decide to reject their match in favor of highly selective independent high schools that are not part of the central system. This provides schools a perverse incentive to shade down their rankings of the most attractive students. Thus, the lack of commitment from even one side can lead to strategic shading and sub-optimal outcome. Indeed, as Roth and Peranson (1999) eloquently put it, while the basic SMP algorithm is theoretically elegant and works well in principle, the actual implementation on the ground requires market designers to modify and accommodate the algorithm for domain-specific factors and engineer practical solutions that work in practice.

# References

A. Abdulkadiroğlu and T. Sönmez. Matching markets: Theory and practice. *Advances in Economics and Econometrics*, 1:3–47, 2013.

A. Abdulkadiroğlu, P. A. Pathak, and A. E. Roth. The New York City High School Match. *American Economic Review*, 95(2):364–367, 2005.

P. D. Allison and N. A. Christakis. Logit models for sets of ranked items. *Sociological methodology*, pages 199–228, 1994.

M. Anderson, E. A. Vogles, and E. Turner. The Virtues and Downsides of Online Dating. *Pew Research Center*, 2020. https://www.pewresearch.org/internet/2020/02/06/the-virtues-and-downsides-of-online-dating/.

N. Arnosti, R. Johari, and Y. Kanoria. Managing Congestion in Matching Markets. *Manufacturing and Service Operations Management, special issue on Sharing Economy and Innovative Marketplaces*, 2019.

G. Baetschmann, K. E. Staub, and R. Winkelmann. Consistent Estimation of the Fixed Effects Ordered Logit Model. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 178(3):685–703, 2015.

R. Bapna, J. Ramaprasad, G. Shmueli, and A. Umyarov. One-Way Mirrors in Online Dating: A Randomized Field Experiment. *Management Science*, 62(11):3100–3122, 2016.

J. E. Biddle and D. S. Hamermesh. Beauty, Productivity, and Discrimination: Lawyers' Looks and Lucre. *Journal of Labor Economics*, 16(1):172–201, 1998.

H. Cai, Y. Chen, and H. Fang. Observational Learning: Evidence from a Randomized Natural Field Experiment. *American Economic Review*, 99(3):864–82, 2009.

J. J. Cameron, D. A. Stinson, and J. V. Wood. The Bold and the Bashful: Self-Esteem, Gender, and Relationship Initiation. *Social Psychological and Personality Science*, 4(6):685–691, 2013.

G. Chamberlain. Analysis of Covariance with Qualitative Data. *The Review of Economic Studies*, 47(1): 225–238, 1980.

Y.-K. Che and Y. Koh. Decentralized College Admissions. *Journal of Political Economy*, 124(5):1295–1338, 2016.

Y. Chen, Q. Wang, and J. Xie. Online Social Interactions: A Natural Experiment on Word of Mouth Versus Observational Learning. *Journal of Marketing Research*, 48(2):238–254, 2011.

J. A. Chevalier and D. Mayzlin. The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal of Marketing Research*, 43(3):345–354, 2006.

E. Choo and A. Siow. Who Marries Whom and Why. *Journal of political Economy*, 114(1):175–201, 2006.

M. Das and A. Van Soest. A Panel Data Model for Subjective Information on Household Income Growth. *Journal of Economic Behavior & Organization*, 40(4):409–426, 1999.

G. Demange, D. Gale, and M. Sotomayor. A Further Note on the Stable Matching Problem. *Discrete Applied Mathematics*, 16(3):217–222, 1987.

S. Dewan, Y.-J. Ho, and J. Ramaprasad. Popularity or Proximity: Characterizing the Nature of Social Influence in an Online Music Community. *Information Systems Research*, 28(1):117–136, 2017.

W. Duan, B. Gu, and A. B. Whinston. Informational Cascades and Software Adoption on the Internet: An Empirical Investigation. *MIS quarterly*, pages 23–48, 2009.

P. W. Eastwick and E. J. Finkel. Sex Differences in Mate Preferences Revisited: Do People Know What they Initially Desire in a Romantic Partner? *Journal of Personality and Social Psychology*, 94(2):245, 2008.

R. Fisman, S. S. Iyengar, E. Kamenica, and I. Simonson. Gender Differences in Mate Selection: Evidence From a Speed Dating Experiment. *The Quarterly Journal of Economics*, 121(2):673–697, 2006.

R. Fisman, S. S. Iyengar, E. Kamenica, and I. Simonson. Racial Preferences in Dating. *The Review of Economic Studies*, 75(1):117–132, 2008.

J. Fong. Search, Selectivity, and Market Thickness in Two-Sided Markets: Evidence from Online Dating.

*Working Paper*, 2020.

D. Gale and L. S. Shapley. College Admissions and the Stability of Marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962.

S. L. Hansen. Dating Choices of High school Students. *Family Coordinator*, pages 133–138, 1977.

G. J. Hitsch, A. Hortaçsu, and D. Ariely. Matching and Sorting in Online Dating. *American Economic Review*, 100(1):130–63, 2010a.

G. J. Hitsch, A. Hortaçsu, and D. Ariely. What Makes You Click?—Mate Preferences in Online Dating. *Quantitative Marketing and Economics*, 8(4):393–427, 2010b.

A. R. Hole, A. Dickerson, L. Munford, et al. A Review of Estimators for the Fixed-Effects Ordered Logit Model. In *United Kingdom Stata Users' Group Meetings 2011*, number 05. Stata Users Group, 2011.

IBISWorld. Dating Services Industry in the US - Market Research Report, 2019. https://www.ibisworld.com/industry-trends/market-research-reports/other-services-except-public-administration/personal-laundry/dating-services.html.

J. Jung, R. Bapna, J. Ramaprasad, and A. Umyarov. Love Unshackled: Identifying the Effect of Mobile App Adoption in Online Dating. *MIS Quarterly*, 43:47–72, 2019.

R. Kurzban and J. Weeden. HurryDate: Mate Preferences in Action. *Evolution and Human Behavior*, 26(3): 227–244, 2005.

D. S. Lee and T. Lemieux. Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2):281–355, 2010.

S. Lee. Effect of Online dating on Assortative Mating: Evidence from South Korea. *Journal of Applied Econometrics*, 31(6):1120–1139, 2016a.

S. Lee. Incentive Compatibility of Large Centralized Matching Markets. *Working Paper*, 2016b.

S. Lee and L. Yariv. On the Efficiency of Stable Matchings in Large Markets. *Working Paper*, 2018.

J. Ludden. Do You Like Me?, 2016. https://www.npr.org/sections/thetwo-way/2016/02/11/466342716/do-you-like-me-swiping-leads-to-spike-in-online-dating-for-young-adults.

E. Moretti. Social Learning and Peer Effects in Consumption: Evidence from Movie Sales. *The Review of Economic Studies*, 78(1):356–393, 2011.

J. Neyman and E. L. Scott. Consistent Estimates Based on Partially Consistent Observations. *Econometrica: Journal of the Econometric Society*, pages 1–32, 1948.

M. Niederle and L. Yariv. Decentralized Matching with Aligned Preferences. Technical report, National Bureau of Economic Research, 2009.

M. Niederle, A. E. Roth, and T. Sönmez. Matching. *The New Palgrave Dictionary of Economics. Palgrave Macmillan*, 2, 2008.

B. Pittel. The Average Number of Stable Matchings. *SIAM Journal on Discrete Mathematics*, 2(4):530–549, 1989.

A. E. Roth. The Economics of Matching: Stability and Incentives. *Mathematics of Operations Research*, 7 (4):617–628, 1982.

A. E. Roth. Two-Sided Matching with Incomplete Information About Others' Preferences. *Games and Economic Behavior*, 1(2):191–209, 1989.

A. E. Roth. Deferred Acceptance Algorithms: History, Theory, Practice, and Open Questions. *International Journal of Game Theory*, 36(3-4):537–569, 2008.

A. E. Roth and E. Peranson. The redesign of the matching market for American physicians: Some engineering aspects of economic design. *American Economic Review*, 89(4):748–780, 1999.

A. E. Roth and M. A. O. Sotomayor. *Two-sided matching*. Cambridge University Press, Cambridge, 1990.

A. E. Roth and X. Xing. Turnaround Time and Bottlenecks in Market Clearing: Decentralized Matching in

the Market for Clinical Psychologists. *Journal of Political Economy*, 105(2):284–329, 1997.

V. Safronova. An Inside Look at Your Favorite Dating Sites, 2018. https://www.nytimes.com/2018/04/11/style/match-shaadi-league-farmersonly-dating-apps.html.

M. J. Salganik, P. S. Dodds, and D. J. Watts. Experimental Study of Inequality and Unpredictability in an Artificial Cultural Market. *Science*, 311(5762):854–856, 2006.

L. Shaw Taylor, A. T. Fiore, G. Mendelsohn, and C. Cheshire. "Out of My League": A Real-World Test of the Matching Hypothesis. *Personality and Social Psychology Bulletin*, 37(7):942–954, 2011.

A. T. Sorensen. Bestseller Lists and Product Variety. *The Journal of Industrial Economics*, 55(4):715–738, 2007.

M. Sun. How Does the Variance of Product Ratings Matter? *Management Science*, 58(4):696–707, 2012.

T. Toch and C. Aldeman. Matchmaking: Enabling Mandatory Public School Choice in New York and Boston. Ideas at Work. *Education Sector*, 2009.

C. Tucker and J. Zhang. How Does Popularity Information Affect Choices? A Field Experiment. *Management Science*, 57(5):828–842, 2011.

L. Y. Wong. Structural Estimation of Marriage Models. *Journal of Labor Economics*, 21(3):699–727, 2003.

H. Yoganarasimhan. Impact of Social Network Structure on Content Propagation: A Study using YouTube Data. *Quantitative Marketing and Economics*, 10(1):111–150, 2012.

H. Yoganarasimhan. The Value of Reputation in an Online Freelance Marketplace. *Marketing Science*, 32(6):860–891, 2013.

J. Zhang. The Sound of Silence: Observational Learning in the US Kidney Market. *Marketing Science*, 29(2):315–335, 2010.

# Appendices

## A   Other Modeling Frameworks

We use a fixed-effects ordered logit model to estimate the effect of the star-rating of a user on the preference ranking that s/he receives. We now explain why two other commonly used approaches, (1) rank-ordered logit with fixed effects and (2) regression discontinuity method, are not appropriate for our setting.

### A.1   Rank-Ordered Logit

The rank-ordered logit model is specified from the perspective of the "rank-giver". As before, the dependent variable in this case is also the preference-ranking, $pref_{ijt}$, which denotes the preference-ranking that player $j$ gives to mate $i$ at game $t$.

Let $u_{ijt}$ be the latent utility that user $j$ expects to receive from being matched with $i$. Following Allison and Christakis (1994), we can write $u_{ijt}$ as a sum of two components such that:

$$u_{ijt} = \mu_{ijt} + \varepsilon_{ijt}, \tag{A.1}$$

where $\varepsilon_{ijt}$ is an idiosyncratic preference shock and $\mu_{ijt}$ is a linear function of user $i$'s observed characteristics to $j$ and us, $i$'s unobserved characteristics to us ($\eta_i$), and user $j$'s characteristics, such that:

$$\mu_{ijt} = \beta star_{it} + \gamma z_i + \eta_i + \theta X_{it} + \alpha z_j + \eta_j. \tag{A.2}$$

Assume that user $j$ is ranking two potential mates $i$ and $k$. Although $u_{ijt}$s are unobserved, we assume that player $j$ gives $i$ a higher preference-ranking than mate $k$ whenever $u_{ijt} > u_{kjt}$. Under the assumption that $\varepsilon_{ijt}$s are IID drawn from a type I extreme value distribution, we can write:

$$
\begin{aligned}
Pr(u_{ijt} > u_{kjt}) &= \frac{exp(\mu_{ijt})}{exp(\mu_{ijt}) + exp(\mu_{kjt})} \\
&= \frac{exp(\beta star_{it} + \gamma z_i + \eta_i)}{exp(\beta star_{it} + \gamma z_i + \eta_i) + exp(\beta star_{kt} + \gamma z_k + \eta_k)}.
\end{aligned} \tag{A.3}
$$

Note that user $j$'s characteristics are canceled out in Equation (A.3). However, the fixed effects for the rank-receivers ($\eta_i$ and $\eta_k$) are not cancelled.

Similarly, when user $j$ ranks four potential mates $\{i, k, l, m\}$ in game $t$ and s/he gives preference-rankings of $\{4,3,2,1\}$ (without loss of generality), we can infer that $u_{ijt} > u_{kjt} > u_{ljt} > u_{mjt}$ and write:

$$
\begin{aligned}
Pr(u_{ijt} > u_{kjt} > u_{ljt} > u_{mjt}) &= \frac{exp(\mu_{ijt})}{exp(\mu_{ijt}) + exp(\mu_{kjt}) + exp(\mu_{ljt}) + exp(\mu_{mjt})} \\
&\times \frac{exp(\mu_{kjt})}{exp(\mu_{kjt}) + exp(\mu_{ljt}) + exp(\mu_{mjt})} \\
&\times \frac{exp(\mu_{ljt})}{exp(\mu_{ljt}) + exp(\mu_{mjt})}.
\end{aligned} \tag{A.4}
$$

Similar to Equation (A.3), if we expand equation (A.4), we will have four receivers' fixed effects ($\eta_i$, $\eta_k$, $\eta_l$, and $\eta_m$). We can write the likelihood of user $j$ giving preference-rankings to his or her potential matches in

game $t$ as:

$$L_{jt} = \prod_{i=1}^{4} \frac{exp(\mu_{ijt})}{\sum_{k=1}^{4} \delta_{ijtk} exp(\mu_{kjt})} \tag{A.5}$$

where

$$\delta_{ijtk} = \begin{cases} 1, & \text{if } pref_{ijt} > pref_{kjt} \\ 0, & \text{else.} \end{cases} \tag{A.6}$$

Now, we can write the log-likelihood of the preference-rankings observed in the data as:

$$\begin{aligned} LL(\beta, \gamma, \eta_1, ..., \eta_N) &= \sum_{j=1}^{N} \sum_{t=1}^{T_j} \ln[L_{jt}] \\ &= \sum_{j=1}^{N} \sum_{t=1}^{T_j} \sum_{i=1}^{4} (\beta star_{it} + \gamma z_i + \eta_i) \\ &- \sum_{j=1}^{N} \sum_{t=1}^{T_j} \sum_{i=1}^{4} \ln[\sum_{k=1}^{4} \delta_{ijtk} exp(\beta star_{it} + \gamma z_i + \eta_i)]. \end{aligned} \tag{A.7}$$

Notice that unlike the ordered logit model, here we cannot condition out the rank-receivers' fixed effects $(\eta_i)$s using CML-style estimators. So there is no way to consistently estimate the effect of the receiver's star-ratings in the rank-ordered logit specification.

## A.2 Regression Discontinuity Design (RDD)

We now briefly explain the main idea behind a Regression Discontinuity Design (RDD) and then discuss why our setting does not satisfy the main assumptions necessary for RDD. In RDD, treatment is determined by comparing the value of an observed running variable to a known threshold. In a valid RD design, treatment effect is identifiable if: (1) individuals just below the threshold are similar to those just above it, and (2) individuals are unable to precisely control their running variable near the threshold Lee and Lemieux (2010). These assumptions provide local randomization around the threshold. So any jump in the outcome variable below and above the threshold represents the treatment effect.

In our setting, $popularity_{it}$ can play the role of the running variable. A user $i$ receives the three-star treatment in game $t$ if her $popularity_{it}$ is equal to or above three, and the two-star treatment if $popularity_{it}$ lies between two and three (see Figure 3). A RD design would typically focus on a sample of observations where the running variable lies within a small bandwidth just above and below the threshold. Here, we focus on a sample of observations where $popularity_{it}$ lies within a small bandwidth around the cutoff three, e.g., [2.95, 3.05]. Although we can claim that users cannot precisely manipulate the running variable ($popularity_{it}$), we cannot claim that the observations on the two sides of the cut-off are similar because of two reasons. First, there is a lot of fluctuation in a user's star-rating in their first few games. The same individual can fall on different sides of the bandwidth at different times. However, as a user plays more games, her/his star-rating starts converging to a stable number. Because the threshold does not distinguish players based on the number of prior games, it will pool players who played a few games and received a popularity in the range of [2.95, 3.05] with those who played many games and have a stable popularity in that range. Therefore, we cannot argue that the observations just below and above the threshold are comparable. Second, the running variable $popularity_{it}$ is calculated based on the previous values of the outcome variable ($pref_{ijt}$), which are influenced by the user's previous star-ratings. This contamination violates the randomization

around the threshold i.e., users around the threshold can differ in their history of prior treatments, which can have a systematic effect on their current star-rating. Thus, the first condition of RDD to identify the treatment effect (similarity of the observations around the threshold) is not satisfied.

# B    Appendix for Robustness Checks

## B.1    Effect of Stars on Preference-Rankings - Linear Model

We consider the following linear model:

$$pref_{ijt} = \beta_1 star1_{it} + \beta_2 star3_{it} + \gamma z_i + \eta_i + \epsilon_{ijt} \tag{A.8}$$

The main difference between these coefficients and those discussed in §6.1 is that these coefficients directly relate to the observed outcome instead of the latent variable $pref^*$. Hence, even though we use the same variable names for expositional convenience, the interpretation of the coefficients in the two models is different. In short, the magnitude of the coefficients from the two models cannot be directly compared.

There are three possible estimation strategies here: (1) pooled OLS, that only includes star-ratings variables as the independent variables but ignores the problem of correlated unobservables, (2) a slightly more elaborate pooled OLS that includes all user-specific control variables ($z_i$), and (3) fixed-effects model, which addressed the omitted variable bias due to $\eta_i$ by employing a "within" transformation to subtract out the time-invariant user-specific variables.

|  | (A1) (OLS) | (A2) (OLS) | (A3) (FE) |
|---|---|---|---|
| $star1_{it}$ | -0.08946*** | -0.08001*** | 0.01776 |
|  | (0.01422) | (0.01412) | (0.01126) |
| $star3_{it}$ | 0.03779*** | 0.04241*** | -0.03100*** |
|  | (0.00971) | (0.01171) | (0.00913) |
| Controls ($z_i$) |  | ✓ |  |
| Constant | 2.50031*** | 2.50558*** | 2.50006*** |
|  | (0.00113) | (0.00740) | (0.00034) |
| Individuals | 24393 | 16461 | 3494 |
| Observations | 2980148 | 2339168 | 630160 |
| R-Squared | 0.00003 | 0.00254 | 0.00002 |

Standard errors (in parentheses) are clustered at the user level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$
Controls in Model A2 include: $age_i$, $college_i$, $graduate_i$, $pic\_score_i$, $num\_pic_i$, $employment_i$, and $bio_i$.

Table A1: Pooled OLS and fixed-effects estimates of the effect of user's star-rating on preference-rankings received.

A pooled OLS estimation strategy consists of pooling all the data across games and users, and running a regression on this data. The results from pooled OLS models are shown in Models A1 and A2 in Table A1.The results from model A1 and A2 are substantively similar to those in model M1 and M2 in Table 4.

Next, we discuss the fixed-effects estimation approach. Here, we start with the following averaging equation for each user $i$:

$$\overline{pref}_i = \beta_1 \overline{star1}_i + \beta_2 \overline{star3}_i + \gamma z_i + \eta_i + \overline{\epsilon}_i, \tag{A.9}$$

where $\overline{pref}_i = \frac{\sum_{t=1}^{T_i} \sum_j pref_{ijt}}{4 \times T_i}$, $\overline{star1}_i = \frac{\sum_{t=1}^{T_i} star1_{it}}{T_i}$, $\overline{star3}_i = \frac{\sum_{t=1}^{T_i} star3_{it}}{T_i}$, and $\overline{\epsilon}_i = \frac{\sum_{t=1}^{T_i} \sum_j \epsilon_{ijt}}{4 \times T_i}$. $z_i, \eta_i$ are constant across time periods, and hence their averages are the same as the variables themselves. Next, we subtract Equation (A.9) from Equation (A.8) as follows:

$$pref_{ijt} - \overline{pref}_i = \beta_1 \left( star1_{it} - \overline{star1}_i \right) + \beta_2 \left( star3_{it} - \overline{star3}_i \right) + \left( \epsilon_{ijt} - \overline{\epsilon}_i \right) \tag{A.10}$$

Note that all the time-invariant user-specific variables are now subtracted out and the new error term, $\epsilon_{ijt} - \overline{\epsilon}_i$, is no longer correlated with the star-ratings variables. The fixed-effects estimator is essentially a pooled OLS estimator for Equation (A.10) and it gives us consistent estimates of $\beta_1$ and $\beta_2$ under the linearity assumption. The results from this model are shown in model A3 in Table A1. Note that to keep the comparisons consistent, we only use the first 100 games of users who saw at least one star change during the observation period. Hence, model A3 in analogous to model M3 in Table 4. The results from model A3 are substantively similar to those in model M3. This suggests that our main results were not an artefact of the parametric specification of the model.

## B.2   Estimation Sample

We present validation checks to confirm that our results in model M3, Table 4, are not driven by the estimation sample (which consists of users who experienced at least one star change during the observation period).

### B.2.1   Pooled ordered logit on users who went through star change

First, we run the pooled ordered logit model for rankings on the subset of users who experienced at least one star change during the observation period (sample used in model M3). As shown in Table A2, the magnitude and direction of the estimates in Model A4 are similar to those for the full population model M1.

|  | (A4) |
|---|---|
| $star1_{it}$ | -0.13888*** |
|  | (0.02369) |
| $star3_{it}$ | 0.05136*** |
|  | (0.01590) |
| $\mu_2$ | -1.09054*** |
|  | (0.00456) |
| $\mu_3$ | -0.00036 |
|  | (0.00420) |
| $\mu_4$ | 1.09067*** |
|  | (0.00446) |
| Individuals | 3494 |
| Observations | 630160 |

Standard errors (in parentheses) are clustered at the user level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A2: Ordered logit estimates of the effect of star-rating on preference-rankings received (without fixed-effects), on the sample used in model M3.

### B.2.2 Variation in Popularity Scores and Star-Ratings over Time

Next, we find that users who experience at least one star change are more likely to be new users who joined the app recently and a vast majority of them had not played any games at the start of the observation period. In contrast, users who do not see a star change are users who had played a large number of games in the past. It is important to note that this difference between new and old users does not reflect inherent differences in users, i.e., differences on user characteristics. Rather, it captures the dynamics of star-ratings. As users play more games, the marginal impact of a new game on their average popularity score is small. Thus, users who have played more games are less likely to experience a star change compared to new users.

We illustrate this point using Figure A1, which shows how the change in users' popularity score in a given game ($\Delta popularity_{it}$) varies as a function of the number of games played ($total\_game_{it}$). Here, $\Delta popularity_{it}$ is the absolute value of change and equals to $|popularity_{it} - popularity_{it-1}|$. Recall that popularity score ($popularity_{it}$) is simply the average of preference-rankings received by $i$ in all her/his prior $t - 1$ games. For the average user, the expected change in popularity score reduces to 0.03 after fifteen games. This is simply due to the Law of Large Numbers – for any user $i$ with a set of characteristics $z_i$, $\eta_i$, the popularity score ($popularity_{it}$) starts converging to a constant value after a few games (i.e., the marginal effect of each new ranking decreases). Thus, the variation in the number of star-changes a user experiences in the observation period is largely a function of whether s/he is new to the app or not.

### B.2.3 Comparison of User-specific Observables for New Users

Of the 3,494 users who experience a star change in our observation period, 3,439 (98%) of them are new users who joined in the observation period ($initial\_game_i = 0$). We now compare the user-specific observables of these 3,439 new users (who went through a star-change) with those of new users who did not go through a star-change during our observation period (3,680 users). The results from this comparison are presented
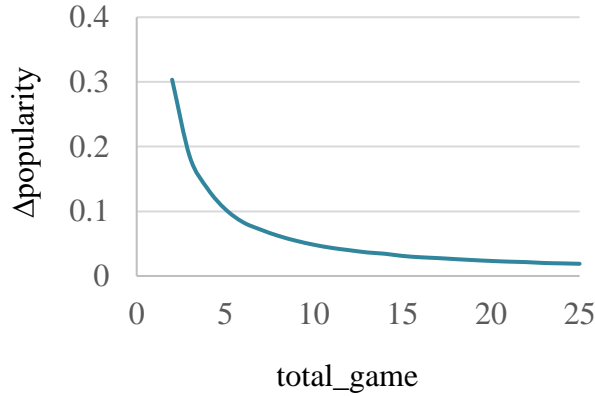
Figure A1: Absolute change in popularity score as a function of number of games played for all the users in our data for the observation period.

| Variables | Star Change | Mean | Std. Dev | Size | $Pr(|T| > |t|)$ |
|---|---|---|---|---|---|
| $age_i$ | No | 21.950 | 7.393 | 2300 | 0.449 |
| | Yes | 22.113 | 7.563 | 2538 | |
| | | | | | |
| $bio_i$ | No | 56.909 | 168.424 | 2715 | 0.368 |
| | Yes | 53.045 | 152.914 | 2920 | |
| | | | | | |
| $education_i$ | No | 1.737 | 0.512 | 2420 | 0.083 |
| | Yes | 1.712 | 0.510 | 2595 | |
| | | | | | |
| $employment_i$ | No | 1.777 | 1.295 | 1614 | 0.758 |
| | Yes | 1.791 | 1.333 | 1727 | |
| | | | | | |
| $num\_pic_i$ | No | 5.355 | 1.393 | 2629 | 0.665 |
| | Yes | 5.338 | 1.426 | 2828 | |
| | | | | | |
| $pic\_score_i$ ($Male$) | No | -0.092 | 0.635 | 1246 | 0.296 |
| | Yes | -0.066 | 0.643 | 1296 | |
| | | | | | |
| $pic\_score_i$ ($Female$) | No | -0.021 | 0.682 | 1066 | 0.077 |
| | Yes | 0.031 | 0.737 | 1246 | |

Table A3: Comparison of attributes between new users who experienced no star change and new users who experienced at least one star change.

in Table A3.[22] Overall, there is sufficient empirical evidence to suggest that new users who experience at least one star change and those who experience no star changes are largely similar. Thus, we expect that the findings from the fixed-effects model to be applicable to the full population of users in the app.

---

[22]For each variable, we only include observations where users reported some value for it. That is why, the size of the observations varies across variables.

### B.2.4 Estimates from the Model M3 on the Sample Used in Model M7

Model M3 included all users who experienced a star-change and model M7 included observations where the user experienced a star-change and also initiated a message. Below, we re-estimate model M3 with the sample used in model M7. We find that the results from this exercise are qualitatively similar to those presented in model M3 (see Table A4 in Appendix §B.2).

|  | (A5) |
|---|---|
| $star1_{it}$ | 0.06800 |
|  | (0.09577) |
|  |  |
| $star3_{it}$ | -0.16914*** |
|  | (0.06463) |
| Individuals | 383 |
| Observations | 21696 |

Standard errors (in parentheses) are clustered at the user level.
\* $p < 0.1$, \*\* $p < 0.05$, \*\*\* $p < 0.01$

Table A4: Ordered logit fixed-effects estimates of the effect of star-rating on preference-rankings received, on the sample used in model M7.

### B.3 Within Game Correlation

|  | (A6) |
|---|---|
| $star1_{it}$ | -0.01546 |
|  | (0.02713) |
|  |  |
| $star3_{it}$ | -0.07380*** |
|  | (0.02135) |
| Individuals | 3430 |
| Observations | 248,944 |

Standard errors (in parentheses) are clustered at the user level.
\* $p < 0.1$, \*\* $p < 0.05$, \*\*\* $p < 0.01$

Table A5: Ordered logit fixed-effects estimates of the effect of star-rating on preference-rankings received, for a subset of games with one competitor who experienced a star change.

### B.4 Star Configuration of the Competitors in a Game

For the set of users in the estimation sample, we calculate the probability of being in a game with a specific configuration of competitors and present these probabilities in Table A6.[23] The first row considers observations where user $i$ has a one-star rating ($star_{it} = 1$). In this case, the probability that s/he is competing

---

[23]In Table A6, we only consider observations for users ($i$) who went through at least one star-change in the observation period (to keep it consistent with our estimation sample in the fixed-effects model M3). Further, we only consider games where all four competitors are shown with a star-rating. Users are not shown with any star-rating in their first game. Further, sometimes, a user may compete with other players who are not shown with any star-rating. Therefore, the total number of observations in Table A6 is smaller than the total number of observations in model M3 in Table 4.

with three users (i.e., the three other players of the same gender as $i$ in game $t$) who all have two stars is 94.17%. Next, in observations where a user $i$ is shown with two stars ($star_{it} = 2$), the probability of competing with three two-star players is 96.77%. And, when user $i$ is shown with three stars ($star_{it} = 3$), this probability is 94.11%. Therefore, regardless of when a given user $i$ decides to play a game, s/he is almost always competing with a similar configuration of players. In particular, s/he is being compared to other two-star users in over 94% of the cases. Thus, the data doesn't show any evidence that users are self-selecting entry time to avoid/obtain certain types of competitors.

| | | Total number of competitors with two stars | | | |
|---|---|---|---|---|---|
| $star_{it}$ | Observations ($ijt$) | 3 | 2 | 1 | 0 |
| 1 | 10,708 | 10,084 (94.17%) | 616 (5.75%) | 8 (0.07%) | 0 |
| 2 | 583,328 | 564,512 (96.77%) | 18,444 (3.16%) | 360 (0.06%) | 12 (0.00%) |
| 3 | 17,644 | 16,604 (94.11%) | 984 (5.58%) | 56 (0.32%) | 0 |
| Total | 611,680 | | | | |

Table A6: Number of observations and probability distribution of the number of two-star competitors that a focal user $i$ faces. (The set of users is the same as the estimation sample, i.e., users who went through at least one star-change ($star_{it}$)).

## C  Conditional Log Likelihood for the Fixed-effects Logit Model

To study the relationship between the users' likelihood of receiving messages and their star-ratings, we consider the following fixed-effects logit formulations:

$$y_{ijt} = \begin{cases} 1, & y_{ijt}^* > 0 \\ 0, & \text{else} \end{cases}$$

where $y_{ijt}$ is a binary variable and it can refer to $first_{ijt}$ or $reply_{ijt}$, and $y_{ijt}^*$ is the corresponding latent variable as follows:

$$y_{ijt}^* = \beta_1 star1_{it} + \beta_2 star3_{it} + \gamma z_i + \eta_i + \epsilon_{ijt}, \tag{A.11}$$

We allow for $\eta_i$ to be arbitarily correlated to $star1_{it}$ and $star3_{it}$. Further, we assume that $star1_{it}$, $star3_{it}$ and $\eta_i$ are independent of $\epsilon_{ijt}$ since users are randomly assigned to games. Assuming that $\epsilon_{ijt}$s are IID and drawn from an Extreme Value Type I distribution, we can write:

$$Pr(y_{ijt} = 1 \mid star1_{it}, star3_{it}, z_i, \eta_i, \beta_1, \beta_2) = \frac{exp(\beta_1 star1_{it} + \beta_2 star3_{it} + \gamma z_i + \eta_i)}{1 + exp(\beta_1 star1_{it} + \beta_2 star3_{it} + \gamma z_i + \eta_i)} \tag{A.12}$$

We can now write the log-likelihoods of $y_{ijt}$ (the first messages or replies) observed in the data as:

$$LL(\beta_1, \beta_2, \gamma) = \sum_{i=1}^{N} \sum_{t=1}^{T_i} \sum_{k=0}^{1} \ln \left[ Pr(y_{ijt} = k \mid star1_{it}, star3_{it}, z_i, \eta_i, \beta_1, \beta_2)^{I(y_{ijt}=k)} \right] \tag{A.13}$$

where $N$ is the total number of users and $T_i$ is the total number of games played by user $i$. Treating the $\eta_i$'s as parameters and maximizing this log-likelihood via Maximum Likelihood Estimator (MLE) is inconsistent with large $N$ and finite $T$ due to the well-known incidental parameters problem (Neyman and Scott, 1948). As a result, the estimate of $\beta_1$, $\beta_2$ from this approach will be inconsistent. However, Chamberlain (1980)

proposes a method to maximize a Conditional Log-Likelihood which gives consistent estimates. Following Chamberlain (1980), we denote $s_i$ as the sum of all received messages (first messages or reply messages) by user $i$ from his/her matches over time, that is:

$$s_i = \sum_{t=1}^{T_i} (y_{ijt} \mid match_{ijt} = 1) \tag{A.14}$$

and, we denote $B_i$ as the set of all possible vectors of length $T_i$ with $s_i$ elements equal to 1, and $T_i - s_i$ elements equal to 0, i.e. all possible ways that user $i$ could receive $s_i$ messages in total over $T_i$ games, that is:

$$B_i = \{d \in \{0,1\}^{T_i} \mid \sum_{t=1}^{T_i} (d_{jt} = s_i \mid match_{ijt} = 1)\} \tag{A.15}$$

For example, if user $i$ plays three games ($T_i = 3$), and receives only one message in total ($s_i = 1$), $B_i$ will be equal to $\{(1,0,0),(0,1,0),(0,0,1)\}$. Now, we can write the conditional probability of $y_i$ given $s_i$ as:

$$Pr\left(y_i \mid star1_{it}, star3_{it}, s_i, \beta_1, \beta_2\right) = \frac{exp\left(y_i.\left(\beta_1 star1_{it} + \beta_2 star3_{it}\right)\right)}{\sum_{d \in B_i} exp\left(d.\left(\beta_1 star1_{it} + \beta_2 star3_{it}\right)\right)} \tag{A.16}$$

Note that this conditional probability does not depend on $\eta_i$'s, i.e. $s_i$ is a sufficient statistic for $\eta_i$. Thus, we can now specify a Conditional Log-Likelihood that is independent of $\eta_i$s as shown below:

$$CLL(\beta_1, \beta_2) = \sum_{i=1}^{N} \sum_{t=1}^{T_i} \ln\left[Pr(y_i \mid star1_{it}, star3_{it}, s_i, \beta_1, \beta_2)\right] \tag{A.17}$$

# D    Validation Check: Truthfulness Assumption

In §9.2, we assumed that users state their preference-rankings truthfully during the game in Assumption 1. This assumption ensures that the relationship between users' latent expected utilities for any pair of potential partners is consistent with their stated preference-ranking over them, i.e., if $EU_{ijt} > EU_{i'jt}$, then $j$ should rank $i$ and $i'$ such that $pref_{ijt} > pref_{i'jt}$. We now present some background for this assumption and empirically validate it.

In our setting, the ranking game resembles a one-to-one marriage SMP, where: (1) agents have to state their strict preference-rankings (i.e., no indifference rankings), (2) agents cannot truncate their list of preference-rankings (i.e., they cannot strategically choose to only rank their top few choices and refuse to rank their bottom choices), (3) agents cannot collude with each other, (4) agents' preferences are private (i.e,. users know their own preferences but not those of others'). Under such circumstances, it has been shown that, when a men-optimal stable matching mechanism is used, it is the dominant strategy for each man to state his true preferences, and any strategy for a woman is dominated if her stated first choice is not her true first choice; and vice-versa for women-optimal stable matching mechanism (Roth, 1989).[24] However, it has been shown that the incentive to manipulate true preferences is negligible for both sides in most real, large markets (Demange et al., 1987; Pittel, 1989; Lee and Yariv, 2018; Lee, 2016b).

Our platform does not use either a men-optimal or a women-optimal matching mechanism. Instead, as

---

[24]The kind of stability studied in the case of incomplete information is ex-post stability, i.e. a stable matching would remain stable even if all the preferences were to become common knowledge (Roth, 1989; Roth and Sotomayor, 1990).

|  | State true | $1^{st}$ and $2^{nd}$ | $2^{nd}$ and $3^{rd}$ |
| Match with | preferences | preference misrepresentation | preference misrepresentation |
|---|---|---|---|
|  | Assuming stated preferences are true preferences | | |
| true $1^{st}$ choice | 49.24 | 28.23 | 49.27 |
| true $2^{nd}$ choice | 28.25 | 49.28 | 14.90 |
| true $3^{rd}$ choice | 15.00 | 14.99 | 28.33 |
| true $4^{th}$ choice | 7.51 | 7.50 | 7.50 |
|  | Assuming true preferences are random | | |
| true $1^{st}$ choice | 49.48 | 28.15 | 49.47 |
| true $2^{nd}$ choice | 28.17 | 49.54 | 14.86 |
| true $3^{rd}$ choice | 14.90 | 14.89 | 28.21 |
| true $4^{th}$ choice | 7.45 | 7.42 | 7.46 |

Table A7: Match results if users misrepresent their preferences.

discussed in §3.2.3, it calculates the set of all possible stable matches and picks the matching with the highest average match-level. Under these conditions, there are no theoretical guarantees on truth-telling for any side of the market. Nonetheless, there are no obvious reasons for users to deviate from truth-telling in our setting. While we cannot theoretically prove this, we now empirically establish that, on average, users cannot gain by mis-representing their preferences in our setting.

We now present two types of deviation checks. In the top panel of Table A7, we start with the assumption that a player's stated preferences are her/his true preferences. The second column represents the average probability of a player being matched with her/his true first, second, third, and fourth choices if the player ranks truthfully (based on the preference-rankings and match levels observed in the data). We find that truthful revelation leads to being paired with the first choice 49.24% of the times, the second choice 28.25% of the times, the third choice 15.00% of the times, and the last choice 7.51% of the times. Next, we consider the following deviation: suppose that in game $t$, everyone except a focal player $j$ plays the same strategy as that observed in the data, and $j$ swaps her/his first and second choices. We then calculate which of her/his true preferences $j$ will be matched with. Then, we aggregate the match outcomes over all players and all games to obtain the average probability of being matched with one's true first choice under this deviation as:

$$\text{Pr(true first choice)} = \frac{\sum_{t=1}^{T} \sum_{j \in t} \text{I}(match\_level_{jt} = \text{true first choice} | pref_{jt}^{12}, pref_{-jt})}{8T}, \quad \text{(A.18)}$$

where $pref_{jt}^{12}$ denotes a strategy where player $j$ swaps her true first and second choices, and $pref_{-jt}$ denotes the preference-rankings observed in the data (i.e., other users' strategies). Similarly, we also calculate the average probabilities of being matched with one's true second, third, and fourth choices.

The results from this simulation exercise are shown in the third column. Notice that misrepresenting preferences makes players strictly worse off. When a player ranks her true first choice as second, the probability of being matched with the true first choice drops to 28.23%. In the fourth column, we show the results from an analogous exercise, when a player misrepresents by swapping her second and third choices, i.e., plays $pref_{jt}^{23}$. Again, note that misrepresenting the preferences makes a player strictly worse off compared to truth-telling. Using similar simulations, it is possible to show that all other deviations also make players strictly worse off, compared to truthful revelation.

One possible critique of the above exercise could be that we started with the assumption that players stated-preferences are their true preferences. Therefore, we also present results from a general case, where

the player's true preferences are drawn randomly (see the bottom panel of Table A7). Again, we find that deviating from truth-telling makes users strictly worse off. In sum, all our tests confirm the validity of the truth-telling assumption in our setting.

Finally, note that there is no need to make any additional assumption on truth-telling for both *first* and *reply* messages since they are both single-agent decisions, and there is no game involved. Therefore, each player only has to follow her/his expected utilities and doesn't have to worry about the strategic behavior of other players. So, by definition, a player's revealed preferences reflect her/his expected utility.

# E Conversation History

| | (A7) |
|---|---|
| $star1_{it}$ | 0.0410055 |
| | (0.0256374) |
| | |
| $star3_{it}$ | -0.0865357*** |
| | (0.0192754) |
| | |
| $successful_{jt}$ | 0.0000904 |
| | (0.0053202) |
| | |
| $star1_{it} \times successful_{jt}$ | -0.0344794 |
| | (0.0366443) |
| | |
| $star3_{it} \times successful_{jt}$ | 0.0736327*** |
| | (0.0270534) |
| | |
| Fixed Effects ($\eta_i$) | ✓ |
| | |
| Individuals | 3494 |
| Observations | 619065 |

Standard errors (in parentheses) are clustered at the user level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A8: Heterogeneous effect of star-ratings based on the rank-giver's conversation history (using ordered logit fixed-effects model).

We now examine the heterogeneous effects of star-ratings based on rank-giver's conversation history, and provide more evidence to show that the negative effect of three-star ratings during the game stems from rejection concerns. We start by defining $conversation_{jt}$ as the average number of successful conversations that user $j$ experienced before game $t$. A successful conversation from $j$'s perspective is defined as one where $j$ either received a first message from the matched partner, or received a reply to a message that s/he had initiated with the match.

Next, we stratify users (rank-givers) based on their conversation history. In our data, a median user experiences an average of 0.016 successful conversations in her/his prior games. Based on this value, we define the binary variable $successful_{jt}$ as one if $conversation_{jt} > 0.016$ and zero otherwise. Next, we add this binary variable and its interactions with receiver's star-rating to Equation (3), and re-estimate the model. The results from this exercise are shown under model A7 in Table A8. The main effect of $star3_{it}$ (when $successful_{jt} = 0$) stays negative and significant. This indicates that when a user is shown with three stars, s/he receives lower preference-rankings from the rank-givers who have not had successful conversations in

the past. However, the interaction effect of $star3_{it} \times successful_{jt}$ is positive and significant, i.e., three stars users receives higher preference-rankings from rank-givers who experienced more successful conversations in the past. This suggests that rank-givers with a successful conversation history are less rejection-averse when they are ranking a popular user.

## F Physical Attractiveness

In this section, we examine the heterogeneous effects of star-ratings based on users' physical attractiveness, and provide additional evidence for strategic shading.

We start by stratifying users (rank-givers) based on their physical attractiveness. As summarized in Table 1, the median user has a standardized $pic\_score$ of -0.09. Based on this value, we define the binary variable $attractive_j$ which equals one if $pic\_score_j > -0.9$ and zero otherwise. Next, we add this binary variable and its interactions with receiver's star-rating to Equation (3) and re-estimate the model. The estimation results are shown in model A8, Table A9. The main effect of $star3_{it}$ (when $attractive_j = 0$) stays negative and significant and the interaction effect of $star3_{it}$ with $attractive_j$ is not statistically significant. This suggests that there is no difference in how rank-givers (attractive or unattractive) rank three-star users.

Next, we further stratify the data based on the physical attractiveness of the rank-receivers. We re-run the analysis separately for attractive receivers ($attractive_i = 1$) in model A9, and unattractive receivers ($attractive_i = 0$) in model A10. In model A9, we find that the main effect of $star3_{it}$ is negative and significant. The main effect (when $attractive_j = 0$) suggests that unattractive users give lower preference-rankings to attractive receivers. We also find that the interaction effect of $star3_{it}$ with $attractive_j$ is positive and significant. The interaction effect (when $attractive_j = 1$) implies that the attractive users give higher preference-rankings to attractive receivers. This suggests that only unattractive users avoid attractive popular users. This is consistent with our hypothesis of strategic shading due to rejection concerns since we expect unattractive users to be more concerned about being rejected, especially when they are ranking attractive users.

Next, in model A10, we re-run the analysis for unattractive receivers ($attractive_i = 0$). However, we find no significant results.[25] Thus, we find no evidence showing that users are concerned about being rejected when ranking an unattractive user.

---

[25]The number of individuals (rank-receivers $i$) in model A8 is greater than the total number of individuals in model A9 and A10 combined. This is because we do not have the attractiveness score for all rank-receivers.

|  | (A8) | (A9) | (A10) |
|---|---|---|---|
|  | All | Attractive | Unattractive |
|  | Rank-Receivers | Rank-Receivers | Rank-Receivers |
| $star1_{it}$ | 0.02746 | -0.04517 | 0.02777 |
|  | (0.02686) | (0.05061) | (0.03848) |
| $star3_{it}$ | -0.07573*** | -0.06781** | -0.05612 |
|  | (0.02093) | (0.03122) | (0.03492) |
| $attractive_j$ | -0.01350*** | -0.01949** | -0.01036 |
|  | (0.00496) | (0.00779) | (0.00727) |
| $star1_{it} \times attractive_j$ | -0.02466 | -0.05755 | -0.02964 |
|  | (0.03788) | (0.07384) | (0.05652) |
| $star3_{it} \times attractive_j$ | 0.04578 | 0.10459** | 0.02747 |
|  | (0.02924) | (0.04332) | (0.05100) |
| Fixed Effects ($\eta_i$) | ✓ | ✓ | ✓ |
| Individuals | 3477 | 1233 | 1354 |
| Observations | 544161 | 223108 | 246059 |

Standard errors (in parentheses) are clustered at the user level.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A9: Heterogeneous effect of star-ratings based on users' physical attractiveness (using ordered logit fixed-effects model).